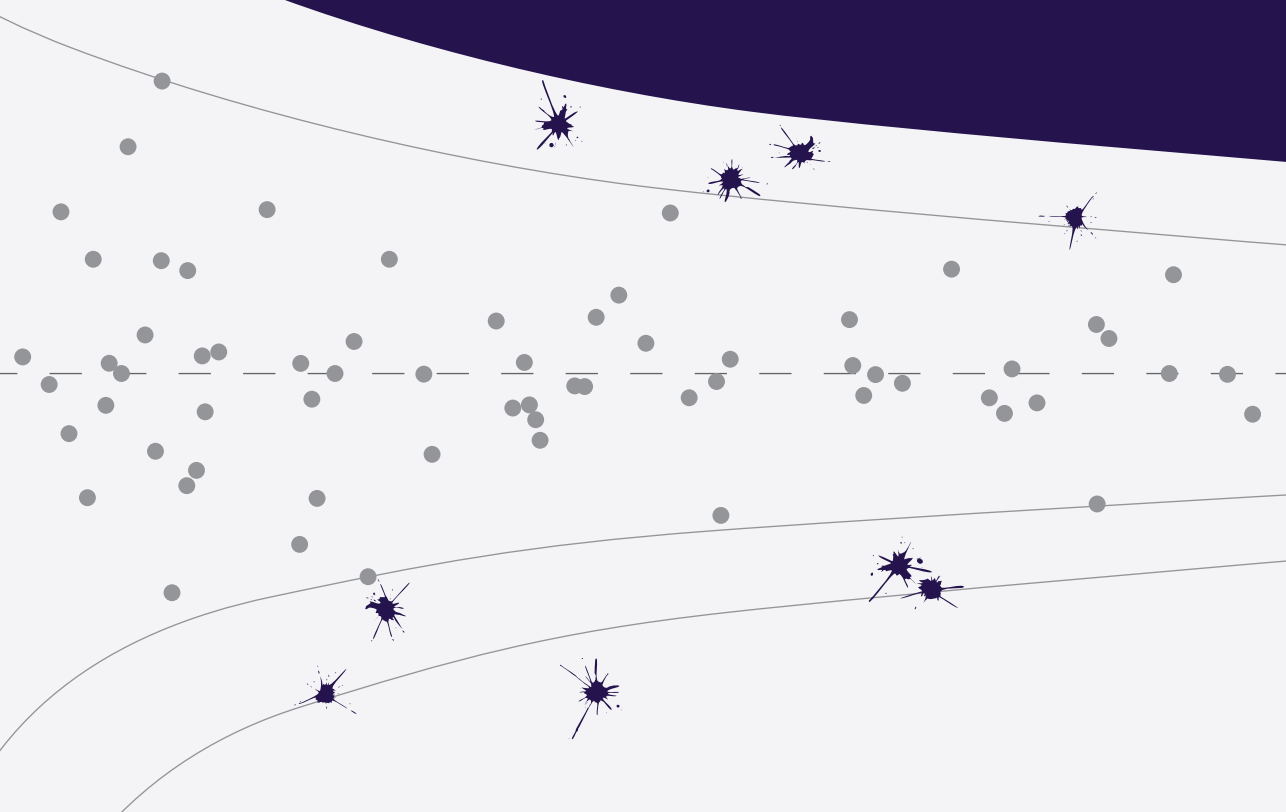
Quality indicators for hospital care

reliability and validity

Claudia Fischer



Quality indicators for hospital care

reliability and validity

Claudia Fischer

The research described in this thesis was supported by the Department of Public health, Erasmus MC Rotterdam and the Department of Public health, Amsterdam Medical Center.

Financial support for the publication of this thesis was kindly provided by the Department of Public health, Erasmus MC Rotterdam.

Design and layout: Irma Theunisse, www.irmatheunisse.nl
Printing: GVO drukkers & vormgevers B.V., Ede, The Netherlands

ISBN: 978-90-6464-921-9

Quality Indicators for Hospital Care:
Reliability and validity

Kwaliteitsindicatoren voor ziekenhuiszorg:
betrouwbaarheid en validiteit

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 9.30 uur

door

Claudia Fischer

geboren te St.Andrae/Lav, Oostenrijk

ERASMUS UNIVERSITY ROTTERDAM

PROMOTIECOMMISSIE

Promotoren: Prof.dr. E.W. Steyerberg
Prof.dr. N.S. Klazinga

Overige leden: Dr. J.A. Hazelzet
Prof.dr. J. Kievit
Prof.dr. J.J.M. Takkenberg

Copromotor: Dr. H.F. Lingsma

TABLE OF CONTENTS

Part I

Introduction

Chapter 1.

General introduction

GENERAL INTRODUCTION

Information on quality of care plays a central role in healthcare nowadays.[1] As a result of the growing demand for health care, increasing costs, and evidence of variation in quality of care, the interest in the quantitative assessment of health care quality has increased.[2,3] Furthermore, today's society demands ever more transparency, which requires the health care sector—as well as other public sectors—to provide insight into their performance.
Different stakeholders may use quality of care information for different purposes. Health care professionals use it to evaluate their performance in order to try to improve quality.[4] Government agencies use quality of care information to monitor and regulate quality of care, while insurance companies use it to select hospitals they want to contract, and patients may use it to compare hospitals in order to make informed treatment decisions.
As quality of care information is used for such a variety of purposes, it is crucial that quality measures are valid and reliable and actually do represent quality of care. However, despite a rapidly growing body of scientific literature on this topic, there is currently no consensus on how to measure quality of care.

In this thesis we focus on the measurement of quality of hospital care for external purposes. Specifically, we will study the reliability and validity of 'quality indicators'.

In this chapter some concepts related to this topic will be introduced and consequently the specific research questions and content of the rest of the thesis will be presented.

MEASURING QUALITY OF HOSPITAL CARE

Various definitions of quality of care have been formulated over the years.[3] One commonly used definition is from the Organisation for Economic Cooperation and Development (OECD) (Box 1).[5]

Box 1. Dimensions of quality of care

Quality of care can be measured on at least three key dimensions: effectiveness, person centeredness and safety.[6] Effectiveness reflects the degree to which processes result error-freely in desired outcomes.[7] Person centeredness or responsiveness means the degree to which a system functions by placing the patient at the centre of its delivery.[6,8] This component is often measured in terms of the care, communication and understanding experienced by the patient in the clinician-patient relationship.[9,10] The degree to which health care processes prevent, and ameliorate adverse outcomes, possibly resulting from the health care processes itself, is referred to as safety.[6,11]

Quality of care research has a long history. In the mid-1800s Florence Nightingale studied quality of care by assessing mortality and infection rates in British military hospitals during the Crimean War. In Austria, Ignaz Semmelweis measured and compared mortality rates related to puerperal fever between maternity clinics in Vienna (1841-1846).[12] In the beginning of the 20th century, the American physician Ernest Codman introduced a system to capture patient outcomes following surgical procedures in US hospitals.[13] These quality assessments were all initiated by health care professionals. The interest in quality and safety of care revived after the publication of the Harvard Medical Practice study in 1991 and the later publication 'To Err Is Human: Building a Safer System' by the Institute of Medicine in America. Since then the importance given to patient safety and quality of care has increased. And specifically the demand for quality assurance and transparency from external stakeholder, such as society and politics, has become stronger.[1,14-16] Nowadays, information on quality of care is used for internal and external purposes. Internal purposes include, for example, the initiation and evaluation of quality improvement programs.[17] External use of quality information involves the public comparison of hospitals based on information regarding the hospitals' quality of care. Such external comparisons may lead to quality improvement through selection of the 'best' providers by patients or payers.[17,18]

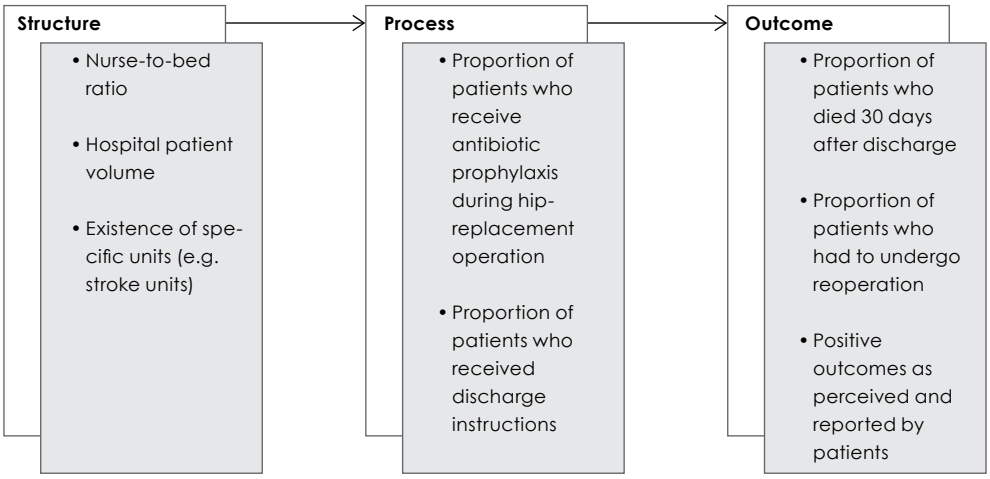
QUALITY INDICATORS

Quality indicators are measurable aspects of quality and are generally used to assess quality of care (Box 2).[19]

Box 2. Definition of quality indicators
Quality indicators are measurement tools, screens, or flags that are used as guides to monitor, evaluate, and improve the quality of patient care, clinical support services, and organizational functions that affect patient outcomes.[20]

Following Avedis Donabedian's framework, quality indicators are often classified into three types: structure, process and outcome indicators. Examples of structure, process and outcome indicators for hospital care are presented in Figure 1. Structure indicators define the characteristics of the health system or the hospital in which the care is provided,[6] such as human resources and organizational factors.[21] Structure indicators are measured at the provider or system level. Process indicators can be measured per patient and refer to the appropriateness of the delivered care,[6] such as guideline adherence. Outcome indicators reflect the end result as a consequence of care,[3] such as clinical care outcomes, adverse events or patient's satisfaction with care.

Figure 1. Structure, process, outcome model [22]



Outcome indicators

While in the scientific literature no consensus exists on which type of quality indicator is best able to assess quality of care,[23,24] there is an increasing interest in outcome indicators.[25] Outcome indicators are thought to matter most to patients and reflect all aspects of care.[24] In this thesis we study mainly outcome indicators.

Outcome indicators are usually expressed as a rate or a percentage. Thus they consist of a denominator and a numerator. The denominator is the total number of patients to which the indicator applies, e.g. all patients admitted with diagnosis x. The numerator is the number of patients with the outcome of interest, e.g. death within 30 days after admission. Examples of commonly used outcome indicators that are studied in this thesis are mortality rates, readmission rates and complication rates. These outcome indicators are expected to reflect—at least to some extent—the quality of care. They are attractive because they are very general, and thus applicable to many different patient groups.

Mortality is a frequently studied outcome indicator as it is the most undesirable outcome of care.[26] In clinical research, it is considered a 'hard' outcome. In many countries, such as the United Kingdom, the United States, Sweden, Canada, France and Australia, mortality is monitored as a measure of quality of hospital care.[27] It is most often defined as death during a hospital stay.[28,29]

Readmissions are of interest as a quality indicator as they incur costs and occur relatively frequently.[30-32] About 20% of hospitalized Medicare patients are readmitted within 30 days.[33] In several countries readmissions are monitored and also used in pay for performance schemes.[34] For example, in the US hospitals with unusually high readmission rates are financially penalized.[35]

Complication rates, such as wound infections or anastomotic leakages, are attractive outcome indicators since they are adverse events that are relatively closely related to the care process. In this thesis we will study anastomotic leakages, which are among the most severe complications in surgery and are associated with increased morbidity, reoperations, and mortality.[36,37]
In many clinical fields remarkable variation in these patient outcomes across hospitals is found, which suggests that there are differences in quality of care between hospitals and that there is room for quality improvement.[38]

PSYCHOMETRIC CRITERIA OF A GOOD OUTCOME INDICATOR

Using outcome indicators as a measure of quality of processes of hospital care requires that differences in outcomes between hospitals represent differences in underlying quality of care. Therefore outcome indicators need to be evaluated in terms of their psychometric criteria. The required quality of an outcome indicator depends on the intended use. The larger the consequences of a specific indicator score—for example, financial penalties for hospitals with higher mortality rates—the more certain it needs to be that the outcome indicator actually reflects the quality of the delivered care. When using quality indicators externally, for comparison between hospitals, at least the relevance, reliability, validity, feasibility and usability of an outcome indicator should be considered (Table 2).

Table 2. Psychometric criteria to evaluate outcome indicators[39-42]

Relevance		The outcome of interest should occur frequently or should represent an improvement opportunity.
Scientific rigor	Reliability	The indicator needs to produce the same result on repeated measurement.
	Validity	The indicator needs to measure what it claims to measure. Different forms of validity can be distinguished: face validity, content validity, construct validity, and criterion validity.
Feasibility		The data used to calculate the indicator need to be feasible to obtain.
Usability		The indicator should be understood by its intended audience and provide the ability to take action to improve the indicator score (actionability).

In this thesis we concentrate on the scientific rigor of outcome indicators for quality of hospital care. We will investigate the reliability and validity of various outcome indicators in different disease fields. Furthermore, we will study different elements that determine the reliability and validity of an outcome indicator: data quality, definitions, statistical uncertainty and case-mix (Table 3). These elements are also represented in the system that is used by the Dutch Quality Institute to judge quality indicators.[43] Below they will be discussed in detail.

Table 3. Psychometric criteria to evaluate an outcome indicator studied in this thesis

Psychometric characteristic		
Reliability:	Data quality:	Uniformity in data (collection)
	Definitions:	Accuracy of the definition of the numerator and denominator and other data elements that are used to calculate the indicator
	Statistical uncertainty:	Random variation caused by low numbers of patients/outcomes
Validity:	Case-mix:	Differences in patient populations between hospitals

RELIABILITY

Reliability refers to the degree to which an outcome indicator is reproducible.[44] The reproducibility of an outcome indicator is threatened by unsystematic errors, such as low data quality, ambiguous indicator definitions and statistical uncertainty due to low event rates, which we will study in this thesis.

Data quality

As mentioned previously, outcome indicators consist of a nominator and denominator that are based on underlying individual patient data. Sometimes other data elements are also used to calculate indicator scores, such as patient characteristics to adjust for case-mix. In order to calculate reliable outcome indicators, the underlying data need to be complete, accurate consistent and reproducible.[45]

Two different types of data sources are generally used to calculate outcome indicators: administrative data and clinical data. For our research questions on reliability we will focus on administrative data. Although not originally set up for this purpose, these data are frequently used to calculate quality indicators.[46] These data sources are attractive because they include large numbers of people (often nationwide), are available in a structured way and patient journeys to other institutions can theoretically be followed.[47-49]

In this thesis we will investigate which aspects of data quality affect the reliability of quality indicators for hospital care. We will review the current knowledge in the scientific literature, and study the effect of data quality empirically by evaluating a hospital quality indicator set of the Dutch Health Transparency Program (DHTP), which is based on administrative data.

Definitions

A second prerequisite for reliable outcome indicators is the clear and unambiguous definition of the indicator elements (nominator, denominator and other data elements that are used to calculate the indicator such as case-mix factors).[50] Different interpretations of an indicator—for example, which patients should be included in the denominator—can significantly alter the indicator score, especially when data is self-reported.[51] Even if the quality of the data is good, if the wrong data elements are delivered, the indicator scores become unreliable.

In this thesis we will investigate the clarity of the definitions of currently used quality indicators for quality of hospital care by reviewing the scientific lit-

erature on the outcome indicator readmission rate in heart failure patients. Heart failure is a complex clinical syndrome, of which exercise intolerance and fluid retention are typical characteristics.[52] In the United States approximately 500,000 new cases are diagnosed each year and 300,000 people die from heart failure annually.[53]

Approximately 800,000 hospital discharges per year can be attributed to heart failure[54] and it is a major contributor to the rising healthcare costs. More than 25% of patients hospitalized for heart failure are readmitted within 30 days after discharge.[55] Therefore readmissions are a highly relevant quality indicator for heart failure.[56] One of the most important predictors associated with adverse outcomes in heart failure patients is symptom status, which is difficult to assess, as a limited number of heart failure symptom instruments are available.[57]

Statistical uncertainty

A certain amount of variation in outcomes between hospitals is simply caused by chance. The smaller the number of treated patients and/or number of outcomes, the larger the impact of chance. As a result, quality of care differences may appear bigger than they in reality are.[58] Statistical uncertainty can be accounted for with so-called random effect regression models.

In this thesis, we will examine how statistical uncertainty affects the reliability of outcome indicators for colon cancer and oesophago-gastric (O-G) cancer surgery. We will use two clinical datasets: the Dutch Surgical Colorectal Audit (DSCA) data and the National Oesophago-gastric Cancer Audit (NOGCA) database. It is estimated that in 2015 there will be 132,700 new cases of colorectal cancer in the United States.[59] Almost 90% of colorectal cancers arise from benign, adenomatous polyps lining the wall of the bowel. When cancer develops, the polyps grow to a larger size and take on a villous appearance or they contain dysplastic cells.[60] Surgery is the main treatment for colon cancer. Usually the part of the colon affected by the tumour is removed as well as close lymph nodes. The two ends of the colon are then reconnected.

The American Cancer Society estimates that in 2015 there will be 93,000 new cases of colon cancer in the United States.[59] The prognosis for O-G cancer patients remains poor with a 5-year survival rate of 5-10% for oesophageal[61] and 10–30% for gastric cancer.[62-64] Treatment options are determined by the disease stage and patient's general health. For local disease surgical resection is regarded as the cornerstone treatment.[65]

For both colon and O-G cancer we will study the most commonly used short-term outcome indicators, namely mortality and the surgical complication anastomotic leakage.

VALIDITY

Validity refers to whether an indicator measures what it claims to measure.[66] An outcome indicator is valid if a hospital that delivers good quality of care—as represented in structure and process indicators—shows desirable outcomes (Figure1).[67]
Different forms of validity can be distinguished. Face validity refers to the extent to which, subjectively viewed, an indicator measures what it is intended to measure. This is usually determined by expert committees.[68] Content validity represents the extent to which an indicator samples the relevant sub-dimensions of quality of hospital care.[41,68] Construct validity describes the extent to which an indicator correlates with another indicator that aims to measure the same underlying construct.[68] Criterion validity reflects the extent to which an indicator shows agreement with a gold standard of the measured domain in quality of care.[68] When no gold standard exists, as is the case with quality of care, the common approach to measure validity is construct validity. Which means for example assessing the correlation between the process indicator 'percentage of patients who received antibiotics prior to surgery' and the outcome indicator 'postoperative infections', as both indicators aim to measure the underlying quality construct safety of surgery.

The validity of an indicator is threatened by systematic errors. An indicator may be reliable, but not valid, for example when there is systematic misclassification. However, if an indicator is not reliable, it cannot have high validity.

Insufficient case-mix correction, a factor that threatens the validity of outcome indicators of hospital care, is studied in this thesis.

Case-mix
Different patients are seen at different hospitals. For example, for cancer surgery high-volume centres or teaching hospitals often treat more difficult patients than small or non-teaching hospitals treat.[69] Therefore, differences in outcome do not only reflect differences in quality of care but also differences in case-mix. The case-mix differences can be accounted for with a 'case-mix correction model', which is a logistic regression model that estimates the probability of the outcome for each patient based on his or her characteristics. These expected outcomes can be compared to the observed outcomes.

In this thesis we will investigate the extent to which case-mix correction affects the validity of outcome indicators by reviewing the scientific literature and conducting empirical research using the two previously mentioned clinical databases: the DSCA on colon cancer and the English audit data on O-G cancer. Using these data we will develop a case-mix model to adjust for differences in patient characteristics when using outcome indicators and we will test the extent of the effect of case-mix correction on the validity of outcome indicators.

AIMS AND OUTLINE

The main aim of this thesis is to expand our knowledge on how to measure quality of hospital care for external purposes. We will study the reliability and validity of outcome indicators. Specifically we will focus on data quality, indicator definitions, statistical uncertainty and case-mix correction.

The specific research questions are:

1. *Which aspects of data quality affect the reliability of quality indicators for hospital care?*
2. *How clear are the definitions of currently used quality indicators for hospital care?*
3. *To what extent does statistical uncertainty affect the reliability of outcome indicators for surgical colon and oesophago-gastric cancer care?*
4. *To what extent does case-mix correction affect the validity of outcome indicators for surgical colon and oesophago-gastric cancer care?*

Outline of this thesis

The thesis is structured in five parts with a total of ten chapters. Part I includes this general introduction (Chapter 1).

Part II and III , the main parts of this thesis, include literature reviews and empirical studies.

Part II (Chapter 2-4) includes literature studies, systematic reviews and a meta-analysis. This part extensively describes the current scientific knowledge on both the reliability and validity of quality indicators for hospital care. In Chapter 2 we investigate the evidence on the validity of the outcome indicator readmission rates in hospitals in the European context. Chapter 3 summarizes the problems with the reliability and validity of readmission rates as an outcome indicator for quality of hospital care in general. Chapter 4 focuses specifically on heart failure.

Part III contains empirical studies on the reliability and validity of quality indicators for hospital care. Chapter 5 describes the extent to which data quality and indicator computation strategies affect the reliability of hospital quality indicators. For this research we use breast cancer and hip replacement indicators from the Dutch Health Care Transparency Program (DHTP) and conduct a survey of 42 Dutch hospitals. In Chapter 6 we investigate the effect of data quality on the correlation between the indicators within the

hip replacement indicator set of DHTP. Chapter 7 examines the effects of statistical uncertainty and case-mix correction on the reliability and validity of outcome indicators for the quality of colon cancer resection in the DSCA data. In Chapter 8 we develop case-mix correction models for comparing 30- and 90-day rates for mortality and for anastomotic leakage after O-G cancer resections between hospitals in the English audit data. Using the same data, Chapter 9 assesses differences in case-mix corrected outcomes between hospitals and surgeons, as well as the correlation between outcomes and the structure indicator volume.

Table 4 presents an overview of the outcome indicators under investigation, the clinical fields, the data sources including year and country, and the chapter in which the indicator is studied.

Table 4. Outcome indicators for the quality of hospital care studied in this thesis

Outcome indicator	Clinical field	Datasource	Country	Year	Chapter
Readmission	General General Heart failure	Literature Literature Literature	International International International	1999-2010 2001-2013 / -2014	2 3 4
Relapse	Breast cancer	Administrative	NL	2009-2011	5
Complications	Hip and knee replacement Colon cancer Oesophago-gastric cancer	Administrative Clinical Clinical	NL NL UK	2009-2011 2011-2012 2011-2013	5 and 6 7 8 and 9
Mortality*	Colon cancer Oesophago-gastric cancer	Clinical Clinical	NL UK	2011-2012 2011-2013	7 8 and 9

*postoperative mortality, 30-day mortality, 90-day mortality.

In the General Discussion (Part IV, Chapter 10) the results of the studies in the previous chapters are discussed as well as their implications for research and policy.

REFERENCES

1. Smith PC ME, Papanicolas I, Leiterman S. Performance measurement for health system improvement. Experiences, challenges and prospects. Mossialos E, editor. USA: Cambridge University Press, New York; 2009.
2. Groene O, Skau JK, Frolich A. An international review of projects on hospital performance assessment. Int J Qual Health Care. 2008 Jun;20(3):162-71.
3. Campbell SM, Roland MO, Buetow SA. Defining quality of care. Soc Sci Med. 2000 Dec;51(11):1611-25.
4. Grol R. Improving the quality of medical care: building bridges among professional pride, payer profit, and patient satisfaction. JAMA. 2001 Nov 28;286(20):2578-85.
5. Arah OA, Westert GP, Hurst J, Klazinga NS. A conceptual framework for the OECD Health Care Quality Indicators Project. Int J Qual Health Care. 2006 Sep;18 Suppl 1:5-13.
6. Keeley E HJ. Health Care Quality Indicators Project: Conceptual Framework Paper, OECD Health Working Papers, No. 23, OECD. 2006; Available from: http://dx.doi.org/10.1787/440134737301 Accessed 01 April 2015.
7. Juran J GB. Juran's Quality Handbook. (New York: McGraw Hill). 2000.
8. World Health Organization. The World Health Report 2000. Health Systems: Improving Performance. Geneva: WHO, 2000.
9. Scott RA, Aiken LH, Mechanic D, Moravcsik J. Organizational aspects of caring. Milbank Q. 1995;73(1):77-95.
10. Roter DL, Stewart M, Putnam SM, Lipkin M, Jr., Stiles W, Inui TS. Communication patterns of primary care physicians. JAMA. 1997 Jan 22-29;277(4):350-6.
11. National Patient Safety Foundation. Agenda for research and development in patient safety. Chicago, IL: National Patient Safety Foundation, 2000.
12. Starr P. The Social Transformation of American Medicine. New York: Basic Books1982.
13. Ross TK. Health care quality management, tools and applications2014.
14. Brennan TA, Leape LL, Laird NM, Hebert L, Localio AR, Lawthers AG, et al. Incidence of adverse events and negligence in hospitalized patients: results of the Harvard Medical Practice Study I. 1991. Qual Saf Health Care. 2004 Apr;13(2):145-51; discussion 51-2.
15. Loeb JM. The current state of performance measurement in health care. Int J Qual Health Care. 2004 Apr;16 Suppl 1:i5-9.
16. Kohn L CJ, Donaldson M., eds. To err is human: building a safer health system. Washington, DC: Committee on Quality of Health Care in America, Instiute of Medicine. National Academy Press, 2000.
17. Agency for Healthcare Research and Quality. Uses of quality measures. [updated May 29. 2014].
18. Hibbard JH. What can we say about the impact of public reporting? Inconsistent execution yields variable results. Ann Intern Med. 2008 Jan 15;148(2):160-1.
19. Spiegelhalter DJ. Handling over-dispersion of performance indicators. Qual Saf Health Care. 2005 Oct;14(5):347-51.
20. Canadian Council on Health Services Accreditation. A guide to the development and use of performance indicators. Ottwaw: Canadian Council on Health Services Accreditation1996.
21. Mainz J. Defining and classifying clinical indicators for quality improvement. Int J Qual Health Care. 2003 Dec;15(6):523-30.
22. Donabedian A. Quality assurance. Structure, process and outcome. Nurs Stand. 1992 Dec 2-8;7(11Suppl QA):4-5.
23. 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.
24. Mant J. Process versus outcome indicators in the assessment of quality of health care. International Journal for Quality in Health Care. 2001;13(6):475-80.
25. Brand CA, Martin-Khan M, Wright O, Jones RN, Morris JN, Travers CM, et al. Development of quality indicators for monitoring outcomes of frail elderly hospitalised in acute care health settings: study protocol. BMC Health Serv Res. 2011;11:281.
26. Lilford R, Pronovost P. Using hospital mortality rates to judge hospital performance: a bad idea that just won't go away. Bmj. 2010;340.

27. Jarman B, Pieter D, van der Veen AA, Kool RB, Aylin P, Bottle A, et al. The hospital standardised mortality ratio: a powerful tool for Dutch hospitals to assess their quality of care? Quality and Safety in Health Care. 2010;19(1):9-13.
28. Walters DM, McMurry TL, Isbell JM, Stukenborg GJ, Kozower BD. Understanding mortality as a quality indicator after esophagectomy. The Annals of thoracic surgery. 2014 Aug;98(2):506-11; discussion 11-2.
29. Metersky ML, Waterer G, Nsa W, Bratzler DW. Predictors of in-hospital vs postdischarge mortality in pneumonia. Chest. 2012 Aug;142(2):476-81.
30. Ross JS, Chen J, Lin Z, Bueno H, Curtis JP, Keenan PS, et al. Recent national trends in readmission rates after heart failure hospitalization. Circ Heart Fail. 2010 Jan;3(1):97-103.
31. Lindenauer PK, Bernheim SM, Grady JN, Lin Z, Wang Y, Wang Y, et al. The performance of US hospitals as reflected in risk-standardized 30-day mortality and readmission rates for medicare beneficiaries with pneumonia. J Hosp Med. 2010 Jul-Aug;5(6):E12-8.
32. Epstein AM, Jha AK, Orav EJ. The relationship between hospital admission rates and rehospitalizations. N Engl J Med. 2011 Dec 15;365(24):2287-95.
33. Jencks SF, Williams MV, Coleman EA. Rehospitalizations among patients in the Medicare fee-for-service program. N Engl J Med. 2009 Apr 2;360(14):1418-28.
34. Desai AS, Stevenson LW. Rehospitalization for heart failure: predict or prevent? Circulation. 2012 Jul 24;126(4):501-6.
35. Joynt KE, Jha AK. A path forward on Medicare readmissions. N Engl J Med. 2013 Mar 28;368(13):1175-7.
36. Rullier E, Laurent C, Garrelon JL, Michel P, Saric J, Parneix M. Risk factors for anastomotic leakage after resection of rectal cancer. The British journal of surgery. 1998 Mar;85(3):355-8.
37. Matthiessen P, Hallbook O, Andersson M, Rutegard J, Sjodahl R. Risk factors for anastomotic leakage after anterior resection of the rectum. Colorectal Dis. 2004 Nov;6(6):462-9.
38. Elferink MA, Wouters MW, Krijnen P, Lemmens VE, Jansen-Landheer ML, van de Velde CJ, et al. Disparities in quality of care for colon cancer between hospitals in the Netherlands. Eur J Surg Oncol. 2010 Sep;36 Suppl 1:S64-73.
39. Dimick JB. What makes a "good" quality indicator? Arch Surg. 2010 Mar;145(3):295.
40. Patwardhan M, Fisher DA, Mantyh CR, McCrory DC, Morse MA, Prosnitz RG, et al. Assessing the quality of colorectal cancer care: do we have appropriate quality measures? (A systematic review of literature). J Eval Clin Pract. 2007 Dec;13(6):831-45.
41. Gooiker GA, Kolfschoten NE, Bastiaannet E, van de Velde CJ, Eddes EH, van der Harst E, et al. Evaluating the validity of quality indicators for colorectal cancer care. J Surg Oncol. 2013 Dec;108(7):465-71.
42. Kimberlin CL, Winterstein AG. Validity and reliability of measurement instruments used in research. Am J Health Syst Pharm. 2008 Dec 1;65(23):2276-84.
43. Zichtbare Zorg Ziekenhuizen. Leeswijzer bij de signaalvlaggen Zichtbare Zorg Ziekenhuizen versie 2010 over het verslagjaar 20092010 versie 16-03-2010. Zichtbare Zorg.
44. Tsuang MT TM, Jones P, . Textbook of psychiatric epidemiology. Sons JW, editor2011.
45. Pringle M, Wilson T, Grol R. Measuring "goodness" in individuals and healthcare systems. Bmj. 2002 Sep 28;325(7366):704-7.
46. Aylin P, Bottle A, Majeed A. Use of administrative data or clinical databases as predictors of risk of death in hospital: comparison of models. Bmj. 2007 May 19;334(7602):1044.
47. Iezzoni LI. Assessing quality using administrative data, measuring quality, outcomes, and cost of care using large databases, The Sixth Regenstrief Conference. Ann Int Med. . 1997;127:666-74.
48. Iezzoni LI. Risk adjustment for measuring health care outcomes In: Press HA, editor. 3rd ed. United States of America2003. p. 83-138.
49. Holt PJ, Poloniecki JD, Hofman D, Hinchliffe RJ, Loftus IM, Thompson MM. Re-interventions, readmissions and discharge destination: modern metrics for the assessment of the quality of care. Eur J Vasc Endovasc Surg. 2010 Jan;39(1):49-54.
50. Arts DG, De Keizer NF, Scheffer GJ. Defining and improving data quality in medical registries: a literature review, case study, and generic framework. J Am Med Inform Assoc. 2002 Nov-Dec;9(6):600-11.
51. Huff ED. Comprehensive reliability assessment and comparison of quality indicators and their components. Journal of clinical epidemiology. 1997 Dec;50(12):1395-404.

52. Jessup M, Abraham WT, Casey DE, Feldman AM, Francis GS, Ganiats TG, et al. 2009 focused update: ACCF/AHA Guidelines for the Diagnosis and Management of Heart Failure in Adults: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines: developed in collaboration with the International Society for Heart and Lung Transplantation. Circulation. 2009 Apr 14;119(14):1977-2016.
53. American Heart Association: Heart Disease and Stroke Statistics-2008 Update. 2009; Available from: http://www.americanheart.org/downloadable/heart/1200082005246HS_Stats%202008final.pdf Accessed on 28 November 2014.
54. Collins SP, Pang PS, Fonarow GC, Yancy CW, Bonow RO, Gheorghiade M. Is hospital admission for heart failure really necessary?: the role of the emergency department and observation unit in preventing hospitalization and rehospitalization. J Am Coll Cardiol. 2013 Jan 15;61(2):121-6.
55. Patel UD, Greiner MA, Fonarow GC, Phatak H, Hernandez AF, Curtis LH. Associations between worsening renal function and 30-day outcomes among Medicare beneficiaries hospitalized with heart failure. Am Heart J. 2010 Jul;160(1):132-8 e1.
56. Jha AK, Orav EJ, Epstein AM. Public reporting of discharge planning and rates of readmissions. N Engl J Med. 2009 Dec 31;361(27):2637-45.
57. Kyoung Suk L. Symptom assessment and management in patients with heart failure. Theses and Dissertations-Nursing. Paper 2. http://uknowledge.uky.edu/nursing_etds/2 [Accessed 15.05.2015]. 2012.
58. Lingsma HF, Steyerberg EW, Eijkemans MJ, Dippel DW, Scholte Op Reimer WJ, Van Houwelingen HC, et al. Comparing and ranking hospitals based on outcome: results from The Netherlands Stroke Survey. QJM. 2010 Feb;103(2):99-108.
59. American Cancer Society. What are the key statistics about colorectal cancer? 10.15.2014 [updated 02.27.2015]; Available from: http://www.cancer.org/cancer/colonandrectumcancer/detailedguide/colorectal-cancer-key-statistics [Accessed: 14.05.2015].
60. Peipins LA, Sandler RS. Epidemiology of colorectal adenomas. Epidemiol Rev. 1994;16(2):273-97.
61. Jemal A, Siegel R, Xu J, Ward E. Cancer statistics, 2010. CA Cancer J Clin. 2010 Sep-Oct;60(5):277-300.
62. Green D, Ponce de Leon S, Leon-Rodriguez E, Sosa-Sanchez R. Adenocarcinoma of the stomach: univariate and multivariate analysis of factors associated with survival. Am J Clin Oncol. 2002 Feb;25(1):84-9.
63. Msika S, Benhamiche AM, Jouve JL, Rat P, Faivre J. Prognostic factors after curative resection for gastric cancer. A population-based study. Eur J Cancer. 2000 Feb;36(3):390-6.
64. Harrison LE, Karpeh MS, Brennan MF. Extended lymphadenectomy is associated with a survival benefit for node-negative gastric cancer. J Gastrointest Surg. 1998 Mar-Apr;2(2):126-31.
65. Koppert LB, Lemmens VE, Coebergh JW, Steyerberg EW, Wijnhoven BP, Tilanus HW, et al. Impact of age and co-morbidity on surgical resection rate and survival in patients with oesophageal and gastric cancer. The British journal of surgery. 2012 Dec;99(12):1693-700.
66. Mainz J. Developing evidence-based clinical indicators: a state of the art methods primer. Int J Qual Health Care. 2003 Dec;15 Suppl 1:i5-11.
67. Shojania KG, Forster AJ. Hospital mortality: when failure is not a good measure of success. CMAJ. 2008 Jul 15;179(2):153-7.
68. Haller G, Stoelwinder J, Myles PS, McNeil J. Quality and safety indicators in anesthesia: a systematic review. Anesthesiology. 2009 May;110(5):1158-75.
69. Kolfschoten NE, Marang van de Mheen PJ, Gooiker GA, Eddes EH, Kievit J, Tollenaar R, et al. Variation in case-mix between hospitals treating colorectal cancer patients in the Netherlands. European Journal of Surgical Oncology (EJSO). 2011;37(11):956-63.

Part II

Current knowledge on the reliability and validity of quality indicators for hospital care

Chapter 2.

The validity of the quality indicator readmission rate in the European literature

Claudia Fischer, Helen A Anema, Niek S Klazinga

This article was published as:
The validity of indicators for assessing quality of care: a review of the European literature on hospital readmission rate.
European Journal of Public Health. 2012; 22(4):484-91.

ABSTRACT

Background: Quality indicators are increasingly implemented in Europe for policy and management purposes. Many of these indicators were initially developed and implemented in the USA. However, the suitability of directly adopting indicators that have been developed in a different health care system can be questioned. Therefore, we investigate the validity behind the readmission rate indicator in the European setting.

Methods: A systematic literature study was conducted to identify the status of scientific research on the validity of this indicator (January 1999 and April 2010). Descriptive information as well as information on the data source, indicator definition, risk adjustment factors, and conclusions was assessed.

Results: The majority of the 486 included studies focused on the actual use of the indicator as an outcome measure in European countries. Only 21 studies specifically addressed its validity, or important prerequisites of validity. There is little consensus over the timeframe used to calculate the indicator, the type of readmission that is included, and the case-mix correction applied.

Conclusions: Despite the increase in Europe of the use of the readmission rate as a measure of quality of care, the amount of research performed on its validity is scarce. Those studies that report on validity replicate earlier, mainly US findings (<1999) of methodological problems and express reservations on its large-scale use. The readmission rate as an indicator should be used with care. Users should address issues related to definition, timeframe and case-mix correction as part of the process to enhance validity in the European settings.

INTRODUCTION

Recent transformations in European (EU) health care systems, such as the change from a government-driven health care to a free market system, ask for an objective assessment of the quality of hospital care.[1] Performance indicators, measurement tools that assess a particular health care structure, process or outcome,[2] are generally expected to provide such objective assessments of health care quality. However, in order to do so, the structures (e.g. organization of care), processes (what has been done to a patient) and outcomes (final health status) that are measured should be interrelated, and should be informative about the same underlying construct.

An important starting point in quality of care research has its origin in the USA in the 1970s. Here, the interest in quality of care increased substantially when the Healthcare Cost and Utilization Project (HCUP) developed a set of health care quality indicators in 1989.[3] In those years, a substantial amount of knowledge became available on the difficulties that arise when developing and maintaining valid and reliable health care quality indicators.[4] Ashton et al.[5], for example, aimed to resolve the (at that time current) controversy in the USA over the validity of early readmissions as a measure of quality of care. Despite the results being somewhat conflicting, the summary odds ratios indicated that the process of medical care does affect the risk for readmission within 30 days. It was suggested however to handle the reported figures cautiously, as a null effect could not be ruled out.

In many health care systems the rate of readmissions (RR) to a hospital has become a well-known indicator which measures how many patients are readmitted to the hospital after they have been discharged. The rationale behind monitoring readmissions is that a readmission is related to substandard care. Within the Donabedian model, this indicator can be considered an intermediate outcome indicator, as it is a proxy for the rate of adverse events or positive outcomes such as increased life expectancy or reduced morbidity (genuine outcome indicator). The fact that hospital readmissions are high in cost,[6] a burden to patients,[7] and can be easily computed from routine statistics of administrative databases,[8] makes it plausible that readmission scores are monitored. Indeed over a decade ago several studies, mostly performed in the USA, report that readmissions were valid measures of quality as they appear to be largely caused by substandard care received during the prior hospitalization.[6,9,10-12] However, other studies failed to confirm a valid relation between the quality of care and the RR.[5,8] In all it seemed that the validity of the RR as a measure of health care quality was still debated.

Within the last decade, Europe followed up on the USA's example and several countries implemented performance indicators that were originally developed there. Since local health care factors, such as the proximity of the hospital and the availability of the beds,[13,14] influence the probability of a readmission, and the European health care systems are substantially different from that of the USA (both at the governmental and hospital level) the validity of the RR might be comprised.

Our aim is to explore the validity behind the RR indicator in the European setting. The validity is assessed by looking at the consistency in the use of definition of the indicator, the readmission timeframe used and the use of case-mix correction.

METHODS

Search strategy and selection criteria

A systematic literature search was conducted in the electronic databases Medline and PubMed, using the following keywords in various combinations: "re-admission", "readmission", "rehospitalisation", "re-hospitalization" and MEDLINE subject heading (MeSH term) "patient readmission", present in either the title and/or the abstract. The search was limited to publications from Europe and to the time period from January 1999 to April 2010. An English abstract was required to be present in order to be able to include or exclude the study. After we screened the abstracts of included studies, two groups of articles that differed largely in magnitude were identified: (i) studies using the RR as an outcome indicator and (ii) studies testing the indicators validity.

Procedure for classification and evaluation

The classification and evaluation process was performed by two independent researchers, who discussed any disagreement. If necessary, a third researcher was consulted. From all included articles, the following data elements were extracted: Year of publication, country of affiliation of the corresponding author and the patient group/disease being focused on. Further, we checked whether they define the RR they are focussing at, and how they define it. To provide more insight into current opinion on the validity, we screened the articles focusing on the validity of the quality indicator in more depth using information from the title, the abstract and full text. The following data elements were abstracted: type of validity investigated, study design used, type of data sources, type of case-mix correction and the conclusion. We distinguished between three types of validity: (i) face validity, the extent to which the measure appears to assess the construct, (ii) construct validity, the degree to which the measure reflects the construct and is related to other variables in predicted ways, (iii) criterion validity, the validity that relates to the ability of a measure to predict an outcome (criterion), this measure ideally gets evaluated against a 'gold standard'.[15]

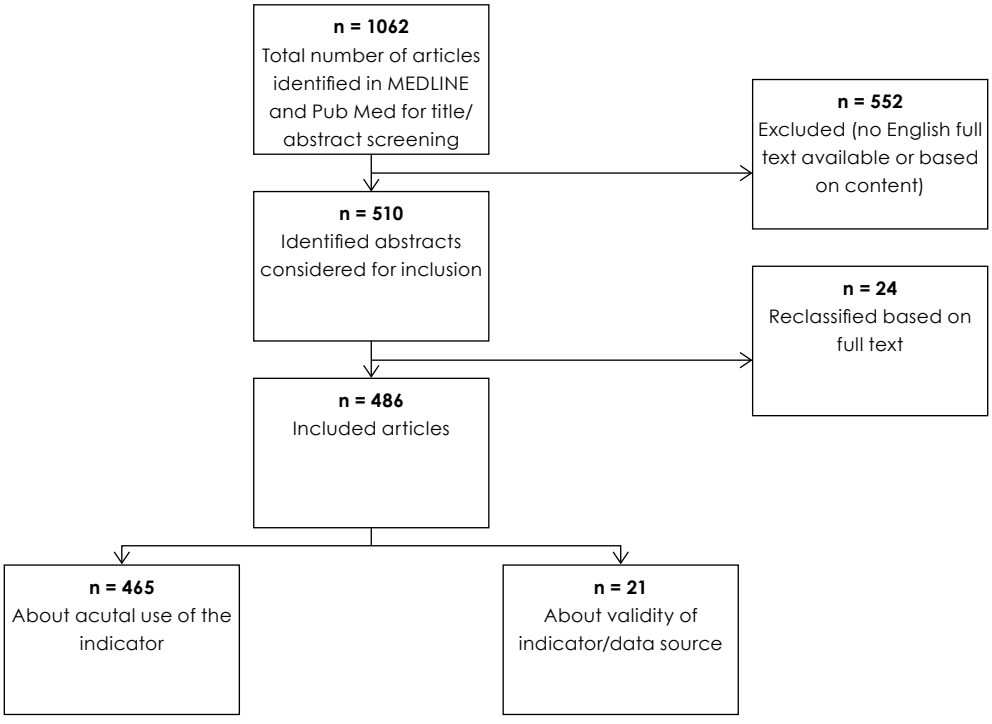
RESULTS

The search identified 1062 publications (see Figure 1 for flow diagram). After the initial screening, 552 articles were excluded on the basis of our inclusion criteria or a missing full English text. Of the 207 studies which did not provide full texts, most studies were published in Germany (n=55), Spain (n=38), France (n=29) and Denmark (n=25). The remaining 510 articles were screened on basis of the full text, 24 articles were reclassified. The resulting 486 articles were evaluated by the reviewers of which 465 articles focused on the actual use of the indicator as an outcome measurement and 21 studies somehow addressed the validity of the indicator, see for further bibliometric results Figure 2 A, B and C.

The RR is a poorly defined but increasingly used outcome indicator

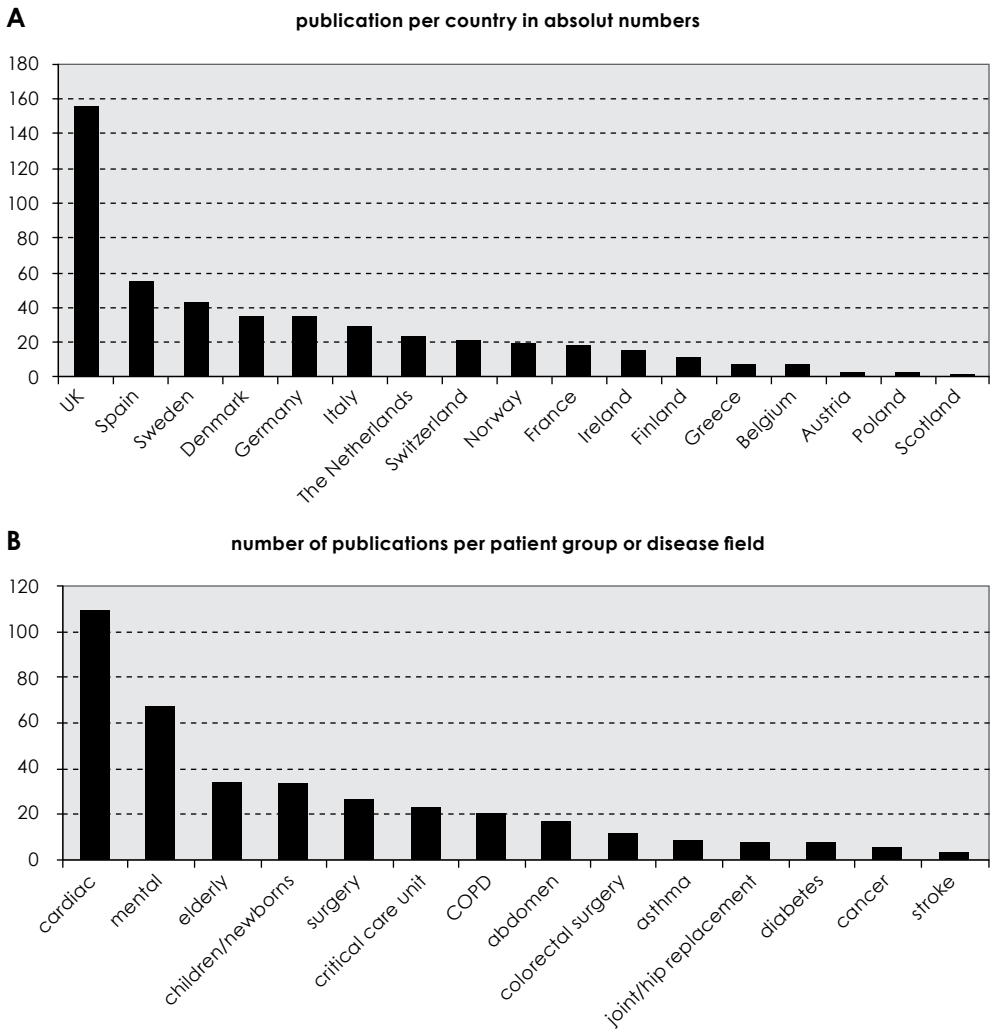
Of the 465 articles which use the RR as an outcome measure in their study, we found 288 (partly) defining what they mean by RR. 177 studies just stated that they used the RR as an outcome measure, without defining it. 263 arti-

Figure 1. Flow diagram of article selection procedure

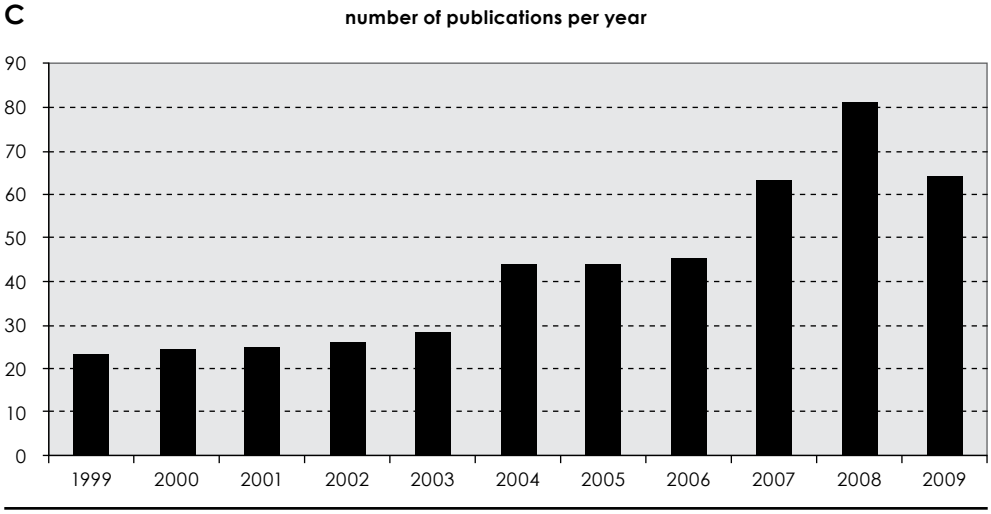


cles out of the 288 defined the timeframe they were investigating, 13 studies looked at readmissions within 14 weeks, while 80 studies used the timeframe 'within 30 days'. However, most of the articles applied 'longer than 30 days' (n=166). Those who defined the type of RR (70 studies out of the 288) used most often 'unplanned/emergency readmission' (n=51). Just 45 studies stated both, the timeframe and type of readmission they included.

Figure 2. (A) number of publications per country in absolute numbers, (B) number of publications per patient group or disease field and (C) number of publications per year



C



In the following paragraph, the 21 validity studies that address the quality of the indicator will be the topic of discussion. The studies will be discussed in regards to the following aspects: definition, type of validity and adjustment for case-mix.

As is described above, 21 European articles could be identified that address validity or validity aspects of the RR. Seven studies were included on basis of their abstract in which they stated to investigate the validity of the indicator by using the term validity or validation.[16-22] The rest was identified on basis of the full-text information. In that case, the authors inferred the validity topic based on whether they questioned validity related factors, such as the influence of case-mix/ hospital factors, or the influence of different definitions on the RR score.

Data sources used varied. Three validity studies used data from an administrative database. Of those three, one study used the MODCOD system (encoded reports), allowing to compute DRGs according to ICD-9-CM codes (CPHA, 1979) and the UNIDOC system, which is an integrated patient report processor.[23] One study used a hospital information database (PAS)[24] and the third used the hospital's 'Clinicom' patient administration system with an in-built 28-day re-admission search tool.[25] Other data source examples are: data collected by statistic offices (n=2) [16,17] hospital discharge data (n=1),[18] hospital information systems (n=1),[26] electronic coding systems (n=1),[27] and cancer registries (n=1).[19] Two studies did not provide details about the source of the studied data.[28,29]

Different ways to address validity

Diverse aspects of an indicator's validity can be addressed (such as face-, construct-, criterion validity, or validity threatening factors such as the time window used, hospital and patient characteristics). The criterion validity of an indicator can be addressed by several ways, such as investigating the predictability of an outcome indicator, or directly comparing performance which is revealed by the performance indicator to a golden standard (expert judgement of care based on record review).

In our review, we observed in total 14 studies that investigated the validity of the RR (13 criterion validity 1 face validity,[30] no construct validity study could be identified) and five studies that did not specifically investigated or discuss valid relations between the RR and other measure of quality of care but investigated factors that might influence validity (see Table 1 for summary).[22,31-34]

The identified criterion studies are discussed below in more detail. Of the 13 criterion validity studies, 11 studies used expert judgments as a golden standard. The quality of care was merely judged by whether a readmission was unplanned/emergency (n=3),[17,23,24] unforeseen and avoidable (n=1)[16] could be avoided (n=1),[18] or unplanned and avoidable (n=3),[25,26,35] or was caused by a complication that was likely related to the surgery or whether it was provoked by the patient himself (n=3).[27-29] In two studies, the golden standard was provided by either an expert judgment about aspects of quality of care based on information from an electronic full text discharge summary,[23] or on process indicators that were derived from evidence based guidelines and were calculated from medical chart information.[20] In the latter study, the authors specifically stated that these process indicators were used as golden standard. Finally one study validated a National Patient Registry against corresponding data from the review of medical records serving as the gold standard.[19] In regards to the study purpose, four studies were set-up to investigate whether the indicator is informative of insufficient care,[18,20,27,29] or if the underlying data source is suitable for calculating the RR (n=9).[16,17,19,23-26,28,35]

This paragraph refers to the 11 studies investigating the validity threatening factors or the feasibility of the RR.[16,17,21,22,27,29,31-34,36] Eight studies[16,17,27,29,31-34] investigated validity threatening factors such as the time window used to determine the readmissions, and the factors that influence readmissions like patient and hospital characteristics. The other three studies merely focused on the feasibility of calculating the indicator,[21,36] or on the validity of linking multiple admissions to one patient without the presence of unique patient identifiers.[22] Finally, one study merely discussed general validity issues of the RR.[30]

Table 1. Validity studies

Validity Type	Author	Study aim	Design	Conclusions[a]
Face Validity	Clarke, 2004, UK[30]	To discuss of the validity of RR. (n.a.)	Opinion paper	RR is an unsatisfactory performance indicator since it seems difficult to select only the unplanned and avoidable RAs, current health care databases are not sufficient to detect RAs to other hospitals, and there is no uniform time window in which RA is measured.
Criterion validity	Kossovsky, 1999, Switzerland[23]	To compare databases to determine which information is needed to distinguish planned from unplanned RA. [b,c]	Retrospective case-control study	Electronic reports based on diagnostic and procedure codes alone are not sufficient to distinguish planned from unplanned RA, automation of detailed (full text) clinical databases seems promising.
	Leng, 1999, UK[17]	To investigate possible factors influencing the RR: frequency of previous admission and cause of RA. [c,d]	Comparative research	RR is unlikely to be a valid outcome indicator, until better routine data for standardization by case-mix is available
	Kossovsky, 2000, Switzerland[29]	To determine relation between early unplanned RAs of heart failure patients and suboptimal in-hospital care, or whether RA relates to clinical and demographic characteristics of the patient. [b,c]	Retrospective case-control study	RR is not a valid measure of the quality of care for heart failure patients; unplanned RAs are more strongly associated with clinical and demographic characteristics than with quality of care. Case-mix correction is needed
	Pearson, 2002, UK[28]	To compare views of general practitioners and hospital staff on the reasons for unplanned RA of elderly people. [d,e]	Observational research	No specific measure for agreement was given. GP's and hospital staff judged most RAs to be caused by a relapse or complication of index illness. Opinions differed most significantly when the reason for RAs was poor health on discharge or inadequate preparation on discharge.
	Halfon, 2002, Switzerland[26]	To develop a computerized method that is able to detect potentially avoidable hospital RAs on basis of routinely collected data and to develop a prediction model that adjusts the RR for case-mix factors. [d,f]	Evaluative retrospective cohort study	Detection of avoidable RA by computerized method was scientifically sound enough to sign quality issues. Medical risk adjusters were more important than non-medical patient characteristics. But RR is sign to gather further information from medical records, and not for public reporting.

	Courtney, 2003, UK[25]	To examine reasons for RA, possible errors in coding and any preventable factors in acutely readmitted patients. [d,f]	3 month retrospective audit	Comparing crude rates of RA does not quantify number of avoidable RAs and is only useful as a sign to conduct local studies to determine avoidable RAs.
	Maurer, 2004, Switzerland[35]	To measure RR and to qualify the RAS as planned, unplanned, avoidable or unavoidable. [b,f]	Prospective study, explorative	No conclusions were drawn regarding the possibility to distinguish between planned, unplanned, avoidable or unavoidable. But, RR should be monitored regularly, with a time frame of 30 days of discharge.
	Jiménez-Puente, 2004, Spain[18]	To identify the frequency and characteristics of potentially avoidable RAs and to compare the quality of care derived from RR to the quality of care judged by experts. [b,g]	Cross-sectional observational study	Most RAs are not avoidable thus RR cannot be considered valid indicators of the quality of care for the set of specialties in a general hospital, RR depends mainly on patients clinical condition and is not uniformly defined.
	Luthi, 2004, Switzerland[20]	To measure the validity and predictive ability of RA. [d,e]	Evaluative, retrospective cohort study	RR, calculated from routinely collected data was not a valid indicator of the quality of heart failure care. Improvement of definition and measurement methods are needed as well as appropriative risk adjustments.
	Halfon, 2006, Switzerland[16]	To evaluate the usefulness of a computerized algorithm to identify avoidable RAs on the basis of minimum bias, criterion validity (expert judgment), and measurement precision. [d,h]	Evaluative retrospective cohort study	Adjusted rates of potentially avoidable RAs are scientifically sound enough to warrant their inclusion in hospital quality surveillance and a high rate acts as a signal to hospitals to evaluate their practices.
	Adeyemo, 2007, UK[24]	To assess the accuracy of hospital unplanned RA data, and identify patterns or possible causes of unplanned general surgical RAs. [c,d]	Retrospective audit of case note records	A hospital coded information database, may not be accurate enough for the calculation of unplanned RR as it allows inclusion of unrelated admissions. Also, factors like age, sex, history of psychiatric disease, number of drugs on discharge could not be used to predict unplanned RA.
	Shalchi, 2009, UK[27]	To formulate an appropriate definition of RA, to investigate the demographics and the predominant cause of readmitted patients (avoidable RA) and to see whether rapid throughput is leading to unacceptably high RRs. [e,i]	Retrospective observational study.	Older patients with more complex care needs are more likely to be readmitted, rapid throughput of patients is not associated with RA. RR needs to be interpreted with caution; it varies with changes in the inclusion criteria. However, on basis of expert judgment it appeared that most RAs can be avoided with better quality care

	Harboe, 2009, Denmark[19]	To investigate the validity of the Danish Cholecystectomy Database (DCD) by evaluating the association between PIs calculated from this database and post-operative complications. [d,e]	Evaluative retrospective cohort study	The DCD is a valid method for the monitoring of quality of care in cholecystectomy, however correct coding is important, especially with administrative data. Length of stay > 3days and/or RA were strongly associated with post-operative complications.
Validity factors	Heggestad, 2003, Norway[31]	To demonstrate the effects of different time intervals to calculate RR. (under investigation, [c])	Modeling approach	RR is susceptible to the choice of time interval. The longer the time interval, the greater the number of unrelated admissions included. The optimal cut-off point in time is additionally dependent on its use.
	Lyratzopoulos, 2005, UK[33]	To examine the effect of patient and disease factors on the risk of emergency medical RA. [b,c]	Prospective observational cohort study	Adjustment for factors like male sex, older age, diagnosis of heart failure, COPD or asthma and patient socio-economic status is necessary. Failure to do so may disadvantage hospitals serving primarily deprived communities.
	Spiegelhalter, 2005[34]	To test different options for handling over-dispersion of performance indicators. [c,d]	Retrospective analysis	To use RR as a valid indicator, the score must be corrected for substantial over-dispersion by using the random effects model.
	Mason, 2006, UK[37]	To identify suitable outcome measures for comparing gynecology performance between hospitals. [c,d]	Descriptive feasibility study	Emergency RRs after day case admissions and after elective abdominal hysterectomy are suitable comparative measures, excluding those records with a cancer diagnosis. However, the measure should not be used to make definitive judgments about hospitals.
	Tromp, 2007, The Netherlands[22]	To describe an efficient, generalizable approach to validate probabilistic record linkage results and to apply this approach to validate linkage of admissions of newborns. (n.a.)	Validation of probabilistic record linkage	The external validation procedure of record linkage is feasible, efficient, and informative about identifying the source of errors.
	Demir, 2008, UK[32]	To develop a modeling approach to tackle the issue surrounding the appropriate choice of a time window as a definition of RA. (under investigation, [c])	Modeling approach	Some support for 28 days as a valid time window for congestive heart failure patients. Little support for 28 days as a valid time window in defining RA for patients with COPD and stroke.

	Bottle, 2009, UK[21]	To apply 10 of the AHRQ indicators for use in English routine hospital admissions data as the first step in validation, and describe their rates in relation to established measures of negative outcome. [c,d]	Descriptive feasibility study	The recommended measures (RR and mortality rate) should not be used to make definitive judgments about hospitals, but as pointers as to where further investigation is needed. They minimize case-mix differences and have sufficient numbers for comparison analyses. Inaccurate and incomplete coding are potential violations of the validity of the indicator.

a:Inferences are made when authors do not describe topic clearly. b: Time window larger than within 30 days. c: Unplanned/emergency readmissions. d: Time window within 30 days. e: All readmissions. f: Unplanned and avoidable. g: Avoidable readmissions. h:Unforeseen and avoidable. i: Time window within 14 days. n.a.= not applicable, RR= Readmission Rate, RA=Readmission.

A definition that often lacks precision

The identified studies learned that the time windows and the criteria 'avoidable' and 'unplanned' were the main issues threatening the validity of this indicator. It seems that there is a large variability in the time range between discharge and readmission as it ranged from '14 days',[27] to 'within 180 days'.[18] The largest number of articles however, calculated readmissions within 30 days (28 to 30; n=11).[16,17,19-21,24-26,28,34,36] The rest used either a time range within 14 days (n=1),[27] or a time frame larger than 31 days (n=5).[18,23,29,33,35] From the total of 19 applicable studies, 10 defined the readmission as unplanned/emergency, [17,21,23,24,29,31-34,36] one defined it as avoidable[18] three studies used both avoidable and unplanned/emergency (see symbols in Table 1).[25,26,35]

Different options for adjusting to patient and disease characteristics

Six studies examined to what extent patient characteristics, demographical, social and medical factors, affected the RR.[16,17,27,29,33,34] Two out of them specifically investigated the increase in risk to be readmitted for various patient dependent variables.[29,33] First, Kossovsky et al.,[29] investigated heart failure patients in a case-control design and observed a significant increase of readmission risk for: previous diagnosis of heart failure, age, history of cardio revascularization. Odds ratio's varied from 1.14 (poor readiness for discharge) to 4.1 (age, for patients > 80). In addition the authors observed an association between readiness for discharge and a subsequent early readmission. The second one[33] conducted a case-control design and observed the following significant predictors of all cause readmission: male sex, age, number of coded co morbidities, admission via GP referral (decrease in risk), primary

diagnosis of heart failure and of chronic obstructive pulmonary disease or asthma, and higher level of deprivation. Hazard ratio's varied from 0.93 (admission via GP referral) to 1.49 (> 4 coded co morbidities).

Three studies investigated the possibility to apply satisfactory risk adjustment for patient-related factors,[16] descriptively compared the rates between different medical specialties,[17] or statistically compared the rates between medical and patient characteristics.[27] In all, there was some evidence for the influence of male gender, age, length of stay, previous hospitalization within 6 months, life threatening diseases that are prone to serious disability or complications (cancer, heart disease, high risk surgery) and disease categories such as nephrology and haematology. The last one addressed the effects of hospital factors on the variability between unadjusted readmission rates from various care institutions (so called over-dispersion), by investigating different options to reduce this variability.[34]

Finally, the resulting two studies investigated the time window to calculate the RR. While one study[32] used a modelling approach on nationally collected hospital data on three disease groups (COPD, stroke and CHF) and Bayesian classification to determine an appropriate time window, the other study applied a conceptual model to analyse all-cause unplanned readmissions (without cancer and obstetrics readmissions) on the basis of the characteristics of the risk, or hazard curve.[31] Whereas Demir et al.[32] observed some support for the 28 days time window in congestive heart failure patients, to Heggestad[31] it seems that the time used to calculate the readmission is largely dependent upon the reason why the RR is measured.

DISCUSSION

In short, the aim of this literature review was to investigate the validity behind the RR indicator, used within the European health care systems. The results revealed a substantial increase of studies reporting on the use of the RR as an outcome since 2004, of which the majority of the papers originated from the UK. However, the amount of research on the validity of the indicator stays relatively behind in comparison with the total increase in studies. Reviewing the content of the articles that used the RR but did not investigate its validity learned that only a small number specifically defined how the RR was measured and even a smaller part of the studies used recommended unplanned readmissions. Exploring the validity studies, however, learned that the majority cast doubt on its validity in the European setting. The studies highlight substantial problems with respect to the adjustment of factors that are beyond medical control but increase the risk to be admitted, the degree to which avoidable readmissions can be accurately detected and the time window that is used to detect relevant readmissions.

Limitations

Some limiting aspects of the methods used in our current review require further discussion. First, the search was limited to PubMed and Medline, other databases were not addressed. Further, in our methods we mentioned that papers were excluded not providing English full texts. As such, the total number of validity studies presented in this literature review is an underestimation of the total number. In our opinion, however, the observations reported in this review are of a robust character and it seems unlikely that inclusion of non-English literature would change the scope and conclusions. Also, as the research field of performance indicators is relatively young, sensitive search terms are not well established yet. Our experience showed that a combination of self-entered search terms and selecting relevant MeSH terms revealed the most appropriate articles.

The time window needs to be in accordance with the type of disease

The reported differences in whether RR is valid or not might be explained by the various ways that were used to measure the RR (e.g time window used). Our review showed similar variety in the use of time windows as the number of days between discharge and readmission ranged from 14 days post-discharge, to 180 days. Most validity studies calculate the RR on basis of early readmissions within 30 days, whereas the majority of the articles using the indicator as an outcome measure use a timeframe > 30 days. According to Heggestad,[31] it seems that the time used to calculate the readmission is

largely dependent upon the reason why the RR is measured, and whether it is important to have a sensitive or a specific measure. Also, the optimal time window is largely dependent on the type of disease the patient was originally treated for.[32]

Readmission needs to be avoidable and unplanned

Not only the time window varied, the type of readmission that is included varied as well, as some counted all and others counted only unplanned readmissions. The importance of this distinction is shown in the conflicting results regarding the validity of the RR, in studies that included all readmissions compared to those only including unplanned readmissions. Despite the fact that unplanned readmissions are more likely to be related to substandard care than planned readmissions, our review showed that in European publications this distinction is not always made. Perhaps this is caused by the difficulty to distinguish between planned and unplanned readmissions. Kossovsky,[23] indeed, concludes that an administrative database that consists of (discharge) diagnosis and procedure codes alone is not sufficient for that purpose. However, even the inclusion of unplanned readmissions might not be valid as unplanned readmissions could well be caused by a new infection that is not related to the previous hospital stay.[26] Instead, it is argued to focus on potentially avoidable readmissions that are unforeseen at time of discharge. In all, on the basis of their literature appraisal, Rumball-Smith and Hider[37] recommend that the RR can be defined as following: 'the number of patients who experienced unintended, acute readmission or death within 30-days of discharge from the index admission, divided by the total number of patients discharged alive within the reference pool' (p.63).

Case-mix factors have to be taken into account

Medical and demographical patient characteristics[9] such as the severity and complexity of the diagnosis of the readmitted patient, the length of stay,[6,13] and age and socio-economic status, seem to influence the validity of the RR as well.[38] The number of factors that could potentially affect the RR seems numerous, but male sex, age, number of coded co morbidities and the disease that a patient was treated for are important factors to take into account when measuring the RR.

Conclusion

With this article, we hope to increase the awareness on the methodological pitfalls of performance indicators and stimulate research activities on their validity in different national set-ups in Europe. Our literature study showed that a performance indicator of US origin, such as the RR, is increasingly used in

Europe. However, the amount of research performed on its validity remained scarce. Those studies that do report on its validity replicate earlier findings from mainly the USA (<1999) of methodological problems such as the lack of a uniform definition, the impact of case-mix factors and the questionable reliability of the databases used to compute the RR.

It seems that a 'best recipe' to calculate the RR in a valid way does not exist. Clarke (2004) nicely summarizes the challenges that arise when calculating the RR. The author suggests stopping using this indicator altogether, particularly when it is calculated from routine data sources and used for the comparison of readmission rates between different hospitals on a macro (national) level. Many databases do not allow for the tracking of patients from one hospital to another, as unique patient identifiers are less frequently used within the health care context. As a consequence, patients that are readmitted to another hospital, something that particularly happens with patients that are dissatisfied with the care they received, are missed in the calculation.[30] Walraven et al.[39] conclude that the true proportion of potentially avoidable readmissions is simply unclear.

Implications

The validity of the measure has to be strengthened for its user's purpose (i.e. managers, policy makers, and developers). Moreover, the actual purpose of the indicator is important when studying its validity as each type of use (accountability, improvement, consumer information, pay-for-performance) may place different demands on the degree of validity.[40,41] Further it is suggested to pay attention to test and enhance the indicators' local validity.

In sum, performance indicators are not those easy obtainable measures of health care quality as most users would like them to be. In fact, insufficient validity of performance indicators seems to be a more common problem[42] as can be learned from the research field of the Hospitalized Standardized Mortality Rates.[43]

Nevertheless, if the RR is measured as an indicator of quality of care, it is best to ensure that these readmissions are related to the index admission, are unplanned or even better, can be identified as avoidable. Secondly, the time window that is used to calculate the readmissions should be adapted to the type of care that is investigated and thirdly, the data used for calculating the indicator should have undergone reliability analysis. The data quality need to be of such high standard that the readmissions can be accurately related to an index admission and that patient-specific information is available to adjust

for patient factors. Only under these circumstances does the RR provide useful information about the quality of care a hospital performs.

Keypoints

Despite the increase (between 1999 and 2009) in the European use of the readmission rate as a measure of quality of care in Europe, the amount of research performed on its validity remains scarce.

- Those studies that do report on the validity of the readmission rate as a quality indicator replicate earlier findings from mainly the USA(<1999), of methodological problems such as the lack of a uniform definition, the impact of case-mix factors and the questionable reliability of the databases used to compute the readmission rate.
- When readmission rates are increasingly used for public reporting or performance payment programs, it seems worthwhile investigating their local validity before using them on a large scale for management and policy purposes.
- Users should address issues related to the indicators' definition, timeframe and case-mix correction as part of the process to enhance validity in the European settings.

REFERENCES

1. Mainz J. Defining and classifying clinical indicators for quality improvement. International Journal for Quality in Health Care 2003;15(6): 523-530.
2. Worning AM, Mainz J, Klazinga N, Gotrik JK, Johansen KS. (Policy on quality development for the medical profession). Ugeskr Laeger 1992; 154:3523-3533.
3. AHRQ Agency for Healthcare Research and Quality. Refinement of the HCUP Quality Indicators: http://www.ahrq.gov/clinic/epcsums/hcupqisum.htm Accessed (27.11.2010).
4. Brook RH, McGlynn EA, Shekelle PG. Defining and measuring quality of care: a perspective from US researchers. Int J Qual Health Care. 2000;12(4):281-295.
5. Ashton CM, Del Junco DJ, Souchek J, Wray NP, Mansyur CL. The association between the quality of inpatient care and early readmission: a meta-analysis of the evidence. Med Care. 1997; 35(10):1044-1059.
6. Ashton CM, Kuykendall DH, Johnson ML, Wray NP, Wu L. The association between the quality of inpatient care and early readmission. Annals of Internal Medicine. 1995. 122(6): 415- 421.
7. Cakir B, Gammon G. Evaluating readmission rates: how can we improve? South Med Journal. 2010. 103(11):1079-1083.
8. Ashton CM, Wray NP. A conceptual framework for the study of early readmission as an indicator of quality of care. Soc. Sci. Med. 1996; 43(11): 1533-1541.
9. Ashton CM, Kuykendall DH, Johnson ML, Wray NP. An empirical assessment of the validity of explicit and implicit process-of-care criteria for quality assessment. Med Care. 1999; 37(8):798-808.
10. Wray NP, Peterson NJ, Souchek J, Ashton CM, Hollingsworth JC. Application of an analytic model to early readmission rates within the department of veterans affairs. Med Care. 1997;35(8):768-781.
11. Weissman JS, Ayanian JZ, Chasan-Taber S, et al. Hospital readmission and quality of care. Med Care. 1999; 37(5):490-501.
12. Hannan EL, Racz MJ, Walford G., et al. Predictors of readmission for complications of coronary artery bypass graft surgery. JAMA 2003;290(6):773-780.
13. LaVela SL, Smith B, Weaver FM, Miskevics SA. Geographical proximity and health care utilization in veterans with SCI&D in the USA. Social Science & Medicine. 2004:59 (11):2387-2399.
14. Fisher ES, Wennberg JE, Stukel TA, Sharp SM. Hospital readmission rates for cohorts of medicare beneficiaries in boston and new haven. The New England Journal of Medicine. 1994:331(15):989-995.
15. Romano P. Selecting Indicators for Patient Safety at the Health Systems Level in OECD Countries. Summary of recent US experience. 2007. Entered at 22.03.2011: http://www.oecd.org/dataoecd/44/29/39495326.pdf
16. Halfon P, Eggli Y, Prêtre-Rohrbach I, Meylan D, Marazzi A, Burnand B. Validation of the potentially avoidable hospital readmission rate as a routine indicator of the quality of hospital care. Medical Care 2006;44(11):972-981.
17. Leng GC, Walsh D, Fowkes FGR, Swainson CP. Is the emergency readmission rate a valid outcome indicator? Quality in Health Care. 1999;8:234-238.
18. Jiménez-Puente A, García-Alegría J, Jorge Gómez-Aracena et al. Readmission rate as an indicator of hospital performance: The case of Spain. International Journal of Technology Assessment in Health Care 2004;20(3):385-391.
19. Harboe K M, Anthonsen K, Bardram L. Validation of data and indicators in the Danish Cholecystectomy Database. International Journal for Quality in Health Care 2009; 21 (3):160-168.
20. Luthi JC, Burnand B, McClellan WM, Pitts SR, Flanders WD. Is readmission to hospital an indicator of poor process of care for patients with heart failure? Qual Saf Health Care 2004;13:46-51.
21. Bottle A, Aylin P. Application of AHRQ patient safety indicators to English hospital data. Qual Saf Health Care 2009; 18(4):303-308.
22. Tromp M, Ravelli ACJ, Méray N, Reitsma JB, Bonsel GJ. An efficient validation method of probabilistic record linkage including readmissions and twins. Methods Inf Med 2008;47:356-363.
23. Kossovsky MP, Sarasin FP, Bolla F, Gaspoz JM, Borst F. Distinction between planned and unplanned readmissions following discharge from a department of internal medicine. Meth Inform Med 1999;38:140-143.
24. Adeyemo D, Radley S. Unplanned general surgical re-admissions – how many, which patients and why? Ann R Coll Surg Engl 2007; 89: 363–367.

25. Courtney ED, Ankrett S, McCollum PT. 28-Day emergency surgical re-admission rates as a clinical indicator of performance. AM R Coll Surg Engl 2003;85:75-78.
26. Halfon P, Eggli Y, Melle G, Chevalier J, Wasserfallen J, Burnand B. Measuring potentially avoidable hospital readmissions. J Clin Epidemiol 2002;55(6):573-587.
27. Shalchi Z, Saso S, Li HK, Rowlandson E, Tennant RC. Factors influencing hospital readmission rates after acute medical treatment. Clinical Medicine 2009; 9(5):426-430.
28. Pearson B, Skelly R, Wileman D, Masud T. Unplanned readmission to hospital: a comparison of the views of general practitioners and hospital staff. Age and Ageing 2002; 31: 141-143.
29. Kossovsky MP, Sarasin FP, Perneger TV, Chopard P, Sigaud P, Gaspoz J. Unplanned readmissions of patients with congestive heart failure: Do they reflect in-hospital quality of care or patient characteristics? The American journal of medicine 2000;109:386-390.
30. Clarke A. Readmission to hospital: a measure of quality or outcome? The value of readmission to hospital as a quality indicator is still debatable. Qual Saf Health Care 2004;13:10-11.
31. Heggestad T, Lilleeng S E. Measuring readmissions: focus on the time factor. International Journal for Quality in Health Care 2003;15(2):147-154.
32. Demir E, Chaussalet TJ, Xie H, Millard PH. Emergency Readmission Criterion: A Technique for Determining the Emergency Readmission Time Window. IEEE Trans Inf Technol Biomed 2008:12(5);644-649.
33. Lyratzopoulos G, Havely D, Gemmell I, Cook GA. Factors influencing emergency medical readmission risk in a UK district general hospital: A prospective study. BMC Emergency Medicine 2005;21:5:1.
34. Spiegelhalter DJ. Handling over-dispersion of performance indicators. Qual Saf Health Care 2005;14:347-351.
35. Maurer PP, Ballmer PE. Hospital readmissions – are they predictable and avoidable? Swiss med wkly 2004;134:606-611.
36. Mason A, Goldacre M, Meddings D, Woolfson J. Use of case fatality and readmission measure to compare hospital performance in gynaecology. BJOG 2006;113:695-699.
37. Rumball-Smith J, Hider P. The validity or readmission rate as a marker of the quality of hospital care, and a recommendation for its definition. New Zeeland Medical Journal 2009 13; 122: 32-39.
38. Corrigan JM, Martin JB. Identification of factors associated with hospital readmission and development of a predictive model. Health Serv Res. 1992; 27(1):82-101.
39. Bennett C, Jennings A, Austin PC, Forster AJ. Proportion of hospital readmissions deemed avoidable: a systematic review. CMAJ. 2011 Apr 19;183(7):391-402.
40. Pronovost PJ, Goeschel CA. Viewing health care delivery as science: challenges, benefits, and policy implications. Health Services Research. 2010;45(5 Pt 2):1508-1522.
41. Klazinga N, Stronks K, Delnoij D, Verhoeff A. Indicators without a cause. Reflections on the development and use of indicators in health care from a public health perspective. International Journal for Quality in Health Care 2001; 13(6): 433-438
42. Stelfox TH, Straus ES, Nathens A, Bobranska-Artiuch B. Evidence for quality indicators to evaluate adult trauma care: A systematic review. Crit Care Med 2011; 39 (4): 1-14.
43. Lingsma HF, Dippel DW, Hoeks SE, Steyerberg EW, Franke CL, van Oostenbrugge RJ, de Jong G, Simoons ML, Scholte op Reimer WJ, and The 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;

Chapter 3.

A systematic review on the reliability and validity of the quality indicator readmission rate

Claudia Fischer, Hester F Lingsma, Perla J Marang-van de Mheen, Dionne S Kringos, Niek S Klazinga, Ewout W Steyerberg

This article was published as:
Is the readmission rate a valid quality indicator? A review of the evidence.
PLoS ONE. 2014; 9(11):e112282.

ABSTRACT

Background: Hospital readmission rates are increasingly used for both quality improvement and cost control. However, the validity of readmission rates as a measure of quality of hospital care is not evident. We aimed to give an overview of the different methodological aspects in the definition and measurement of readmission rates that need to be considered when interpreting readmission rates as a reflection of quality of care.

Methods: We conducted a systematic literature review, using the bibliographic databases Embase, Medline OvidSP, Web-of-Science, Cochrane central and PubMed for the period of January 2001 to May 2013.

Results: The search resulted in 101 included papers. We found that definition of the context in which readmissions are used as a quality indicator is crucial. This context includes the patient group and the specific aspects of care of which the quality is aimed to be assessed. Methodological flaws like unreliable data and insufficient case-mix correction may confound the comparison of readmission rates between hospitals. Another problem occurs when the basic distinction between planned and unplanned readmissions cannot be made. Finally, the multi-faceted nature of quality of care and the correlation between readmissions and other outcomes limit the indicator's validity.

Conclusions: Although readmission rates are a promising quality indicator, several methodological concerns identified in this study need to be addressed, especially when the indicator is intended for accountability or pay for performance. We recommend investing resources in accurate data registration, improved indicator description, and bundling outcome measures to provide a more complete picture of hospital care.

INTRODUCTION

Readmissions cause a high burden to healthcare systems and patients. In the US nearly 20% of Medicare patients are readmitted within 30 days after hospital discharge, associated with an estimated annual cost of 17billion.[1] Readmissions are thought to be related to quality of care, for instance due to postoperative complications. As readmissions vary widely across countries, regions and centers, at least part of them might be avoidable.[2-6] As a consequence, there is a high interest in the readmission rate as an indicator of quality of hospital care. Nevertheless, the actual way this indicator is used in different countries varies widely.

In the US, since 2009 all-cause hospital readmission rates for pneumonia, congestive heart failure, and acute myocardial infarction are publically reported by the Centers for Medicare and Medicaid Services (CMS).[7] In 2010, the Patient Protection and Affordable Care Act (ACA) introduced the Hospital Readmissions Reduction Program (HRRP), for cost controlling. The program included financial penalties for hospitals having high readmission rates, which will be extended in the coming years.[8] In the UK, readmission rates for specific diseases have been published since 1998 by the National Centre for Health Outcomes Development (NCHOD) to improve quality.[9] It was found that the crude emergency readmission rate had increased from about 8% in 1998 to about 10% in 2006.[9] In response, the NHS started a new regulation for reimbursement payments in 2011: hospitals receive no reimbursement for emergency readmissions within 30 days of discharge following an elective admission. All other emergency readmissions are reimbursed for only 25%.[10] Since the year 2006 also the Australian government monitors 28-day readmission rates to gain more insight in quality of care.[11]

Readmissions are used for different aims, such as cost control or as balancing measure for length of hospital stay or other outcome measures. However, in recent years the focus has primarily been on using it as an easily available measure of the quality of hospital care. Despite its use by policymakers for both quality improvement and cost control, the validity of readmission rates as a measure of quality of hospital care is not evident.[12]

However, in order to consider a quality indicator to evaluate care for external purposes it needs to fulfill certain criteria in regards to its reliability and validity. An indicator needs to show relevance, based on its impact on health, its importance for policy and its susceptibility to being influenced by the health care system. The assessment of an indicator needs to be feasible.

The data needed to calculate an indicator need to be available, reliable and need to be seen in relation to the burden of reporting. Further, an indicator needs to show scientific soundness.[13] In the case of the readmission rate, this suggests, that readmissions are determined by quality of hospital care, measured by structures and processes. This implies that we are interested in avoidable readmissions.

We aim to give an overview of the different methodological aspects in the definition and measurement of readmission rates that need to be considered when interpreting readmission rates as a reflection of quality of hospital care for external purposes.

METHODS

A systematic computerized literature search was applied in the bibliographic databases Embase, Medline OvidSP, Web-of- Science, Cochrane central and PubMed for the period of 1st January 2001 to 27th May 2013.

With the search terms we aimed to cover quality indicators, quality measurement and readmission. This resulted in the following search strategy, which was adapted for the different bibliographic databases: ('clinical indicator'/de OR 'performance measurement system'/exp OR 'quality control procedures'/de OR 'quality control'/de OR 'medical audit'/de OR (((qualit* OR perform* OR safet* OR governance) NEAR/3 (indicat* OR measure* OR assessment* OR control* OR marker* OR metric*)) OR ((clinical OR medical) NEAR/3 (indicator* OR audit*))):ab,ti) AND ('hospital readmission'/de OR (readmiss* OR rehospital* OR ((re OR return) NEAR/3 (hospital* OR admiss*))): ab,ti).

Studies were included when they were written in English, focused on methodological aspects of readmission rates as a quality indicator for hospital care and full texts were available. We included only studies in major disease fields. Hence, studies focusing on rare diseases, just describing readmission rates over time or using readmissions as outcome measures of interventions were excluded.

Of the references identified in the literature search, titles and abstracts were screened and articles that did not meet the inclusion criteria were excluded. The full text of the remaining potentially eligible articles was reviewed to assess whether they should be included. In case of doubt, the article was discussed among the authors and if necessary, an independent researcher was consulted.

We discuss the methodological aspects that emerged from the literature review that are important for the validity of the readmission rates as an indicator of quality of care.

RESULTS

Our search strategy resulted in 1609 unique references of which titles and abstracts were screened. Based on title and abstract 1189 studies were excluded. Of the remaining 420 articles another 319 were excluded based on full text review (Figure 1). We provide a detailed description of the included studies in the appendix, and below we discuss the most important findings.

The context in which readmission rates are used

Prior to using the readmission rate as a measure of quality of care, the context in which the indicator will be used needs to be clearly defined. The rationale for using readmission rates is one aspect of this context. The readmission rate can be used with the primary aim to improve quality of care or rather for reasons of cost control. Next, specification of the clinical processes of which quality of care is assessed is important. Currently, readmission rates are mostly intended to measure quality of care in hospitals. Which implies that the risk of being readmitted is determined by the quality of care delivered during the hospital stay. Yet, literature shows that the conditions after patients' discharge, like the presence of a social network after discharge[14] as well as patients' capacity for managing their own care, influence the likelihood of being readmitted.[15,16] As a result, hospitals pay attention to improving transitional care,[17-24] for instance by patient education to prepare the patient for discharge and to coordinate outpatient follow up.[23] Although such a transition phase may help, the actual post-discharge phase is not really in a hospital's reach anymore. Another example are readmissions in chronic diseases, such as heart failure. These patients are readmitted often because of their comorbidities or because their condition becomes too severe to be treated by the general practitioner, irrespective from the quality of delivered care during their hospital stay.[1] Hence, the quality of care processes captured by readmission rates will often be broader than only in-hospital care.[25]

In summary, using readmission rates as a quality measure requires a clear definition of the context, including the rationale of measuring readmissions, the related care processes and the patient groups.

Methodological aspects

Based on the literature we defined several methodological aspects that need to be considered when using the readmission rates as a quality indicator (Table 1). These range from fundamental issues like the definition and the effect of competing outcomes, to more practical issues as the possibility

to adjust for case-mix and the data quality. These issues and their effects will be described in the next paragraphs. In the final paragraph we will focus on studies that have specifically tested the validity of readmission rates as a quality indicator.

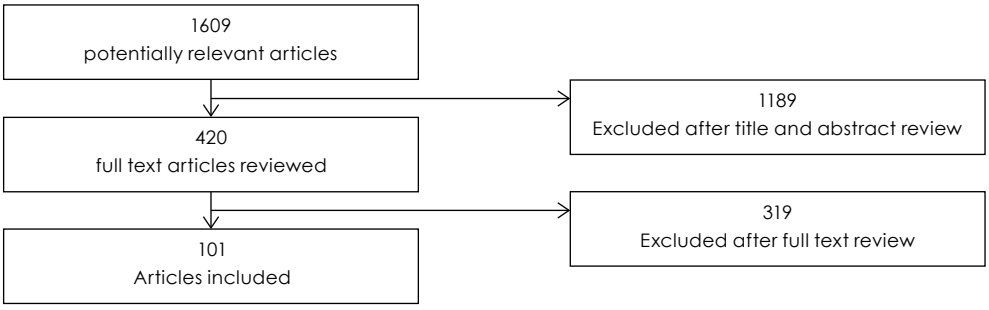
Indicator definition

Type of readmission

The definition of readmissions determines the number of readmissions that will be counted (numerator). Planned procedures, such as staged operations, are readmissions that are not determined by quality of care and therefore should not be included in the numerator of the quality indicator.[26,27] However, this basic distinction is not always made.[28] Hence, capture quality of care related readmissions requires a more specific definition (such as disease specific or emergency readmissions) rather than all-cause readmissions.[29]

A frequently suggested alternative is to count unplanned readmission rates. However, not all unplanned readmissions are a result of poor quality of care as certain complications cannot be avoided. Research has shown that just about 25% of all readmissions are avoidable/preventable. Therefore, ideally, the addition on whether a readmission was avoidable/preventable. Although high variation in overall readmission rate can be observed, this is not the case for the rate of preventable readmissions.[2,29] Therefore, ideally it is defined, whether a readmission was avoidable/preventable (through proper care delivery)[28,30] but the judgment on the preventability of a readmission remains subjective.[2]

Figure 1. Flow chart of inclusion process



Time window

The time window after the index admission in which admissions are regarded as readmissions is not consistently defined in the literature. The indicator is generally calculated on basis of readmissions within one month (28 days UK, 31 days USA) regardless of the patient group and condition.[28,31-33] When choosing a time window, it needs to be considered that a too short time window might miss related readmissions while a large one increases the likelihood of included admissions unrelated to the index admission. For example, in cancer surgery a longer time frame would allow to provide a better overview of actual costs, but it would also include readmissions due to disease progression instead of poor quality of surgery.[25] Clearly, the type of disease the patient was originally treated for is largely influencing the optimal timeframe.[32] Therefore the timeframe for readmissions should be defined per disease.

The effect of competing outcomes

Association with (in-hospital) mortality

Mortality can be seen as a competing endpoint for readmissions: patients who die will not be readmitted.[34,35] Therefore patients who died during their hospital stay need to be excluded from the denominator of the readmission rate. Further, hospitals with high 30-day in-hospital mortality rates are not necessarily outliers on the readmission rate as well.[36] Research showed that the link between high readmission rates and mortality rates on hospital level is limited. A 'modest' inverse relationship was merely found for heart failure patients, and no relation could be observed for pneumonia and acute myocardial infarction, suggesting that the two indicators measure different aspects of quality of care, which are not strongly related.[37] Therefore different outcome measures, such as the readmission rate and the mortality rate should be brought in relation with each other to gain insight in total hospital performance.[36,37]

Association with length of in hospital stay

Length of stay is generally decreasing, partly because of efficiency gaining interventions, such as a 'just-in-time bed availability system' to increase the bed turnover ratio.[38,39] Research suggests a link between length of stay and the risk of being readmitted.[39-45] For each day shorter in hospital, a 6% increase in likelihood of readmission was found.[40] Other studies fail to confirm this link,[24,42,46-50] which might be due to inappropriate adjustment for disease severity.[41,51]

Table 1. Overview of methodological aspects challenging the validity of readmission rates for benchmarking

	Methodological aspect	Problem	Potential solution
Clinical setting	readmission rates are thought to reflect quality of hospital care[17-24]	care after discharge also influences readmission[14,25]	clear definition of the indicator, the patient group and the clinical setting(hospital care, integrated care) aimed to measure increase insight in influence of post discharge phase/social factors on readmissions[14] relate readmission to other outcome measures such as mortality, emergency department and observation service use[14] evaluate home health care/nursing home information [15]
Indicator definition			
Type of readmission	missing distinction between planned/unplanned procedures [2,26-29]	inclusion of readmissions unrelated to quality of care into the numerator[26] leads to overestimation of the rate of readmission [69]	specify definition of the indicator,[27,28,31] define disease-specific/emergency readmissions instead of overall readmissions[2] include indication on preventability/avoidability of readmission in definition[2,28-30]
Time window	no consistent definition of the time window in which admission is considered as readmission[28] generally 28-31 day time frame used regardless of patient group/condition[31-33]	although 30 days seem generally sufficient,[31,33] for certain conditions it is a too short time window, while for others it increases the likelihood of including admissions unrelated to index admission[25,32]	evaluate time frame based on condition under evaluation
Effect of competing outcomes			
Association with (in-hospital) mortality	a group of patients who receive poor quality of care are not readmitted, because they die or recover nevertheless[31,34,35]	not excluding patient who died from the denominator leads to a potential underestimation of rate of qoc related readmission	exclude patients who died during hospital stay from denominator link hospital data with death statistics, exclude patients from denominator who die outside hospital relate the readmissions with mortality rate in order to understand total hospital performance[36,37]
Association with length of in-hospital mortality	a decreased length of hospital stay increases readmissions[38,39]	the exact mechanism with readmission is inconclusive[24,39,40,42-50]	further research to understand the mechanism between length of stay and readmission

qoc – quality of care

Case-mix correction using administrative data vs. clinical data	no consensus on which patient characteristics affect readmission likelihood[27,52] two high risk groups defined: the sickest and the poorest[2,51,53,54]	these factors are not standard variables in risk prediction models as often not available in administrative databases[36] current risk prediction models perform moderately[39,40,55-57]	apply proper case-mix correction for patient characteristics including socio-economic status and disease severity[39,51] further research on risk prediction models including linkage of primary care data and data information
Data reliability			
Missing readmissions to other institutions	patients are readmitted to institutions other than index hospital[25,35]	patients cannot always be followed between centers; only readmissions to same institutions are measured assessing 'same hospital' readmissions, might be underestimation of the real number of readmissions[25]	further research on the proportion of patients readmitted to other hospitals than index hospital unique patient information to follow patients between centers
Coding	coding practice influences the validity readmission rates[30,58,77-84] no conclusion on how to register readmissions potentially related to qoc in reliable way[2,59-61]	missing distinction between planned/unplanned procedures leads to overestimation of real readmission variation in coding leads to biased comparison between hospitals	increase investment in performance measurement systems[16] research on data reliability[28] standardized data registry (electronic data systems)[16,62] engagement of the provider in measurement, analysis and interpretation of the indicator[16,64]
Completeness and accuracy of data source	reliable data collection systems are lacking[38] Readmissions are mainly calculated based on administrative data[16,63] administrative data suffer from inaccuracy, like non-exact/ incomplete registration of variables not relevant for financial concerns[39,40,64-69]	incomplete registration may lead to over/underestimation of real readmission inaccurate indication of readmissions related to qoc may lead to overestimation of readmissions[64,65]	aim for minimum data set with complete registration registration of unique patient identifying information to enhance possibility for linking data(such as pharmacy data)[70] enhancing linkage opportunities increases possibility for better case-mix correction
Validity of readmission rates as a quality measure	no gold standard on how to assess qoc in the literature huge variation in conclusions in regard to the validity of the readmission indicator[71-113]	potentially invalid conclusions on qoc	above described methodological conditions need to be taken into account when further investigating readmissions as a quality indicator additional data gathering for further investigation of outlier hospitals[93]

Case-mix correction

The likelihood that a patient is readmitted is not only affected by quality of care but also by characteristics of the patient. Between-hospital differences in readmission rates may be caused by differences in patient population and therefore readmission rates need to be adjusted for patient characteristics. Although many case-mix correction models for readmissions have been developed, there is little consensus on which patient characteristics affect the likelihood of a readmission.[27,52] Numerous studies, varying in their methodology, geographical characteristics, patient groups and considered variables, find different factors that increase the risk of re-admission. In general, two patient groups seem to be at a high risk of being readmitted: the sickest and poorest patients.[2,20,51,53,54] However, these factors are often not included as standard variables in case-mix correction models, as these models are often based on administrative data and therefore miss detailed clinical information.

In a review that evaluated 30 validated readmission risk prediction models, the authors concluded that most models had poor predictive ability. Almost all studies had c-statistics less than 0.70,[55] possibly due to missing demographic or clinical variables. In a more recent paper, the prediction model reached a higher predictive ability (c-statistic = 0.80).[41] The authors concluded however that information on demographics, SES, prior utilization and diagnosis still had restricted predictive power.[41] Thus, current research provides limited guidance on which variables should be included in models to adjust for case-mix.[41,55-57]

Data reliability

Missing readmissions to other institutions

Not all patients are readmitted to the same center where they had their index admission. This is mainly due to the centralization of complex operations in tertiary centers, such as in oncology.[25] When patients unexpectedly develop complications and are readmitted in their local center, they are not captured when only readmissions to the ''same hospital'' are counted.[25] Missing these patients leads to an underestimation of the true overall readmission rate.

Coding

The coding practice within a hospital has an essential impact on the validity of readmission rate as a quality indicator.[58] The way a ''planned'' procedure is defined is crucial for the comparability between hospitals. Ideally a planned readmission is coded in the registration system, for example,

with an additional coding element ''staged'' at the index admission, which would indicate that a follow-up procedure is planned.[59]

Urgent readmissions are sometimes considered as a potential proxy for the relatively subjective 'avoidable readmissions', as these are coded, for example an admission through the ER. Although low urgent readmission rates showed to be related to low avoidable readmission rates[60] it was shown that the ''avoidability'' of urgent readmissions also significantly varied by the time from discharge, with early readmissions being more likely to be avoidable.[2,61]

Other causes for biased comparisons between hospitals are the different and unspecific definitions of the type of readmissions assessed, and variation in coding between hospitals. It is essential who is in charge of the coding process. For example administrative staff at the department or hospital level, the treating clinician, or specialized data coders. The variation in coding practice may affect both the readmission rates and the case-mix variables.

Completeness and accuracy of data source

Electronic health records and health information exchange networks result in more accurate and complete clinical data.[62] The major information source to calculate the readmission rate is administrative data. The advantage of administrative data is that this data is standard available and patient journeys can be followed (within hospitals).[63] Nevertheless, one major limitation of administrative data is the data inaccuracy,[64] which includes the non-exact or incomplete registration of variables that are not relevant for financial concerns.[38,40,41] Research showed that to a certain degree administrative data captures similar information compared to medical records, for example on all-cause readmissions.[65-68] However more specific information, like the identification of unplanned readmissions or index procedure related readmissions, showed to be more difficult to extract.[66,69] An accurate indication of whether a readmission is a part of treatment or due to a cancelled procedure and not a readmission related to a quality of care problem, would enhance the reliability of the data source.[64,65]

The case-mix correction variables that have been investigated so far are most often present in administrative databases. However, clinical information such as disease severity is often lacking limiting case-mix correction possibilities. The addition of a unique patient identifier across different databases would enhance the possibility for linking data, such as pharmacy

data[70] or clinical data. This would largely improve the possibilities for more precise definitions of readmissions and better case-mix correction.

Validity of readmission rates as a quality measure

No gold standard exists on how to assess quality of care. Usually different hospital structures and processes and their relation with patient outcomes are measured. The different definitions and proxies used in studies to quantify quality of hospital care influence whether an association between the readmission rate and 'quality' is found. For example, we found studies that relate readmissions to hospital volume, but neither can be regarded as a 'gold standard' of hospital quality.

Furthermore, the methodological aspects we discussed have a potential influence on the validity of the readmission rates as a quality indicator. These may contribute to the huge variation in conclusions with regard to the validity of readmission rates found in the literature. Different studies in different patient groups and conditions come to the conclusion that lower quality of hospital care is linked to a higher number of readmissions.[71-94] Especially safety-related events (such as postoperative complications) show a relation with readmissions.[71,95] Rosen and colleagues, who evaluated the correlation between patient safety indicators and readmissions, showed that patients who experienced a patient safety event had an increased risk of readmission.[71] Nevertheless, there are also studies that are inconclusive,[96-101] show an inverse relationship[102,103] or no relationship at all between readmission rate and in-hospital quality of care.[98,104-113] Analysis of additionally collected data could help to gain insight into outlier hospitals in order to understand driving mechanisms behind high readmission rates.[93]

DISCUSSION

This review aimed to summarize the methodological aspects that need to be considered when using the readmission rate as a measure for quality of hospital care for external purposes. We found that the validity of readmission rates as a quality indicator is influenced by the clinical process that is assessed, the indicator definition, the extend of case-mix correction, the effect of competing outcomes and the data reliability. Ignoring or poorly handling these aspects may lead to a biased estimation of the overall readmission rate and a biased comparison of readmission rates between hospitals. As a result of variance in handling these methodological threats, studies on the validity of readmission rates as a quality indicator reach conflicting conclusions. We conclude that given the limitations of readmission rates, they need to be used with caution as a measure of in-hospital quality, even more when used as a tool for a pay for performance scheme.

Some of the discussed factors concerning the readmission rate could in principal be improved by investing resources in accurate data registry and refinements of indicator description. For instance, by using unique patient identifiers to follow patients across centers. That would help to avoid missing readmissions to other institutions. Another option would be to flag planned admissions, which are a part of the treatment plan or due to cancelled procedures, to measure just the quality of care related readmissions.

Other problems, such as the competing endpoint 'mortality' are more complex. Patients who died in hospital need to be excluded from the patient group forming the denominator to calculate the readmission rate, as they are not at risk any more to be readmitted. These deaths are captured in the mortality rate. Therefore it is essential to combine outcome in order to provide a more complete picture of the quality of hospital care.

Nevertheless there are theoretical considerations whether a readmission is an indication of bad quality of care. First, a readmission is obviously a more positive outcome than dying. Secondly, if there is for example a chance of six percent that a complication occurs after discharge, it would mean that 100 patients need to be admitted longer, to avoid a complication in six patients.[114] It can be questioned whether by a longer length of hospital stay a complication really can be avoided or only detected at an earlier stage. It is also possible to inform the patient on the risk of developing a complication and decide together how to continue. Furthermore, it needs to be taken into account that readmissions are not always solely determined by quality of hospital care. For

certain diseases, like heart failure, the patient's condition is the major driver behind repeated admissions. Patients with low socio-economic status, elderly and patients with co-morbidities are at high risk of getting readmitted. Therefore case-mix correction is essential. Furthermore, the role of facilities outside the hospital and after the 30-day time window, like community services, need to be involved in the conceptual framework of readmissions. When aiming to improve quality of care (in and outpatient) increased integration and cooperation between primary and secondary care is needed.

The literature study revealed inconclusive results for some methodological aspects, such as the relation with length of stay, or patient characteristics. The studies we assessed investigated different patient populations and often were based on hospital administrative data. A recent high quality study which was not included in our review investigated surgical readmissions of 479 471 patients from 3004 hospitals. The authors found that higher surgical volume was significantly related with lower composite readmission rates (upper volume quartile 12.7% vs. lower volume quartile 16.8% P,0.0001), and hospitals with the lowest surgical mortality rates had significantly lower readmission rates (lower mortality quartile 13.3% vs. upper mortality quartile 14.2% P,0.0001). But high adherence to surgical process measures was only marginally linked with lower readmission rates (highest quartile vs. lowest quartile, 13.1% vs. 13.6%; P = 0.02), showing that it is still unclear whether low readmission rates are the result of good quality.[115]

Furthermore, the risk of getting readmitted is also varying between patient groups and conditions. This supports the idea that outcome measures, like the readmission rate, are not a one size fits all measure. Even if quality of hospital care and the transition phase can potentially be improved, readmissions might be a more applicable measure for certain diseases than for others. For chronic diseases, where planned admissions are part of treatment strategies, readmissions are a less suitable performance measure. At least not until generally used data systems can identify planned admissions with high certainty. It requires clinical knowledge to determine whether (avoidable) readmissions may theoretically represent poor quality of care for specific diseases. Consequently more research is needed to build reliable algorithms to identify avoidable readmissions.

In sum, avoidable readmissions are of high relevance, as they are an adverse event to patients and family and are a high financial burden for healthcare systems. The assessment of the indicator shows difficulties, as the indicator definition is often not explicit enough to identify readmissions related to qual-

ity of care (avoidable readmissions). The data used to calculate the indicator is mainly administrative data, which generally includes incomplete and inaccurate data elements and lacks clinical information. Furthermore, in many countries readmissions to other institutions cannot be followed. Readmission rates are influenced also by other factors than quality of hospital care, which include length of stay, (in-hospital) mortality and patient characteristics. The magnitude of influence is partly not know as data is missing to investigate the association (e.g. no post discharge mortality, no clinical characteristics). Further, the scientific evidence of the indicator is limited, as existing research shows conflicting results with regard to the influence of quality of hospital care on the readmission rate (see Appendix 1). This, however, could be related to the prior mentioned methodological aspects that are variously.

Using outcome measures externally to measure and compare hospital performance has consequences. When financial consequences are linked to the outcome, unintended effects could occur. For example, hospitals may try to reduce their readmission to escape the penalty of exceeding the readmission rate by lowering admissions, moving readmissions after the 30-day window, or risk-avoidance in regards to high risk groups. These gaming efforts might reduce the focus on the actual intention: improving quality of hospital care.

A measure used for external purposes should be underpinned with solid evidence for its validity. However, the link between readmissions and the quality of hospital care seems not to be fully explained yet. Still, this does not imply that there is no room for improvement for hospitals in their readmission rate and the indicator could not be useful for internal use. Research should continue to gain insight in the driving mechanisms behind readmissions for the different conditions to improve our understanding how the readmission rate is a part of the quality of hospital care picture. In addition, the readmission rate needs to be brought into relation with other outcome indicators, and hence considered as part of a bundle, to understand all aspects of hospital performance.[36]

The methodological aspects we identified need to be considered when using readmission rates as quality indicator. The use of readmission rates for external quality purposes, such as for pay for performance requires strict methodological criteria to avoid confounding. At its current state the rate of readmission does not fulfill the methodological requirements of a reliable and valid indicator. Therefore the indicator should not be used for external purposes. As this is nevertheless currently happening, readmission rates should be interpreted with great caution.

REFERENCES

1. Jencks SF, Williams MV, Coleman EA (2009) Rehospitalizations among patients in the Medicare fee-for-service program. N Engl J Med.360(14): 1418-28.
2. Van Walraven C, Bennett C, Jennings A, Austin PC, Forster AJ (2011) Proportion of hospital readmissions deemed avoidable: A systematic review. CMAJ. 183(7):E391-E402.
3. Krumholz HM, Merrill AR, Schone EM, Schreiner GC, Chen J, et al. (2009) Patterns of Hospital Performance in Acute Myocardial Infarction and Heart Failure 30-Day Mortality and Readmission. Circulation-Cardiovascular Quality and Outcomes. 2(5):407-13.
4. Jensen PH, Webster E, Witt J (2009) Hospital type and patient outcomes: an empirical examination using AMI readmission and mortality records. Health Econ. 18(12):1440-60.
5. Bernheim SM, Grady JN, Lin ZQ, Wang Y, Wang YF, et al. (2010) National Patterns of Risk-Standardized Mortality and Readmission for Acute Myocardial Infarction and Heart Failure Update on Publicly Reported Outcomes Measures Based on the 2010 Release. Circulation-Cardiovascular Quality and Outcomes. 3(5):459-67.
6. Schiotz M, Price M, Frolich A, Sogaard J, Kristensen JK, et al. (2011) Something is amiss in Denmark: a comparison of preventable hospitalisations and readmissions for chronic medical conditions in the Danish Healthcare system and Kaiser Permanente. BMC Health Serv Res. 11:347.
7. Axon RN, Williams MV (2011) Hospital readmission as an accountability measure. JAMA. 2;305(5):504-5.
8. Desai AS, Stevenson LW (2012) Rehospitalization for heart failure: predict or prevent? Circulation. 24;126(4):501-6.
9. Department of Health. Emergency readmission rates (2008) London: Department of Health.
10. Department of Health. The operating framework. 2010. Available: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/151906/dh_122736.pdf.pdf. Accessed (2013 Jul 8..
11. Australian Government. Australian Institute of Health and Welfare. National Healthcare Agreement: PI 23-Unplanned hospital readmission rates, 2013 QS. Available: http://meteor.aihw.gov.au/content/index.phtml/itemId/507456. Accessed 2013 Jul 11.
12. Gu Q, Koenig L, Faerberg J, Steinberg CR, Vaz C, et al. (2014) The medicare hospital readmissions reduction program: potential unintended consequences for hospitals serving vulnerable populations. Health Serv Res.49(3):818-37.
13. Mattke S, Kelley E, Scherer P, et al. (2006) "Health Care Quality Indicators Project Initial Indicators Report", OECD Health Working Papers. Paris.
14. Kangovi S, Grande D (2011) Hospital readmissions - Not just a measure of quality. J Am Med Assoc. 306(16):1796-7.
15. Spector WD, Mutter R, Owens P, Limcangco R (2012) Thirty-day, all-cause readmissions for elderly patients who have an injury-related inpatient stay. Med Care. 50(10):863-9.
16. Forster AJ, van Walraven C (2012) The use of quality indicators to promote accountability in health care: the good, the bad, and the ugly. Open Med. 6(2):e75-9.
17. Shepperd S, Lannin NA, Clemson LM, McCluskey A, Cameron ID (2013) Discharge planning from hospital to home. Cochrane Database Syst Rev. 1:CD000313.
18. Lambrinou E, Kalogirou F, Lamnisos D, Sourtzi P (2012) Effectiveness of heart failure management programmes with nurse-led discharge planning in reducing re-admissions: A systematic review and meta-analysis. Int J Nurs Stud. 49(5):610-24.
19. Wong FKY, Ho MM, Yeung S, Tam SK, Chow SK (2011) Effects of a health-social partnership transitional program on hospital readmission: A randomized controlled trial. Soc Sci Med. 73(7):960-9.
20. Demir E, Chaussalet T, Adeyemi S, Toffa S (2012) Profiling hospitals based on emergency readmission: A multilevel transition modelling approach. Comput Methods Programs Biomed. 108(2):487-99.
21. Jha AK, Orav EJ, Epstein AM (2009) Public reporting of discharge planning and rates of readmissions. New Engl J Med. 361(27):2637-45.
22. Benbassat J, Taragin MI (2013) The effect of clinical interventions on hospital readmissions: a meta-review of published meta-analyses. Isr J Health Policy Res. 2(1):1.
23. Burke RE, Kripalani S, Vasilevskis EE, Schnipper JL (2013) Moving beyond readmission penalties: Creating an ideal process to improve transitional care. J Hosp Med. 8(2):102-9.

24. Chan FW, Wong FY, Yam CH, Cheung WL, Wong EL, et al. (2011) Risk factors of hospitalization and re-admission of patients with COPD in Hong Kong population: analysis of hospital admission records. BMC Health Serv Res. 11:186.
25. Rochefort MM, Tomlinson JS (2012) Unexpected Readmissions After Major Cancer Surgery. An Evaluation of Readmissions as a Quality-of-Care Indicator. Surg Oncol Clin North Am. 21(3):397-405.
26. McCormack R, Michels R, Ramos N, Hutzler L, Slover JD, et al. (2013) Thirty-day readmission rates as a measure of quality: causes of readmission after orthopedic surgeries and accuracy of administrative data. J Healthc Manag. 58(1):64-76; discussion -7.
27. Brooke BS, De Martino RR, Girotti M, Dimick JB, Goodney PP (2012) Developing strategies for predicting and preventing readmissions in vascular surgery. J Vasc Surg. 56(2):556-62.
28. Fischer C, Anema HA, Klazinga NS (2012) The validity of indicators for assessing quality of care: a review of the European literature on hospital readmission rate. Eur J Public Health. 22(4):484-91.
29. Van Walraven C, Jennings A, Forster AJ (2012) A meta-analysis of hospital 30-day avoidable readmission rates. J Eval Clin Pract. 18(6):1211-8.
30. Courtney EDJ, Ankrett S, McCollum PT (2003) 28-Day emergency surgical re-admission rates as a clinical indicator of performance. Ann R Coll Surg Engl. 85(2):75-8.
31. Rumball-Smith J, Hider P (2009) The validity of readmission rate as a marker of the quality of hospital care, and a recommendation for its definition. New Zealand Med J. 122(1289).
32. Demir E, Chaussalet TJ, Xie H, Millard PH (2008) Emergency readmission criterion: a technique for determining the emergency readmission time window. IEEE Trans Inf Technol Biomed. 12(5):644-9.
33. Maurer PP, Ballmer PE (2004) Hospital readmissions - are they predictable and avoidable? Swiss Medical Weekly. 134(41-42):606-11.
34. Gheorghiade M, Vaduganathan M, Fonarow GC, Bonow RO (2013) Rehospitalization for heart failure: Problems and perspectives. J Am Coll Cardiol. 61(4):391-403.
35. Grocott MPW (2010) Monitoring surgical outcomes: How and why? Curr Anaesth Crit Care. 21(3):129-36.
36. Almoudaris AM, Burns EM, Bottle A, Aylin P, Darzi A, et al. (2013) Single measures of performance do not reflect overall institutional quality in colorectal cancer surgery. Gut. 62(3):423-9.
37. Krumholz HM, Lin Z, Keenan PS, Chen J, Ross JS, et al. (2013) Relationship between hospital readmission and mortality rates for patients hospitalized with acute myocardial infarction, heart failure, or pneumonia. JAMA. 309(6):587-93.
38. Basu A, Howell R, Gopinath D (2010) Clinical performance indicators: Intolerance for variety? Int J Health Care Qual Assur. 23(4):436-49.
39. Heggestad T (2002) Do hospital length of stay and staffing ratio affect elderly patients' risk of readmission? - A nation-wide study of Norwegian hospitals. Health Services Research. 37(3):647-65.
40. Kaboli PJ, Go OT, Hockenberry J, Glasgow JM, Johnson SR, et al. (2012) Associations between reduced hospital length of stay and 30-day readmission rate and mortality: 14-year experience in 129 veterans affairs hospitals. Ann Intern Med. 157(12):837-45.
41. Shulan M, Gao K, Moore CD (2013) Predicting 30-day all-cause hospital readmissions. Health Care Manag Sci. 16(2):167-75.
42. Kramer AA, Higgins TL, Zimmerman JE (2013) The association between ICU readmission rate and patient outcomes. Crit Care Med. 41(1):24-33.
43. Ahmed J, Khan S, Lim M, Chandrasekaran TV, MacFie J (2012) Enhanced recovery after surgery protocols - compliance and variations in practice during routine colorectal surgery. Colorectal Disease. 14(9):1045-51.
44. Dobrzanska L, Newell R (2006) Readmissions: A primary care examination of reasons for readmission of older people and possible readmission risk factors. J Clin Nurs. 15(5):599-606.
45. Schneider EB, Hyder O, Brooke BS, Efron J, Cameron JL, et al. (2012) Patient readmission and mortality after colorectal surgery for colon cancer: Impact of length of stay relative to other clinical factors. J Am Coll Surg. 214(4):390-8.
46. Dunlay SM, Weston SA, Killian JM, Bell MR, Jaffe AS, et al. (2012)Thirty-day rehospitalizations after acute myocardial infarction: A cohort study. Ann Intern Med. 157(1):11-8.
47. Dallal RM, Trang A (2012) Analysis of perioperative outcomes, length of hospital stay, and readmission rate after gastric bypass. Surg Endosc Interv Tech. 26(3):754-8.

48. Lau ACW, Yam LYC, Poon E (2001) Hospital re-admission in patients with acute exacerbation of chronic obstructive pulmonary disease. Respiratory Medicine. 95(11):876-84.
49. Kiran RP, Delaney CP, Senagore AJ, Steel M, Garafalo T, et al. (2004) Outcomes and prediction of hospital readmission after intestinal surgery. J Am Coll Surg. 198(6):877-83.
50. Kociol RD, Liang L, Hernandez AF, Curtis LH, Heidenreich PA, et al. (2013) Are we targeting the right metric for heart failure? Comparison of hospital 30-day readmission rates and total episode of care inpatient days. Am Heart J.
51. Johnson T, Bardhan J, Odwazny R, Harting B, Skarupski K, et al. (2012) Hospital care may not affect the risk of readmission. Qual Manag Health Care. 21(2):68-73.
52. Van Walraven C, Wong J, Hawken S, Forster AJ (2012) Comparing methods to calculate hospital-specific rates of early death or urgent readmission. CMAJ. 184(15):E810-E7.
53. Khawaja FJ, Shah ND, Lennon RJ, Slusser JP, Alkatib AA, et al. (2012) Factors associated with 30-day readmission rates after percutaneous coronary intervention. Arch Intern Med. 172(2):112-7.
54. Goldfield NI, McCullough EC, Hughes JS, Tang AM, Eastman B, et al. (2008) Identifying potentially preventable readmissions. Health Care Financ Rev. 30(1):75-91.
55. Kansagara D, Englander H, Salanitro A, Kagen D, Theobald C, Kripalani S (2011) Readmission risk modeling: A systematic review. J Gen Intern Med. 26:S125.
56. Correll PK, Xuan W, Williamson M, Sundararajan V, Ringland C, et al. (2007) Reattendance at hospital for asthma in two Australian states, 2000-2003. Respirology. 12(2):220-6.
57. Giamouzis G, Kalogeropoulos A, Georgiopoulou V, Laskar S, Smith AL, et al. (2011) Hospitalization epidemic in patients with heart failure: Risk factors, risk prediction, knowledge gaps, and future directions. J Card Fail. 17(1):54-75.
58. Cram P, Ibrahim SA, Lu X, Wolf BR (2012) Impact of alternative coding schemes on incidence rates of key complications after total hip arthroplasty: a risk-adjusted analysis of a national data set. Geriatr orthop surg rehabil. 3(1):17-26.
59. Hannan EL, Zhong Y, Lahey SJ, Culliford AT, Gold JP, et al. (2011) 30-Day readmissions after coronary artery bypass graft surgery in New York State. JACC Cardiovasc Interventions. 4(5):569-76.
60. van Walraven C, Austin PC, Forster AJ (2012) Urgent readmission rates can be used to infer differences in avoidable readmission rates between hospitals. Journal of Clinical Epidemiology. 65(10):1124-30.
61. Bianco A, Mole A, Nobile CGA, Di Giuseppe G, Pileggi C, et al. (2012) Hospital Readmission Prevalence and Analysis of Those Potentially Avoidable in Southern Italy. Plos One. 7(11).
62. Ben-Assuli O, Shabtai I, Leshno M (2013) The impact of EHR and HIE on reducing avoidable admissions: controlling main differential diagnoses. BMC Med Inf Decis Mak. 13:49.
63. Holt PJE, Poloniecki JD, Hofman D, Hinchliffe RJ, Loftus IM, et al. (2010) Re-interventions, Readmissions and Discharge Destination: Modern Metrics for the Assessment of the Quality of Care. Eur J Vasc Endovasc Surg. 39(1):49-54.
64. Amin BY, Schairer W, Tu TH, Ames CP, Berven S, et al. (2012) Improving benchmarking in spine surgery: Suggested modifications of the UHC algorithm for calculating readmission rates following spine surgery. Neurosurgery. 71(2):E575-E6.
65. Lindenauer PK, Normand SLT, Drye EE, Lin Z, Goodrich K, et al. (2011) Development, validation, and results of a measure of 30-day readmission following hospitalization for pneumonia. J Hosp Med. 6(3):142-50.
66. Sellers MM, Merkow RP, Halverson A, Hinami K, Kelz RR, et al. (2013) Validation of New Readmission Data in the American College of Surgeons National Surgical Quality Improvement Program. Journal of the American College of Surgeons. 216(3):420-7.
67. Krumholz HM, Wang Y, Mattera JA, Wang YF, Han LF, et al. (2006) An administrative claims model suitable for profiling hospital performance based on 30-day mortality rates among patients with heart failure. Circulation. 113(13):1693-701.
68. Wallmann R, Llorca J, Gomez-Acebo I, Ortega AC, Roldan FR, et al. (2013) Prediction of 30-day cardiac-related-emergency-readmissions using simple administrative hospital data. Int J Cardiol. 164(2):193-200.
69. Adeyemo D, Radley S (2007) Unplanned general surgical re-admissions - How many, which patients and why? Ann R Coll Surg Engl. 89(4):363-7.
70. Parker JP, McCombs JS, Graddy EA (2003) Can pharmacy data improve prediction of hospital outcomes? Comparisons with a diagnosis-based comorbidity measure. Med Care. 41(3):407-19.

71. Rosen AK, Loveland S, Shin M, Shwartz M, Hanchate A, et al. (2013) Examining the Impact of the AHRQ Patient Safety Indicators (PSIs) on the Veterans Health Administration: The Case of Readmissions. Med Care. 51(1):37-44.
72. Mokhtar SA, El.Mahalli AA, Al-Mulla S, Al-Hussaini R (2012) Study of the relation between quality of in-patient care and early readmission for diabetic patients at a hospital in the eastern province of Saudi Arabia. East Mediterr Health J. 18(5):474-9.
73. Kergoat MJ, Latour J, Lebel P, Leclerc BS, Leduc N, et al. (2012) Quality-of-care processes in geriatric assessment units: principles, practice, and outcomes. J Am Med Dir Assoc. 13(5):459-63.
74. Auerbach AD, Hilton JF, Maselli J, Pekow PS, Rothberg MB (2009) Shop for quality or volume? Volume, quality, and outcomes of coronary artery bypass surgery. Ann Intern Med. 150(10):696-704+W-123.
75. Stukel TA, Alter DA, Schull MJ, Ko DT, Li P (2010) Association between hospital cardiac management and outcomes for acute myocardial infarction patients. Med Care. 48(2):157-65.
76. Chen JY, Ma Q, Chen H, Yermilov I (2012) New bundled world: quality of care and readmission in diabetes patients. J Diabetes Sci Technol. 6(3):563-71.
77. Weber RS, Lewis CM, Eastman SD, Hanna EY, Akiwumi O, et al. (2010) Quality and performance indicators in an Academic Department of Head and Neck Surgery. Arch Otolaryngol Head Neck Surg. 136(12):1212-8.
78. Kent TS, Sachs TE, Callery MP, Vollmer Jr CM (2011) Readmission after major pancreatic resection: A necessary evil? J Am Coll Surg. 213(4):515-23.
79. Barbieri CE, Lee B, Cookson MS, Bingham J, Clark PE, et al. (2007) Association of procedure volume with radical cystectomy outcomes in a nationwide database. Journal of Urology. 178(4):1418-21.
80. Chung ES, Guo L, Casey Jr DE, Bartone C, Menon S, et al. (2008) Relationship of a quality measure composite to clinical outcomes for patients with heart failure. Am J Med Qual. 23(3):168-75.
81. Judge A, Chard J, Learmonth I, Dieppe P (2006) The effects of surgical volumes and training centre status on outcomes following total joint replacement: analysis of the Hospital Episode Statistics for England. J Public Health (Oxf). 28(2):116-24.
82. Boulding W, Glickman SW, Manary MP, Schulman KA, Staelin R (2011) Relationship between patient satisfaction with inpatient care and hospital readmission within 30 days. Am J Managed Care. 17(1):41-8.
83. Youn YJ, Yoo BS, Lee JW, Kim JY, Han SW, et al. (2012) Treatment performance measures affect clinical outcomes in patients with acute systolic heart failure: Report from the Korean Heart Failure Registry. Circ J. 76(5):1151-8.
84. VanSuch M, Naessens JM, Stroebel RJ, Huddleston JM, Williams AR (2006) Effect of discharge instructions on readmission of hospitalised patients with heart failure: Do all of the Joint Commission on Accreditation of Healthcare Organizations heart failure core measures reflect better care? Qual Saf Health Care. 15(6):414-7.
85. Showalter JW, Rafferty CM, Swallow NA, DaSilva KO, Chuang CH (2011) Effect of Standardized Electronic Discharge Instructions on Post-Discharge Hospital Utilization. Journal of General Internal Medicine. 26(7):718-23.
86. Polanczyk CA, Newton C, Dec GW, Di Salvo TG (2001) Quality of care and hospital readmission in congestive heart failure: An explicit review process. J Card Fail. 4:289-98.
87. Luthi JC, Lund MJ, Sampietro-Colom L, Kleinbaum DG, Ballard DJ, et al. (2003) Readmissions and the quality of care in patients hospitalized with heart failure. International Journal for Quality in Health Care. 15(5):413-21.
88. Joynt KE, Orav EJ, Jha AK (2011) The association between hospital volume and processes, outcomes, and costs of care for congestive heart failure. Ann Intern Med. 154(2):94-102.
89. Horwitz LI, Wang YF, Desai MM, Curry LA, Bradley EH, et al. (2012) Correlations among risk-standardized mortality rates and among risk-standardized readmission rates within hospitals. Journal of Hospital Medicine. 7(9):690-6.
90. Hernandez AF, Hammill BG, Peterson ED, Yancy CW, Schulman KA, et al. (2010) Relationships between emerging measures of heart failure processes of care and clinical outcomes. Am Heart J. 159(3):406-13.
91. Bottle A, Aylin P (2009) Application of AHRQ patient safety indicators to English hospital data. Qual Saf Health Care. 18(4):303-8.
92. Halfon P, Eggli Y, van Melle G, Chevalier J, Wasserfallen JB, et al. (2002) Measuring potentially avoidable hospital readmissions. Journal of Clinical Epidemiology. 55(6):573-87.

93. Halfon P, Eggli Y, Pretre-Rohrbach I, Meylan D, Marazzi A, et al. (2006) Validation of the potentially avoidable hospital readmission rate as a routine indicator of the quality of hospital care. Med Care. 44(11):972-81.
94. Maeda JLK (2010) Evidence-Based Heart Failure Performance Measures and Clinical Outcomes: A Systematic Review. Journal of Cardiac Failure. 16(5):411-8.
95. Encinosa WE, Hellinger FJ (2008) The impact of medical errors on ninety-day costs and outcomes: An examination of surgical patients. Health Serv Res. 43(6):2067-85.
96. Shahian DM, Meyer GS, Mort E, Atamian S, Liu X, et al. (2012) Association of National Hospital Quality Measure adherence with long-term mortality and readmissions. BMJ Qual Saf. 21(4):325-36.
97. Fonarow GC, Abraham WT, Albert NM, Stough WG, Gheorghiade M, et al. (2007) Association between performance measures and clinical outcomes for patients hospitalized with heart failure. J Am Med Assoc. 297(1):61-70.
98. Stefan MS, Pekow PS, Nsa W, Priya A, Miller LE, et al. (2012) Hospital Performance Measures and 30-day Readmission Rates. J Gen Intern Med. 1-9.
99. Nathwani D, Williams F, Winter J, Winter J, Ogston S, et al. (2002) Use of indicators to evaluate the quality of community-acquired pneumonia management. Clin Infect Dis. 34(3):318-23.
100. Morse RB, Hall M, Fieldston ES, McGwire G, Anspacher M, et al. (2011) Hospital-level compliance with asthma care quality measures at children's hospitals and subsequent asthma-related outcomes. J Am Med Assoc. 306(13):1454-60.
101. Schopfer DW, Whooley MA, Stamos TD (2012) Hospital compliance with performance measures and 30-day outcomes in patients with heart failure. Am Heart J. 164(1):80-6.
102. Mansi IA, Shi R, Khan M, Huang J, Carden D (2010) Effect of compliance with quality performance measures for heart failure on clinical outcomes in high-risk patients. J Natl Med Assoc. 102(10):898-905.
103. Marcin JP, Romano PS (2004) Impact of between-hospital volume and within-hospital volume on mortality and readmission rates for trauma patients in California. Crit Care Med. 32(7):1477-83.
104. Patterson ME, Hernandez AF, Hammill BG, Fonarow GC, Peterson ED, et al. (2010) Process of care performance measures and long-term outcomes in patients hospitalized with heart failure. Med Care. 48(3):210-6.
105. Roccaforte R, Demers C, Baldassarre F, K.Teo K, Yusuf S (2005) Effectiveness of comprehensive disease management programmes in improving clinical outcomes in heart failure patients. A meta-analysis. Eur J Heart Fail. 7(7):1133-44.
106. Gawlas I, Sethi M, Winner M, Epelboym I, Lee JL, et al. (2012) Readmission After Pancreatic Resection is not an Appropriate Measure of Quality. Ann Surg Oncol. 1-7.
107. Jimenez-Puente A, Garcia-Alegria J, Gomez-Aracena J, Hidalgo-Rojas L, Lorenzo-Nogueiras L, et al. (2004) Readmission rate as an indicator of hospital performance: The case of Spain. Int J Technol Assess Health Care. 20(3):385-91.
108. Ricciardi MJ, Selzer F, Marroquin OC, Holper EM, Venkitachalam L, et al. (2012) Incidence and predictors of 30-day hospital readmission rate following percutaneous coronary intervention (from the national heart, lung, and blood institute dynamic registry). Am J Cardiol. 110(10):1389-96.
109. Auerbach AD, Maselli J, Carter J, Pekow PS, Lindenauer PK (2010) The relationship between case volume, care quality, and outcomes of complex cancer surgery. J Am Coll Surg. 211(5):601-8.
110. Rumball-Smith J, Blakely T, Sarfati D, Hider P (2013) The mismeasurement of quality by readmission rate: How blunt is too blunt an instrument?: A quantitative bias analysis. Med Care. 51(5):418-24.
111. Luthi JC, Burnand B, McClellan WM, Pitts SR, Flanders WD (2004) Is readmission to hospital an indicator of poor process of care for patients with heart failure? Qual Saf Health Care. 13(1):46-51.
112. Mayer EK, Bottle A, Aylin P, Darzi AW, Athanasiou T, et al. (2011) The volume-outcome relationship for radical cystectomy in England: An analysis of outcomes other than mortality. BJU Int. 108(8 B):E258-E65.
113. McConnell KJ, Lindrooth RC, Wholey DR, Maddox TM, Bloom N (2013)Management practices and the quality of care in cardiac units. JAMA Intern Med. 173(8):684-92.
114. Marang-van de Mheen PJ, Dijs-Elsinga J, Otten W, Versluijs M, Smeets HJ, et al. (2010) The importance of experienced adverse outcomes on patients' future choice of a hospital for surgery. Qual Saf Health Care. 19(6):e16.

115. Tsai TC, Joynt KE, Orav EJ, Gawande AA, Jha AK (2013) Variation in surgical-readmission rates and quality of hospital care. N Engl J Med. 369(12):1134-42.

APPENDIX

Appendix 1. Descriptive information of included studies

Author et al.	Year	Study type	Condition/ population group	Main data source and year	Definition incl. time frame	Risk adj.	Nr. of patients incl. (denominator)	
The context in which readmission rates are use								
Kangovi[14]	2011	discussion paper	/	/	/	/	/	
Spector[15]	2012	retrosp. cohort study	elderly/ injury	Healthcare Cost and Utilization Project State Inpatient Databases 2006	30-day, all-cause inpatient readmission	yes	224193	
Shep-perd[17]	2013	s.rev	/	/	/	/	/	
Lambri-nou [18]	2012	s.rev	hf	/	all-cause readmissions	/	/	
Wong[19]	2011	RCT	/	/	28 days rr		555	
Demir[20]	2012	multilevel transition model	/	1997/2004 national hospital episodes statistics dataset	test different time windows	yes	More than 5 million patient / 1000 random samples	
Jha[21]	2009	retrosp. cohort study	chf	CMS Hospital Quality Alliance data 2008	all-cause 30-day readmission	yes	2222 hospitals	
Ben-bassat[22]	2013	meta-review of published s.rev	hf, coronary heart disease and bronchial asthma, stroke, chronicm disorders	/	/	/	/	
Burke[23]	2012	discussion paper	/	/	/	/	/	

	Conclusion on the validity of the indicator and on the management/governance purpose of the indicator
	Social factors need to be considered when determining readmission risk. Policy makers should monitor the consequences of readmission penalties. Next to 30-day rrs, 90- and 180-day rates, ED and observation service use, and mortality should be monitored. The readmission penalty incentivizes hospitals to expand their scope beyond traditional health services. Although hospitals cannot assume responsibility for all patients social needs, they can serve as a vehicle for connecting high-risk patients to resources like addiction counseling and community centers. Readmission policy could reward hospitals that address the root causes of readmission. The penalties also have the potential to create new access barriers.
	Hospital experiences and injury characteristics influence rrs. a strategy to reduce rrs should not only focus on hospitals but also nursing homes and home health care.
	Discharge plan tailored to the individual patient probably brings about reductions in rrs for older people admitted to hospital with a medical condition;
	Hf management programs have potential to reduce readmissions; more research is needed
	Using volunteers as substitutes for some of the professional care, may be effective for general medical patients
	4 out of 5 worst performing hospitals are hospitals treating cancer patients. These hospitals are known to be the leading NHS Trusts in England, providing diverse range of services to complex patients and therefore it is inevitable to expect higher numbers of emergency readmissions.
	Current effort to collect and publicly report data on discharge planning are unlikely to yield large reductions in unnecessary readmission.
	Efficiency of in hospital interventions aimed at reducing rr need further study; for patients with heart diseases and bronchial asthma, rr may be considered as a publicly reported qi of community care, provided that future research confirms that efforts to reduce rr do not adversely affect other patients outcomes, such as mortality.
	Controversy exists over the penalties fairness and likelihood of driving appropriate behavior. We call for development and use of multifaceted, collaborative transitions interventions that span settings, risk-adjustment models that allow for fairer comparisons.

Indicator definition								
Wal-raven[2]	2011	s.rev*	/	/	avoidable readmissions			
Roche-fort[25]	2012	lit. rev/ discussion paper	major cancer surgery	/	/	/	/	
McCor-mack[26]	2013	retrosp. database analyses	orthopedic surgeries	hospital admin. data 2007/9	planned/un-planned 30-day readmissions	/	/	
Brooke[27]	2012	discussion paper	vascular surgeries	/	/	/	/	
Fischer[28]	2011	s.rev.	/	/	/	/	/	
Van Wal-raven[29]	2012	meta-analysis	/	/	30-day urgent readmissions	/	/	
Court-ney[30]	2003	retrosp. case-note review	general surgical division	case notes 2000	28-day emergency readmissions	/	1914	
Rumball-Smith[3]	2009	lit. review	/	/	unintended, acute readmission within 30-days	/	/	
Demir[32]	2008	modeling approach to define time window	copd, hf, stroke, congestive hf	National HES data (England) 1997/2004	/	/		
Maurer[33]	2004	pilot study	internal medicine	1998		/	884	
The effect of competing outcomes								
Gheorg-hiade[34]	2013		hf	OPTIMIZE-HF registry 1993/2006	30-day readmissions	yes	41267	
Almouda-ris[36]	2012		colorectal cancer resection	NHS data England 2000/1, 2007/8	unplanned readmission within 28-days	yes	144542	
Krum-holz[37]	2013		hf, AMI, pn	Medicare data 2005/8	30-day readmissions	yes	586027	
Grocott[35]	2010	lit. review	elective surgery	/	/	/	/	

	True proportion of potentially avoidable readmissions is unknown. Given the large variation in the proportion of avoidable readmissions between studies using primary data, 'avoidability' cannot accurately be inferred based on diagnostic codes for the index admission and the readmission.
	The proposed congressional plan to use rrs to assess hospital performance and determine reimbursement should be pursued with extreme caution, pending further investigation
	Careful data collection and abstraction when calculating early rr are important. Preventing surgical site infection and better coordinating care between orthopedic surgeons and primary care and medical subspecialty physicians may significantly reduce rr.
	Risk factors associated with readmission remain poorly defined, and further research is needed to propose a conceptual model to explain the driving forces behind readmissions in vascular surgery.
	The rr as an indicator should be used with care. Users should address issues related to definition, timeframe and case-mix correction as part of the process to enhance validity in the European setting
	Health system planners need to use caution in interpreting all cause readmission statistics as they are only partially influenced by qoc
	Accurate emergency 28-day readmission rates can potentially highlight where care has been suboptimal, and readmissions and even deaths prevented. However, just comparing rates between trusts does not allow for inaccuracies in coding and makes no attempt to quantify the number of avoidable readmissions which, with better qoc, could be avoided.
	Rr as a marker of the quality of hospital care has been used internationally and nationally, although its validity has only been partially substantiated. While prone to confounding, it remains a valuable indicator due to its ease of collection and its ability to be able to be combined with other variables.
	Given that the NHS performance rating framework regards rr as one of its key indicators, the research suggests that some hospitals may be disadvantaged by the use of one single number to define a time window.
	Most readmissions occur within 30 day timeframe, 90 days consideration did not change the overall conclusion
	A reduction in rrs, but this will require integration of these efforts on clinician/hospital/system level to improve overall outcomes.
	Appraising institutional service quality in colorectal surgery is complex. High mortality outlier status does not necessarily predict poor performance across other domains. A clear understanding of the scope of each outcome metric used to reflect performance is required if these are to be openly reported.
	Our findings indicate that many institutions do well on mortality and readmission and that performance on one does not dictate performance on the other. From a policy perspective, the independence of the measures is important. A strong inverse relationship might have implied that institutions would need to choose which measure to address.
	Use PROMS, post-discharge follow up, simple questions about admission to other institutions, track mortality to identify cases when patient die at home/attend a different medical institution other than the index institution for complications.

Basu[38]	2010	retrosp. analysis	/	clinical performance data 2006/7	28-day readmissions	/	/	
Hegge-stad[39]	2002	cross-sectional study	elderly patient	national patient database of admissions to all acute care Norwegian hospitals 1996	30-day unplanned readmission	yes	113055	
Kaboli[40]	2012	observa-tional study	hf, COPD, AMI, pn, gastro intestinal hemorrhage	observational data 1997/2010	30-day readmissions	yes	4124907	
Shulan[39]	2013	retrosp. database analysis	/	Veterans Healthcare Network Upstate New York data 2011/12	30-day readmissions	yes	8718	
Kramer[42]	2013	retrosp. cohort study	patients discharged from ICU	admin. data (APACHE) system 2002/10	ICU readmissions	yes	369129	
Ahmed[43]	2011	s.rev	/	/	/	/	/	
Dobrza-na[44]	2006	desc. data analysis	elderly patients	patient records 2002/3	28-day emergency readmissions	yes	4222	
Schnei-der[45]	2012	retrosp. data analysis	colorectal cancer surgery patients	SEER Medicare database 1986/2005	30-day readmissions	yes	149622	
Dunlay[46]	2012	retrosp. cohort study	AMI	population based registry 1987/2010	30-day readmissions	yes	3010	
Dallal[47]	2012	retrosp. multi-variable analysis	gastric bypass patients	medical record review 2006/10	30-day readmissions	yes	1065	
Lau[48]	2001	retrosp. study	COPD	patient chart extraction 1997	/	yes	551	

	This study highlights the difficulty in comparing the performance of hospitals due to the inherent lack of consistency. Therefore it is apparent that any 'reward-rebuke' system linked to performance should interpret the data with caution. It is therefore suggested that easy to control single value activities and standardized routine activities could be used to measure hospital performance. The hospital could compare with its statistics from previous years.
	With optimal hospital care and planning of aftercare, it is possible to prevent some new admissions or prolong the time period from discharge to a readmission. The consequences or costs of readmissions may be experienced differently by individual patients and their families compared with what is the case at the hospital or system level. So when monitoring system performance, readmission measures are relevant, particularly when considering vulnerable groups such as the elderly.
	Payors of health care and quality organizations should carefully examine (un)intended consequences of financial incentives for episodes of care in our often fragmented health care system.
	Demographics, socio-economic variables, prior utilization and Diagnosis-related Group (DRG) all have limited predictive power; more sophisticated patient stratification algorithm or risk adjuster is desired for more accurate readmission predictions. Even though our model can be used in both hospital comparisons and patient identification for readmission reduction, caution needs to be exercised. High predictive power is a necessary but not a sufficient condition for valid comparisons; unlike predictions, hospital comparisons demand more delicate use of confounders (e.g. health literacy) and more complete data. For predictions, the model should not and cannot replace clinical judgment.
	We found no association between unit readmission rate and quality of care as reflected by aggregate mortality and lengths of stay when adjusted for patient case-mix. the use of readmis¬sion as a quality measure should only be implemented if patient case-mix is taken into account.
	Higher compliance with after surgery protocols was associated with a reduced length of hospital stay. However, reduced length of hospital stay was associated with a high rate of readmission.
	Institutionalized patients were readmitted sooner than those who lived at home: those discharged home vs. other sources and agreeing to increased social service provision had longer stays on readmission. Shorter los on index admission (up to 72 hours) was associated with increased likelihood of earlier readmission.
	Rrs after colectomies have increased during the past 2 decades and mean LOS after this operation has declined. More research is needed to understand the balance and possible tradeoff between these hospital performance measures for all surgical procedures.
	Despite advances in medical care following MI, readmission risk has not declined. Angiography, reperfusion and revascularization are mainstays of therapy in acute MI, and complications are associated with a large risk of readmission. Comorbidities appear to play a central and increasing role in readmissions. Comprehensive strategies of care in patients with MI need to be deployed that incorporate the treatment of both cardiovascular and non-cardiovascular disease in order to prevent future hospitalizations.
	Early discharge on postoperative day 1 is possible but no modifiable, random patient factors challenge predictable discharge planning. Reliable discharge on postoperative day 1 is not likely with current technologies.
	Following factors were associated with shorter time to readmission: hospital admission within 1 year before index admission, los in index admission 45 days, nursing home residency, dependency in self-care activities, right heart strain pattern on electrocardiogram, on high dose inhaled corticosteroidandactualbicarbonatelevel425mmoll. They may be relevant in the future planning of healthcare utilization for COPD patients.

Chan[24]	2011	retrosp. study	COPD	analysis of admission episode data of acute medical wards 2006/7	unplanned 30-day readmissions	yes	65497	
Kiran[49]	2004	retrosp. data analysis	intestinal surgery patients	data from hospital Computer records 2002	30-day readmissions	yes	553	
Kociol[50]	2013	retrosp. data analysis	hf	clinical data from GWTG-HF quality improvement registry linked to Medicare claims 2005/6	30-day readmissions	yes	17387	
Johnson[51]	2012	retrosp. data analysis	/	single hospital data analysis 2004/6	30/120 same-center all-cause readmissions to general medicine units	yes	4151	
Case-mix correction using administrative data vs. clinical data								
Van Walraven[52]	2012	retrosp. data analysis	/	discharge abstract database of 162 hospitals 2005/10	urgent readmissions within 30-days	yes	3148648	
Khawaja[53]	2012	analysis of prosp. data registry	PCI	prospective registered database 1998/2008	30-day readmissions	yes	15498	
Goldfield[54]	2008	retrosp. data analysis	/	hospital admin. database analysis 2004/5	potentially preventable readmissions	yes	5.02 million	
Kansagara[55]	2011	s.rev.	/	/	/	/	/	

	A systematic approach in program provision in the community and a good discharge planning process targeting on COPD patients who are at high risk of unplanned readmission are essential.
	Early readmission is an unpredictable sequel of major bowel operations; it does not correlate with shorter los. Identification of unpredictable complications does not adversely affect clinical outcomes. 30 days readmissions of patients who have attained standardized discharge criteria may not be a valid indicator of qoc.
	Although hospital 30-day rr was poorly correlated with LOS, quality measures, and 30-day mortality, better performance on the EOC metric was associated with better 30-day survival. Total inpatient days during a 30-day EOC may more accurately reflect overall resource use and better serve as a target for quality improvement efforts.
	The findings suggest that more hospital care may not affect the likelihood of readmission and thus denying payment for readmission may be unwarranted.
	The study found notable variation in rates of urgent readmission within 30 days based on the extent of adjustment for confounders and the unit of analysis. Slight changes in the methods used to calculate hospital-specific rrs influence their values and the consequent rankings of hospitals. Our results highlight the caution required when comparing hospital performance using rates of death or urgent readmission within 30 days.
	About 1 in 10 PCI procedures resulted in a readmission within 30 days, most readmissions were due to a cardiovascular cause. Patients readmitted within 30 days of discharge were at an increased risk of 1-year mortality.
	MedPAC has proposed that Medicare should reduce payments to hospitals with high rrs. From a policy perspective the key challenge is to establish the extent of the payment reduction for a readmission. For true medical errors that are clearly related to mistakes in the delivery of care, not paying for the readmission may be justified. However, most readmissions are not so clearly linked to medical errors, and, although they may possibly relate to errors in judgment or lapses in execution that reflect poor quality care, they cannot be considered always preventable. A balance between the relative preventability of a readmission and the extent of the payment reduction associated needs to be achieved. The financial consequences of a readmission need to be significant enough to motivate hospitals to reduce rrs, without penalizing hospitals for events over which they have limited control. MedPAC is essentially proposing that the extent of the payment reduction for a readmission be set separately for each hospital based on its risk adjusted rr. Hospitals with lowest risk adjusted rrs would have a small reduction in payment for readmissions, while hospitals with high rrs would have a larger payment reduction. The advantage to this approach is that an estimate of the relative preventability of readmissions does not have to be made. The amount of the payment reduction is based on the relative overall performance of hospitals in terms of their risk adjusted rr.
	Most current readmission risk prediction models, whether designed for comparative or clinical purposes, perform poorly. Though in certain settings such models may prove useful, efforts to improve their performance are needed as use becomes more widespread.

Correll[56]	2007	retrosp. data analysis	asthma	linked hospital and ED data 2000/3	readmissions to hospital or emergency department within 28-day for asthma	yes	139043	
Gia-mouzis[57]	2011	lit. review	hf	/	/	/	/	
Data reliability								
Cram[58]	2012	admin. data analysis	THA	Medicare Provider Analysis/ Review (MedPAR) 2007/8	30-day readmissions	yes	202773	
Hannan[59]	2011	retrosp. data analysis	CABG surgery	Cardiac Surgery Reporting System 2005/7	30-day readmissions	yes	33936	
Wal-raven[60]	2012	retrosp. data analysis	/	hospital admin. data	urgent 30-days readmissions	/	1666	
Bianco[61]	2012	cross sectional data analysis	/	medical records review 2005/7	30-day readmissions	yes	2252	
Ben-Assuli[62]	2013	Track log-file analysis	/	electronic health records information system data 2004/7	7-day readmissions	yes	281750	
Holt[63]	2010	retrosp. data analysis	AAA repair	HES data 2003/8	all-cause emergency readmissions	yes	143237	

	There is evidence that programs to improve asthma care in the hospital setting result in reduced hospital readmissions. However, the likelihood of readmission for asthma is not only related to the qoc in hospital, but also to the qoc in the community, in particular, from the patient's general practitioner or specialist.
	30-day post discharge HF rrs are currently used as quality measures. This consideration raises several concerns. The published risk-adjusted hf readmission predictions models are not optimal for accurate risk prediction for reasons stated above. Performance of other models that are currently unpublished in peer-reviewed journals cannot be judged. There are no gold standard rules for when a person should be admitted with hf, and patients may be hospitalized for borderline clinical/nonclinical social reasons. Finally, qis such as hf rr might precipitate provider behaviors regarding resisting admissions; this may be a safety concern.
	The study found important differences in the rates of THA complications depending upon the coding algorithms and time frame employed. Our results suggest that admin. data can be used to evaluate THA complications but that methodology should be carefully considered.
	The authors speculate that readmissions have not decreased over time despite decreasing mortality rates, because efforts to coordinate in/outpatient care have been insufficient. Since the announcement by the Centers for Medicare/Medicaid Services that reimbursement for readmissions would be curtailed or eliminated, numerous organizations have sponsored symposia for coordinating inpatient and outpatient care, it is hypothesized that hospitals that follow these suggestions will experience considerably reduced rrs.
	Urgent rrs can be used to estimate the probability that avoidable rrs differ significantly between two hospitals. Hospitals with significantly lower urgent rrs frequently have negligibly lower avoidable rrs than comparator hospitals. These results highlight that urgent rrs should be used cautiously to compare hospital qoc
	Avoidable readmissions are influenced by factors at the patient, organizational, and environmental levels. Not all of these factors are actually in the hospitals' control, although the cost of avoidable readmissions will be borne principally by hospitals that can modify their structure and processes. More qualitative information may improve the patient's ability to manage aspects of his care after discharge, and for the elderly through interviews with the patients, family members, and caregivers about their needs. Management is one of the reasons that justifies vertical integrated, moreover, active post-discharge delivery systems that are poised to facilitate transitions between inpatient and outpatient care.
	Viewing medical history via an EHR IS and using HIE network led to a reduction in the number of 7 day readmissions and single-day admissions for all patients. Using external medical history may imply a more thorough patient examination that can help eliminate unnecessary admissions. In most instances physicians did not view medical history at all.
	Readmissions at either 30-days or 1-year were associated with high mortality rates. This suggests that readmission is a useful outcome measure both for the burden on emergency services and because there is the underlying cause of the readmission, which has a direct impact on survival.

Forster[16]	2012	discussion paper	/	/	all-cause urgent rr within 30-days	/	/	
Amin[64]	2012	chart review	spine surgery	admin. hospital data 2007/11	30-days all-cause readmission	yes	5780	
Lin-denauer[65]	2011	retrosp. cohort study	elderly patients/pn	Medicare claims 2005/6	hospital-specific 30-day readmissions	yes	453251	
Sellers[66]	2013	retrosp. cohort study	surgical patients	The American College of Surgeons National Surgical Quality Improvement Program data 2011	30-day postoperative all-cause readmissions	/	1748	
Krum-holz[67]	2013	retrosp. data analysis	AMI	Medicare claims 2006	30-day all-cause readmissions	yes	279152	
Wall-mann[68]	2013	cohort study	cardiovascular disease	hospital admin. data 2003 /9	30-day cardiac-related readmissions	yes	37381	
Ade-yemo[69]	2007	retrosp. audit case note records	general surgery	case not review 2003	adjusted unplanned re-admission within 28 days	yes	2652	

	Current approaches to measure qoc cannot fully meet the public's expectation, that providers should, with fair methods, be held accountable. For this reason, we suggest several actions. First, we need a better understanding of the limitations of current indicators. To achieve this goal, the health sector will need to make modest investments in research on qis; this would cost significantly less than many other interventions in the health care system. To avoid perceived conflicts of interest, this investment should be directed toward experts who are able to critically assess qis but are not themselves responsible for holding people accountable. Second, providers need to be fully engaged in the measurement, analysis, and interpretation of indicator data. To achieve this goal, we recommend a gradual transition in funding methodology such that providers would be more accountable for their actual performance. Third, the entire health care system needs to increase investment in performance measurement systems. The current approach—measuring what we can instead of what we should—is inadequate, especially in light of the large investments in health care services.
	Findings identify the potential pitfalls in the calculation of rrs from admin. data sets. Benchmarking algorithms for defining hospitals' rrs must take into account planned staged surgery and eliminate unrelated reasons for readmission. When this is implemented in the calculation method, the rr will be more accurate. Current tools overestimate the clinically relevant rr and cost.
	Rehospitalization within 30 days of treatment for pneumonia is common, and rates vary across hospitals. A risk standardized measure of hospital rrs derived from admin. claims has similar performance characteristics to one based on medical record review. The development of a valid measure of hospital performance and public reporting are important first steps towards focusing attention on this problem. A necessary prerequisite for public reporting of readmission is a validated, risk-adjusted measure that can be used to track performance over time and can facilitate comparisons across institutions.
	Admin. data accurately captured all-cause readmissions, but could not identify unplanned readmissions and less consistently agreed with chart review on cause. The granularity of clinically collected data offers tremendous advantages for directing future quality efforts targeting surgical readmission. If surgical rr is to be used as a quality metric, clinical data are likely a better source to use in order to more closely target preventable readmissions.
	Rrs are influenced by the quality of inpatient and outpatient care, the availability and use of effective disease management programs, and the bed capacity of the local healthcare system. Some of the variation in readmissions may be attributable to delivery system characteristics. Additionally, interventions during and after a hospitalization can be effective in reducing rrs in geriatric populations generally 18-20 and for patients with AMI specifically.
	The study provides a prediction model for 30-day cardiac-related diseases based on available admin. data ready to be integrated as a screening tool. It has reasonable validity and can be used to increase the efficiency of case management. There were series of factors that might be important for predicting readmission but could not be considered because of the design of the study. The admin. data did not contain several clinical variables that could only be obtained by chart review, quality of life measurements, psychosocial and behavioral factors and use of primary care services or outpatient management. It is likely that a more powerful prediction model may be derived from a combination of all these dimensions, but availability of this data is very difficult, costly and time consuming. A more comprehensive hospital information system would allow exploring more variables and fit the prediction model in the future.
	Unplanned general surgical re-admission rates collated from hospital PAS systems may be inaccurate. Nearly half of 'genuine', unplanned re-admissions involved patients with chronic and/or recurrent symptoms, which are predictable and may be preventable. Significant postoperative complications accounted for few re-admissions in this study. While it is likely that these errors in the information database are wide-spread in the NHS, the magnitude may vary between hospital trusts (perhaps even with time) and may lead to breach(es) of set targets and (any consequent) penalties.

Parker[70]	2011	retrosp. data analysis	/	automated hospital and pharmacy data 1993/5	14-day readmissions 30-day unplanned readmissions; unplanned readmissions	yes	6721	
Validity of readmission rates as a quality measure								
Rosen[71]	2013	retrosp. admin. data analysis	/	VA Patient Treatment File-admin. database 2003/7	30-day readmissions	yes	2332794	
Mokhtar[72]	2012	retrosp. cohort and case-control study	diabetic patients	medical records 2000/8	28-day readmissions	/	1125	
Kergoat[73]	2012	retrosp. data analysis	geriatric patients	medical records review 2002/3	hospital readmissions	yes	934	
Auer-bach[74]	2009	obser-vational cohort	CABG	hospital discharge data 2003/8	30-day readmissions	yes	81289	
Stukel[75]	2010	population-based longitudenal cohort study	AMI	patients records review 2000/6	30-day and 1-year readmissions	yes	89115	
Chen[76]	2012	retrosp. data analysis	diabetic patients	admin. claims data 2009/10	2-30 day readmissions	yes	30139	
Weber[77]	2010	retrosp. data analysis	head and neck surgery	institutional enterprise information warehouse 2004/8	30-day readmissions	yes	2618	

	A pharmacy-based measure of comorbid illness provided predictive performance similar to that of an ICD-9-based measure of comorbidity for hospital readmission and LOS. This is noteworthy because data contained in secondary diagnoses are recorded by hospital medical personnel for the purpose of documenting prognostic illness in patients and pharmacy claims were never intended for that purpose. Researchers concerned about the overall reliability of hospital coding of patients' multiple conditions may be able to use equally predictive pharmacy data. Deficiencies in a diagnosis-based measure of risk may be somewhat ameliorated by including outpatient pharmacy data. This combined admin. data strategy is certainly cost-effective and may prove increasingly practical as more medical data become automated. Groups that have access to both pharmacy and other automated medical data are already using both sources of information to better understand and predict utilization within their patient populations.
	Interventions that focus on minimizing preventable inpatient safety events as well as improving coordination of care between and across settings may decrease the likelihood of readmission. Although previous studies have linked readmissions with the quality of inpatient care during the index hospitalization, our study suggests that focusing on AEs provides another important way of viewing readmission as an indicator of qoc.
	Adherence to American Diabetes Association guidelines for admission work-up and readiness for discharge criteria were significantly more likely to decrease the risk of readmission within 28 days. Quality of inpatient care exerts a substantial influence on the risk of readmission. Hospital should improve the qoc delivered to diabetic patients, special emphasis should be placed on attendance at outpatient clinics for follow up after discharge.
	A large gap between geriatric care principles and practice in GAUs has been observed. Improvement in care processes may be translated to decreased short-term health services use and mortality. This underlines the importance to examine structural and patient-related factors that promote qoc processes.
	Overall qoc was not associated with readmission risk - a particularly striking finding when juxtaposed with the association between quality and 30-day mortality rates. Although our data do not allow us to directly test these hypotheses, readmission risk may be more influenced by care delivered at discharge, such as care planning or the presence of support during the post discharge period, whereas in-hospital death is dependent more on decisions and care provided earlier in the hospital stay, such as medications.
	Hospitals with higher levels of both medical and interventional management and higher quality initial ED assessment had better outcomes. Readmissions were particularly sensitive to care processes. In the face of the unwarranted variations in outcomes across hospitals, strategies that promote better ED and inpatient management of AMI patients are needed.
	Receipt of LDL testing and adherence to statin medications were effective in decreasing the likelihood of 30-day hospital readmission and may be considered as elements of a quality focused incentive-based health care delivery package for diabetes patients. Findings of this study also suggest that incentive-based quality programs to increase the use of and adherence to statins among all patients with diabetes regardless of LDL level may be worthy of consideration.
	Other factors that affect outcomes must be included in performance comparison. Two important factors are procedure acuity and the patient's comorbid conditions. Although both factors have significant effects, the acuity of the procedure is the strongest determinant of the incidence of negative quality and pis and should be heavily weighted in physician comparisons. These data may serve as a tool to evaluate performance, to positively impact outcomes through identifying best practices and providing data to individual surgeons who may positively affect their patients' outcomes.

Kent[78]	2011	prosp. collected data analysis	pancreas resection	prospectively entered data 2001/9	30-day readmissions	yes	578	
Barbieri[79]	2007	retrosp. database analysis	cystectomy	university Health System Consortium Clinical Database 2002/5	/	yes	6728	
Chung[80]	2008	retrosp. database analysis	hf	medical records review 2004/6	all-cause readmissions	yes	400	
Judge[81]	2006	retrosp. database analysis	hip/knee joint replacements	HES data 1997/2002	1-year readmissions	yes	492459	
Boul-ding[82]	2011	obser-vational analysis	AMI, hf, pn	Hospital Compare data 2005/8	30-day risk-standardized readmissions	yes	2373082	
Youn[83]	2012	retrosp. database analysis	hf	Heart Failure Registry 2004/9	cause-specific readmissions	yes	1527	
Van-Such[84]	2013	retrosp. database analysis	hf	admin. and medical record data 2002/3	hf readmissions; all-cause readmissions	yes	1121	

	Readmissions after pancreatic resection were frequent, early, costly, and largely related to procedure-specific complications. As initial hospital stay continues to decline in high-acuity surgery, readmissions might be required for optimal management of complications, which often manifest later in the recovery course. Clinical pathway deviations predict potential readmissions, and might prompt adjustments in management and disposition of patients at risk for returning to the hospital. Use of readmission data as a quality measure for surgical patients must follow a decision on how to interpret such readmissions: are they failed discharges or are they rescues? The rrs reported here and by others must be considered in the context of the complexity of pancreatic disease and surgery, and do not necessarily reflect poorly on the qoc. As the focal caregivers for our patients, we as surgeons should promote caution when using rrs as qis for our performance. At the very least, we should have a stake in the discourse, as public policy is developed in this domain.
	Even among academic medical centers hospitals with a higher volume of cystectomies were associated amongst others lower rehospitalization rates. These data may provide a framework for self-assessment and help establish criteria for performance evaluation.
	Adherence to a composite of HFPM appears to be related to a reduction in all-cause readmissions for patients with hf but not short-term mortality. Given the resources required for data collection and reporting, it is important to further evaluate whether or not these measures are associated with near-term clinical benefit before they can be used as an appropriate and meaningful surrogate measure for qoc that is subsequently linked to reimbursement.
	In England, there are fewer adverse events following TJR in high volume centers and in orthopedic training centers. Standardization of procedures may account for this finding. The data have implications for private practice in the UK and for the current move to undertake TJRs in Independent Sector Treatment Centers.
	Higher overall patient satisfaction with discharge planning are associated with lower 30-day risk-standardized hospital rrs after adjusting for clinical quality. This suggests that patient centered information can have an important role in the evaluation and management of hospital performance. Despite dramatic improvements in clinical process performance for hf, there has been virtually no reduction in these rrs or costs. Our findings confirm the lack of association between hf clinical measures and rrs. Conversely, we found that patient-reported measures were highly associated with 30-day rrs. Therefore, patient perceptions about hospital care in general and discharge planning specifically may provide an important new tool for measuring the quality of transitions of care.
	For patients with LVSD in Korea, adherence to treatment performance measures is associated with improved clinical outcomes. Although it is expected that application of these measures would result in substantial improvement in outcomes for hf patients, guideline recommendations for these measures are based on expert opinion.
	There are likely many other factors which influence readmission that are not dealt with by the hf core measures. Further studies are needed to examine the additional aspects of both hospital care and care after discharge for patients with hf, to determine the best practices that should be endorsed to assure the highest qoc. This study presents stronger evidence for the use of discharge instructions as an evidence-based measure than has been produced previously. Documentation of discharge information and patient education does appear, in fact, to be associated with reductions in both mortality and readmissions.

Showalter[85]	2011	retrosp. pre- and post-implementation comparison cohort study	/	data from electronic discharge document 2005/6 2007/8	readmissions or ED visit within 30-days	yes	34088	
Polanczyk[86]	2001	prosp. collected data analysis	chf	prosp. collected data 1997	3-months readmissions	yes	205	
Luthi[87]	2003	retrosp. data analysis	hf	medical records 1995/6	Readmissions up to 21 months	yes	2943	
Joynt[88]	2011	retrosp. data analysis	chf	Medicare claims 2006/7	30-day risk-adjusted readmissions	yes	1029497	
Horwitz[89]	2012	cross-sectional study	AMI, hf, pn	Medicare claims 2007/9	30-day risk-adjusted readmissions	yes	3013616	
Hernandez[90]	2010	retrosp. data analysis	hf	OPTIMIZE-HF data 2003/4	1-year cardiovascular readmissions	yes	20441	
Bottle[91]	2009	retrosp. data analysis	/	HES data 2005/6	unplanned readmissions within 28-days	yes	13974949	
Halfon[92]	2002	retrosp. data analysis	/	hospital system data 1997	foreseen/ unforeseen 1-year readmissions	yes	3474	
Halfon[93]	2006	retrosp. data analysis	elective surgery	medical records 2000	30-day adjusted rates of potentially avoidable readmissions	yes	131809	
Maeda[94]	2010	s. rev.	hf	/	/	/	/	

	Having a standardized discharge instructions process will not sufficiently address the complex issues around post-discharge hospital utilization. Multi-faceted interventions that comprehensively address numerous aspects of the discharge process will be more likely to have a meaningful impact on post-hospital utilization, such as discharge conversations between health care providers and patients/caregivers, patient health literacy, appropriate post-hospital follow-up, and communication with outpatient providers. Although health IT initiatives are likely a necessary building block for achieving national health care quality goals, whether they result in intended outcomes needs to be evaluated closely. Additional work is needed to further understand the reasons for preventable hospital readmissions and post-discharge ED visits.
	Hospital qoc for patients with chf is independently associated with 3-month rrs, and cardiologist involvement during hospitalization is associated with overall qoc.
	Results show that ACEI use at discharge in patients with LVSD is associated with decreased rrs. This suggest that compliance with the ACEI prescribing recommendations listed in clinical practice guidelines for patients with hf due to LVSD confers benefit.
	Experience with managing chf, as measured by an institution's volume, is associated with higher qoc and better patient outcomes, but at a higher cost. Understanding which practices employed by high-volume institutions account for these advantages can help improve qoc and clinical outcomes for all chf patients. Admin. data cannot fully account for variations in severity of illness across hospitals. However, although admin. data are imperfect, they are standardized, validated, increasingly used, even for public reporting.
	Risk-standardized rrs are moderately correlated within hospitals, as are risk-standardized mortality rates. This suggests that there may be common hospital-wide factors affecting hospital outcomes.
	Several evidence-based processes of care are associated with improved outcomes, can discriminate hospital-level qoc, and could be considered as clinical performance measures. Given the moderate associations between individual process measures and clinical outcomes, it may be appropriate to include multiple new measures in hospital profiling.
	The move towards ''payment for performance''- led systems such as the UK's Payment by Results initiative should improve recording because of the financial incentive to accurately capture comorbidity and other clinical details—as is already the case in the US with DRGs. The Institute for Healthcare Improvement believes that data on safety are best used internally by the hospitals rather than being used to make judgments on them by external bodies.
	It is likely that other factors (which are not routinely collected) could bias the detection pf potentially avoidable readmissions. E.g. abnormal laboratory values, failure to ambulate, mental status, marital status, living situation, or income. Nevertheless, most readmissions are caused by the relapse of the original condition and by complications linked to inappropriate inpatient care
	Adjusted rates of potentially avoidable readmissions are scientifically sound enough to warrant their inclusion in hospital quality surveillance. Indicator variability should not result from practices unrelated to qoc. E.g. hospitals may decrease their potentially avoidable rrs by failing to readmit (or postponing) patients requiring inpatient care, or by upcoding unsupported risk factors. There is not much evidence that this may occur when the indicators are not subjected to public reporting. A strong perverse incentive would be for insurers not to reimburse early readmissions, as sometimes suggested. An explicit and automated algorithm to compute the indicator protects against falsification. However, the issue revolves mainly around the quality of the diagnoses and procedures coding.
	The findings from this systematic review suggest that an increase in compliance with the hf performance measures leads to a consistent positive impact on patient outcomes although the strength, magnitude, and significance of this effect is variable across the individual pi.

Encinosa[95]	2008	matched case-control study	surgical patients	MarketScan Commerical Claims and Encounter data 2001/2	readmission 90-days after surgery	yes	161004	
Shahian[96]	2012	retrosp. cohort study	AMI, hf, pn	hospitals data 2004/7	all-cause 90-day readmissions	yes	Between 374 to 3020	
Fonarow[97]	2007	retrosp. data analysis	hf	The Organized Program to Initiate Lifesaving Treatment in Hospitalized Patients with HF registry 2003/4	60- to 90-day mortality and combined mortality/ rehospitalization rates.	yes	5791	
Stefan[98]	2012	cross-sectional study	AMI, hf, pn	Hospital Inpatient Quality Reporting Program 2007	adjusted 30-day all-cause readmissions	yes	972-2940	
Nathwani[99]	2002	audit study	pn	audit data 1999/2000	readmissions within 2 weeks	yes	205	
Morse[100]	2011	cross-sectional study	asthma	admin. data 2008/10	asthma-related readmissions at 7,30,90 days	yes	37267	
Schopfer[101]	2012	retrosp. database analysis	hf	Hospital compare database	3-0-day readmissions	yes	3655 hospitals	
Mansi[102]	2010	retrosp. cohort study	hf	data collected for quality-management administration 2003/4	90-day readmissions	yes	357	
Marcin[103]	2004	population-based non-concurrent cohort study	trauma patients	computerized hospital discharge abstracts 1995/99	non-scheduled/ unknown 30-day readmissions	yes	102008	

	Medicare will stop reimbursing hospitals for the extra costs of eight patient safety events beginning in 2008. At that time, hospitals will have to revisit and re-evaluate their business case for increasing investments in patient safety improvements. One vital component of the business case is an accurate estimate of the potential returns to patient safety interventions.
	Adherence with recommended AMI and pn care processes is associated with improved long-term outcomes, whereas the results for hf measures are inconsistent. The evidence base for all process measures must be critically evaluated, including the strength of association between these care processes and outcomes in real-world populations. Some currently recommended processes may not be suitable as accountability measures.
	Current hf performance measures, aside from prescription of an ACEI or ARB at discharge, have little relationship to patient mortality and combined mortality/ rehospitalization in the first 60 to 90 days after discharge. Additional measures and better methods for identifying and validating hf performance measures may be needed to accurately assess and improve care of patients with hf.
	Hospitals with greater adherence to recommended care processes did not achieve meaningfully better 30-day hospital rrs compared to those with lower levels of performance.
	Regular evaluation of clinical performance through audit and feedback of information to stakeholders as a means of promoting change is now encouraged as part of the quality improving agenda of Clinical Governance within the UK. The emphasis is on implementation and evaluation of clinical care. The study has confirmed the value of qis in evaluating cap management and has stimulated the development and implementation of a local hospital–based integrated care pathway.
	No association between CAC-3 compliance and subsequent ED visits and asthma-related readmissions was found. The CAC-3 measure in its current form may not meet the criteria outlined by the Joint Commission for accountability measures. Until CAC-3 compliance can be linked to improved outcomes, the Joint Commission should reconsider whether the CAC-3 component of the measure set is appropriately classified as an 'accountability measure' suitable for public reporting, accreditation, or pay for performance. The Joint Commission embraces accountability measures as appropriate for use for public reporting and pay for performance. While the health of children hospitalized with asthma clearly improves when they receive relievers (CAC-1) and systemic corticosteroids (CAC-2), the CAC-3 measure, in its current form, may not meet the criteria set out for accountability measures.
	After adjusting for socio-economic factors and hospital volume, only two (out of four) hf process indicators were associated with lower readmissions.
	Compliance with the Joint Commission on Accreditation of Healthcare Organizations' TJC quality measures for hf was associated with higher rrs for hf. Several factors may explain this trend, including patient characteristics and focus on national reporting benchmarks rather than patient centered health care.
	Relationships between trauma volume and outcomes exist but depend on which patient populations are studied and how the data are analyzed. Furthermore, trauma centers may be subject to the detrimental effects of high temporal volume overextending existing services and capacity. Since this study found that both between-hospital volume and within-hospital volume measures are associated with outcomes, we recommend that both measures be included in future volume outcome investigations.

Patterson[104]	2010	retrosp. data analysis	hf	national clinical registry data 2003/4	1-year readmissions	yes	22750	
Roccaforte[105]	2005	s.rev.	hf	/	/	/	/	
Gawlas[106]	2012	prosp. populated database analysis	pr	case review 2006/10	30-day readmissions	yes	787	
Jimenez-Puente[107]	2004	cross-sectional observational study	/	clinical records review 1997	avoidable readmissions within 6-months	yes	784	
Ricciardi[108]	2012	retrosp. database analysis	percutaneous coronary intervention	Dynamic Registry 1997/2006	30-day readmissions	yes	10965	
Auerbach[109]	2010	observational cohort study	complex cancer surgery	Perspective database 2003/5	adjusted 30-day readmissions	yes	14170	
Rumball-Smith[119]	2013	retrosp. data analysis	surgical patients	hospital admin. data 2002/8	acute public hospital admissions within 30days	yes	89090	
Luthi[111]	2004	retrosp. cohort study	hf	hospital admin. data 1999	readmissions within 30-days	yes	1055	
Mayer[112]	2011	retrosp. data analysis	radical cystectomy	hospital admin. data 2000/1 2006/7	emergency readmissions within 30-days	yes	/	
Mc-Connell[113]	2013	survey; retrosp. data analysis	AMI	survey data and admin. data 2010	30-day readmissions for AMI	yes	597	

	Hospital process performance for hf as judged by current CMS measures is not associated with patient outcomes within 1 year of discharge, calling into question whether existing CMS metrics can accurately discriminate hospital qoc for hf.
	Disease management programs reduce mortality and hospitalizations in hf patients.
	The study found the vast majority of readmissions after pr were to manage complications related to the operation and were not due to poor coordination of care/ discharge planning. Because evidence-based measures to prevent these surgical complications do not exist, we cannot support the use of rrs as a qi after pr.
	Hospitals and health administrations of both continents include the rr among the indicators regularly monitored. Most of the hospital readmissions are not avoidable with the modification of medical care during the index episode. Thus, rrs cannot be considered valid qoc indicators for the set of specialties in a general hospital. It is proposed to continue evaluating whether they can be valid qoc indicators within specific diagnoses.
	Although in-hospital mortality and los after PCI have decreased over time, the observed 30-day cardiac rr was highly variable and risk of readmission was more closely associated with underlying patient characteristics than procedural characteristics.
	Although hospital and surgeon volume were not associated with outcomes, lower overall adherence to quality measures is associated with higher costs, but not improved outcomes. This finding might provide a rationale for improving care systems by maximizing care consistency, even if outcomes are not affected.
	The study found substantial error when using readmission as a marker of quality, and suggests that differences in readmission between populations are more likely to be due to factors other than qoc.
	Readmission did not predict and was not a valid indicator of the qoc for hf patients. Early readmission is sometimes interpreted as a problem following discharge due to inadequate care during the hospital stay, but post discharge factors may contribute to readmission. The current limitations of routinely available outcome measures such as readmissions and the lack of valid risk adj. methods, the AHA/ACC Scientific Forum on Quality of Care and Outcome Research in Cardiovascular Disease and Stroke does not recommend the use of readmission for comparing hospitals for patients with hf. However, this outcome measure might be documented and recorded over time and included as part of quality improvement projects within institutions. Adjustment for severity, comorbidity, other patient risk factors is crude and limited.
	The study found no statistically significant relationship between volume and complication or rrs. it may be that readmission as a performance measure is more suitable when assessing a relatively homogenous patient population and when the severity of the condition shows little variance, such as in day surgery setting.
	Management practices were not associated with lower rrs, a finding that may be consistent with evidence suggesting that 30-day rr are driven primarily not by hospital practice but by a hospital's patient population and the resources of the community in which it is located

* AAA repair- abdominal aortic aneurysm repair; ACEI- angiotensin-converting enzyme inhibitor; admin data- administrative data; aes-adverse events; arb-angiotensin receptor blocker; AMI-acute myocardial infarction; CABG-coronary artery bypass graft surgery; cap- Community-Acquired Pneumonia; CHF-congestive heart failure; COPD-chronic obstructive pulmonary disease; ed-emergency department; eoc-episode of care; GAUs-Geriatric Assessment Units; HES-Hospital Episode Statistics; hf-heart failure; HFPM-heart failure performance measures; ICU- intensive care unit; lit. review-literature review; los-length of stay; pci-percutaneous coronary intervention; pi-performance indicator; pn-pneumonia; prosp - prospective; pr-pancreatic resection; rct-randomized controlled trial; retrosp.-retrospective; risk adj.-risk adjustment; rr-readmission rate; rrs-readmission rates; s.rev- systematic review; tha- total hip arthroplasty; qi-quality indicator;

Chapter 4.

A systematic review on the validity of readmission rates as a quality indicator for heart failure

Claudia Fischer, Ewout W. Steyerberg, Gregg C. Fonarow, Theodore G. Ganiats, Hester F. Lingsma

This article was published as:
A systematic review and meta-analysis on the association between quality of hospital care and readmission rates in patients with heart failure.
American Heart Journal 2015;0:1-13.e2.

ABSTRACT

Introduction: In recent years, readmission rates have been increasingly used as a measure of quality of hospital care for patients with heart failure. The aim of this systematic review is to assess the scientific evidence regarding the relation between hospital readmission rates and quality of hospital care for patients with heart failure.

Methods: We defined quality of hospital care for patients with heart failure by adhering to the performance measures developed by the American College of Cardiology (ACC)/American Heart Association (AHA). Relevant articles published in English and indexed in the bibliographic databases Embase, Medline OvidSP, Web of Science, Cochrane Central, and PubMed were reviewed.

Results: Of the 2,638 studies identified, 18 were included. They varied widely in their methodology, data sources used, and study populations. We found mixed but rather limited evidence that there is a relationship between the ACC/AHA process measures and the rate of readmission. Four of 10 studies showed a significant correlation of readmission rate with "angiotensin-converting enzyme inhibitor/angiotensin receptor blocker use." Three of 9 studies showed a significant correlation between readmission rates and "evaluation of left ventricular systolic function." One of 7 studies showed a significant correlation with "smoking cessation counseling," and 2 of 8 showed a significant correlation with "providing discharge instructions." No evidence was found for a relationship between readmission rates and the performance measure "warfarin for atrial fibrillation."

Conclusions: Readmission rates after heart failure are mostly not related to the evidence-based ACC/AHA in-hospital process indicators for heart failure. It is unclear whether in-hospital quality of care is the key determinate of the readmission rate or whether readmissions are likely influenced more by postdischarge care. Further research is needed to clarify whether the readmission rate is a reflection of hospital care or quality of care on a larger level, especially when it is used for a pay-forperformance scheme to measure quality of hospital care.

BACKGROUND

In recent years, progressively sophisticated efforts have been undertaken to increase the transparency and accountability of the quality of hospital care. Hospital readmission rates are increasingly receiving attention, as they result in high health care costs[1] and as they are thought to represent quality of hospital care.[2]

For some years, hospital risk-adjusted readmission rates have also been used for policy purposes.[3] The Hospital Readmissions Reduction Program was set up by the Centers for Medicare and Medicaid Services (CMS) as a pay-for-performance scheme in which hospitals with an excessive readmission rate, compared with the national mean, are financially penalized by being paid less.[4,5]

Heart failure, for which readmissions are common, is one condition included in this CMS pay-for-performance scheme.[6] Furthermore, heart failure forms a growing public health problem in the United States, due to its high prevalence,[7] high mortality risk,[8] and high costs.[1] It is estimated that $39.2 billion was spent on care for patients with heart failure in the United States in 2010.[9]

The use of risk-standardized readmission rates in pay-for-performance schemes is based on the theory that hospitals that deliver low-quality hospital care have higher readmission rates. In the scientific literature, the debate regarding whether or not the rate of readmission is ready to be used for pay-for-performance schemes is still ongoing because there are several concerns about the validity of risk-standardized readmission rates as a measure of hospital care quality.[10,11]

If hospitals' readmission rates are a valid indicator of the quality of delivered hospital care, based on Donabedian's classic structure-process-outcome framework,[12] they are expected to correlate with other hospital quality indicators.

The American College of Cardiology/American Heart Association (ACC/AHA) has developed a core set of evidence-based in-hospital heart failure process measures (Table I).[13,14] For example, the beneficial effect of the administration of angiotensin-converting enzyme inhibitor (ACEI), on outcome was shown in randomized controlled trials.[15,16]

The aim of this study is to evaluate whether the evidence-based ACC/AHA process indicators are related to the readmission rate. Therefore,we performed a systematic review to assess the scientific evidence on the relation between readmission rates and quality of hospital care for patients with heart failure,which was defined as adherence to the ACC/AHA process indicators for heart failure.

Table I. ACC/AHA process indicators based on the guidelines for the diagnosis and management of HF

Performance measure name	Description
ACEI or ARB for LVSD	HF patients with LVSD and without both ACEI andARB contraindicationswho are prescribed an ACEI or ARB at hospital discharge
Evaluation of LVS function	HF patients with documentation in the hospital record that LVS function was assessed before arrival, and during hospitalization, or is planned after discharge
Adult smoking cessation advice/counseling	HF patients with a history of smoking cigarettes, who are given smoking cessation advice or counseling during hospital stay
Discharge instructions	HF patients discharged home with written instructions or educational material given to patient or caregiver at discharge or during the hospital stay addressing all of the following: activity level, diet, discharge medications, follow-up appointment, weight monitoring, and what to do if symptoms worsen
Anticoagulant at discharge for HF patients with AF	HF patients with chronic/recurrent AF and without warfarin contraindications who are prescribed warfarin at discharge

Abbreviations: HF, Heart failure; LVSD, left ventricular systolic dysfunction; LVS, left ventricular systolic; AF, atrial fibrillation.

METHODS

We carried out a systematic literature review on the relation between the ACC/AHA in-hospital process indicators for heart failure and readmission rates. This article was drafted following the Preferred Reporting Items for Systematic Reviews and Meta-Analysis checklist[17] (Appendix A).

Data sources and search strategy

With the help of a medical information specialist, 2 researchers (C.F. and H.L.) developed and carried out a broad systematic computerized literature search in the bibliographic databases Embase, Medline OvidSP, Web of Science, Cochrane Central, and PubMed. No time restrictions were set. All articles must be published in English, and they were included until 14th April 2014. MeSH search terms were related to hospital quality indicators, quality control procedures, medical audit, measurement, and readmission/hospitalization. The fullsearch strategy, which was adapted for the several databases searched, can be found in Appendix B. Furthermore, the reference lists of relevant articles were screened for additional articles. In addition, a free literature search was conducted.

Selection process

First, the titles and abstracts of all identified articles were screened by one reviewer (C.F.). Of the articles possibly eligible for inclusion, full texts were obtained. The full text was screened by 2 reviewers independently (H.L., C.F.), using a screening form. The interrater reliability was measured by calculating the reviewers' agreement percentage on the inclusion of the eligible full-text articles. In case of disagreement, studies were discussed among the 2 researchers. If no agreement could be reached, a third researcher was consulted (E.S.).

Eligibility criteria

The selected articles had to meet specific criteria. They had to be peer reviewed. The patient population under investigation had to be patients with heart failure. Studies had to investigate the relation between the readmission rate and at least one of the heart failure hospital process indicators described by the ACC/AHA core set 2005 (Table I).[14] Studies with various outcome definitions, disease-specific or all-cause readmission, and time intervals between leaving the hospital and readmission from 28 days, 30 days, or up to 1 year were considered. Randomized controlled trials that evaluated the process measures were not included because we aimed to test the indicators in the context where performance measurement takes place,

namely, in the real-world setting, without randomization, perfect data quality, blinded outcome assessment, and others.

Data abstraction

From each included study, we collected descriptive information, including authors' names, publication year, study design, and data source used in the study. Furthermore, the information on the patient population, like the number of centers and patients and specific diagnoses included was assessed. To be able to compare the studies, judge their quality, and interpret results, we abstracted information on the methodology used in the studies. The definition of the outcome measure was assessed, including type of readmission (disease-specific, all-cause readmission, planned/unplanned) and time interval between leaving the hospital and readmission. We assessed how the authors handled the competing outcome of mortality. Regarding data reliability and methodology, we assessed whether the studies applied case-mix correction and whether the studies provided information on the data they used (eg, regarding the completeness and accuracy and by whom and in which way coding was done). We also checked whether readmissions to hospitals other than the one where the index admission took placewere captured in the data. In case of doubt during the data abstraction, cases were discussed among the 2 researchers (C.F., H.L.). If doubt persisted or if information was missing, we contacted the authors of the original studies.

Data analysis

Of the studies providing effect estimates, we pooled the results using random-effect meta-analysis. Results for each of the process measures were pooled separately, and also studies analyzing readmissions as a dichotomous outcome (resulting in odds ratios [ORs]) were pooled separately from those analyzing time to readmission (resulting in hazard ratios [HRs]). Studies providing other numbers, such as correlation coefficients or percentages of adherence across quartiles of outcome,were not included in themeta-analysis but described in a qualitative way. Moreover, studies that used a combined outcome (eg, death or readmission) from which readmissions could not be deducted and studies that used a composite quality measure instead of the separate process measures were excluded from the meta-analysis. Some studies analyzed nonadherence vs adherence to the process indicators. In that case, we transformed the OR or HR presented into the OR or HR for adherence vs nonadherence using the formula: 1/OR or 1/HR. Cochran Q test was used to assess the homogeneity across the studies, and it was quantified by the I^2 statistic, which reflects the amount of

heterogeneity that is attributable to the variation across studies, rather than chance.[18] Meta-analyses are presented as forest plots createdwith the metaphor package in R statistics version 3.0.1 (The R Foundation for Statistical Computing, Vienna, Austria).

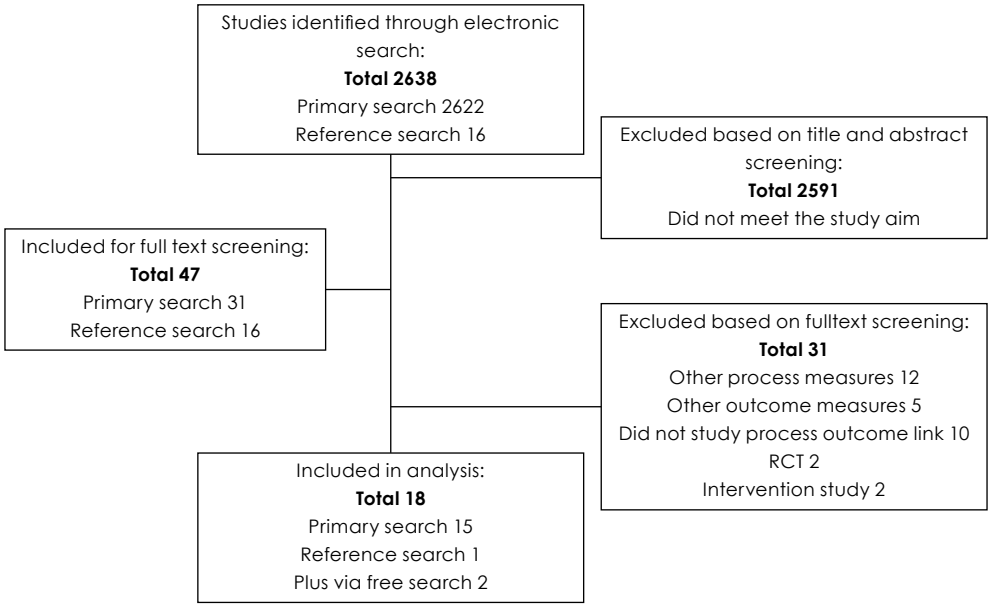
Funding
No extramural funding was sought or used to support this work. Wichor Bramer, Medical Information Specialist at the Library of the Erasmus MC University Medical Center, helped to develop the search strategy.

RESULTS

Search result

Figure 1 shows that our search strategy yielded 2,638 articles, 2,622 through the initial search and 16 by reference review. Of those, 2,591 articles were excluded after title and abstract review. The full text of the resulting 47 articles was screened, which resulted in exclusion of another 31 articles. Twelve studies were excluded because they measured other process indicators than those recommended by the ACC/AHA. Five studies were excluded because they considered other outcome indicators than the rate of readmission. Ten studies focused on the ACC/AHA process and outcome indicators, but did not investigate the link between them. Another 4 studies were excluded because they used the indicators in an randomized controlled trial or intervention study setting. This resulted in 16 articles eligible for inclusion. The free search resulted in 2 additionally included articles. Thus, 18 studies were included in our final analysis. The interrater agreement for the selection process was 89%.

Figure 1. Flowchart of article inclusion process



Description of the studies included

Of the 18 included studies, 15 were conducted in the United States.[19–33] Five studies were single center,[19,24,28,31,32] and the maximum number of centers was 3,655.[27] The number of analyzed patients ranged from 239[31] to 328,830[29] across the studies. Most studies (n = 14) were based on retrospective data (Table II). Seven studies defined adherence to the process indicators on the hospital level (as percentage adherence), whereas the other 11 studies defined adherence on the individual patient level (adherence to ACC/AHA process indicators per patient: yes/no).

Half of the included studies (n = 9) considered a time interval of leaving the hospital and readmission of around 30 days.

Most of the studies included readmissions for all causes, whereas only 6 studies defined cause-specific readmissions as an outcome.[20,24,26,32,34,35] Case-mix correction was executed in all but one study.[25] A number of studies stated that they excluded in-hospital deaths, transferred patients, or patients who left the hospital against medical advice from the denominator. However, most of the studies did not provide information on such selections (Table II). In total, 8 studies could not be included in the meta-analysis and were only described qualitatively because they used combined outcomes and/or combined process measures.

Associations between process indicators and readmissions

Angiotensin-converting enzymeinhibitor/angiotensin receptor blocker use

The strongest evidence was found for an association between the use of ACE/angiotensin receptor blocker (ARB) blocker in eligible patients and the rate of readmission. Ten studies tested the association, of which 8 could be included in the pooled analysis. The pooled OR for the effect of administrating ACEI or ARB blocker on readmissions was 0.7 (95% CI 0.4-1.23). The pooled HR indicated the same trend (HR 0.8, 95% CI 0.71-1.06), despite large heterogeneity across the studies. In general, administrating ACEI or ARB blocker seems to be associated with lower readmission rates.

Evaluation of left ventricular systolic function

The evidence for a relation between "evaluation of left ventricular systolic function" and the rate of readmission wasmixed. Of 9 studies that tested the association, 6 studies were pooled in the meta-analysis.[25–28,31,36] The pooled OR for the effect of evaluating the left ventricular systolic function on readmissions was 0.98 (95% CI 0.72-1.34) (Figure 2). The meta-analysis of time to readmission even indicated an adverse effect of adhering to the process indicator (HR 1.04, 95% CI 1.01-1.07) (Figure 3). However, for both meta-analyses, the I^2 indicated a high level of heterogeneity across the studies. Two of the 3 studies that could not be included in the meta-analysis indicated a significant association between the adherence to this indicator and the rate of readmission.[30] Nevertheless, the overall evidence for this indicatorwas very heterogeneous, as supported by the fact that 2 studies reported a significant negative effect of adhering to the indicator on readmission.[21,26]

Smoking cessation counseling

Limited evidence was found for a relation between smoking cessation counseling and readmissions. Six studies tested the association between providing smoking cessation counseling and the rate of readmission. Of those, 4 studies could be included in the meta-analysis. The pooled effects estimates showed no effect of providing smoking cessation counseling on (time to) readmission (OR 0.99 [95% CI 0.98-0.99], HR 0.99 [95% CI 0.99-1.00]). This effect was very consistent over the studies, as shown by the low I^2. Also, the 2 studies not included showed no significant association between the smoking cessation counseling and readmission.[21,33]

Discharge instructions

Limited evidence was found for an association between the indicator "providing discharge instructions" and readmissions. A total of 9 studies tested the association, of which 6 could be included in the meta-analysis. The pooledOR showed no effect of providing discharge instructions on readmission rates (OR 1.0, 95% CI 0.98-1.01) (Figure 2). The pooled HR showed a positive but nonsignificant effect of providing discharge instructions on time to readmission (HR 0.93, 95% CI 0.78-1.01) (Figure 2). The I^2 indicated a low level of heterogeneity across the studies. The 3 studies not included in the pooled estimate of the OR had similar findings.

Table II. Descriptive information and methodological aspects of included studies

			Author, year, country	
			Chung et al, 2008 USA,[19]	
Descriptive information	Study type		Case-control study	
	Data source		Inpatient medical records	
	Centers		1	
	Patients		400	
	Level of definition of process measures		Patient	
	Included patient population		Primary diagnosis of HF	
Methodological aspects	Indicator definition	Type of readmission	Time to all-cause readmission	
		Time frame	Within 6 mo	
	Effect of competing outcomes	Association with mortality	Patients were right-censored for readmission/death	
	Data reliability	Case-mix adjustment	+	
		Readmissions to other hospitals	+	
		Coding		
		Completeness and accuracy of data source	Medical records were assessed by the quality assurance department in accordance with CMS standards	
	Reported data		HR	

	Author, year, country		
	Fonarow et al, 2007, USA[33]	Hernandez et al, 2010, USA[20]	Hernandez et al, 2011, USA,[21]
	Prospective cohort study	Retrospective cohort study	Retrospective cohort study
	Web-based registry optimize HF	OPTIMIZE-HF linked to Medicare claims data	GWTG-HF registry linked to in-patient claims data from CMS
	91	141	176
	5791	20,441	19,952
	Patient	Hospital	Hospital
	Primary diagnosis of HF, hypertensive heart disease with HF, or hypertensive heart and renal disease with HF; Age, >18 y discharged home	Discharge diagnosis of HF	Admission for episode of worsening HF or developed significant HF symptoms during hospitalization with HF as primary discharge diagnosis; *ICD-9-cm* 428.x, 402.x1, 404.x1, 404.x3
	All-cause rehospitalization combined	Time to cardiovascular readmission	/
	60-90 d	60-d and 1 y	30 d
	Death included in outcome	In-hospital deaths excluded	In-hospital deaths excluded
	+	+	+
	+	+	+
	Admission staff, medical staff, or both recorded race/ethnicity, usually as the patient was registered	Web-based case report form; system used automatic electronic data checks to prevent out-ofrange and duplicate entries, an audit of the database used pre specified criteria to verify the data against source documents for a 5% random sample of the first 10000 patients	Data were collected via a Web-based registry
	Participating hospitals supplied data on eligible admissions according to established JCAHO methods; The protocol was approved by each participating center's institutional review board or through use of a central institutional review board; Excluded patients discharged to skilled nursing facilities or other acute-care hospitals	The institutional review board approved the protocol; Patients transferred to another acute care hospital, discharged to hospice or a federal hospital, or left the hospital against medical advice excluded	Hospitals with b15 cases were excluded; contraindications or other medical exceptions for that therapy, as well as those documented to be comfort care only excluded; Discharge from the hospital with palliative care or hospice care also excludes a patient from each discharge process measure; For outcome measures, transfer-outs were also excluded
	OR	HR	Correlation coefficient

			Author, year, country	
			Jha et al, 2009, USA,[22]	
Descriptive information	Study type		Retrospective cohort study	
	Data source		HQA data linked to AHA annual survey	
	Centers		4031	
	Patients		/	
	Level of definition of process measures		Hospital	
	Included patient population		Patients with CHF	
Methodological aspects	Indicator definition	Type of readmission	Time to all-cause readmissions for CHF	
		Time frame	30 d	
	Effect of competing outcomes	Association with mortality	In-hospital deaths excluded	
	Data reliability			
		Case-mix adjustment Readmissions to other hospitals	+ +	
		Coding		
		Completeness and accuracy of data source	Excluded hospitalreferral regions with fewer than 100 patients admitted for pneumonia or congestive HF in the year under evaluation; Included all patients discharged from hospitals in each hospital-referral region during the first 11 mo of the year under evaluation and calculated the proportion of patients who were readmitted within 30 d	
	Reported data		Variation across percentiles of readmission/adherence	

	Author, year, country			
	Kociol et al, 2013, USA,[23]	Luthi et al, 2004, Switzerland[36]	Mansi et al, 2010, USA,[24]	Mazimba et al, 2013, USA,[25]
	Retrospective cohort study	Retrospective cohort study	Retrospective cohort study	Retrospective cohort study
	AHA's GetWith The Guidelines-HF Registry linked to Medicare claims data	Medical records electronic medical records hospital administrative data	Recorded data for quality-management administration	Medical records
	149	3	1	4
	173,871	1055	357	6063
	Hospital	Patient	Patient	Patient
	Primary diagnosis of HF; Age, >65 y	Primary or secondary diagnosis of *ICD-10* codes: 150.0, 150.1, 150.9, 111.0, 113.0, 113.2	Primary diagnosis of HF*	Primary diagnosis of HF by *ICD-9* codes designated for CHF; Age, >18 y
	Time to re-hospitalization	All-cause	Time to readmissions (only admissions for HF exacerbation)	/
	30 d	30 d	90 d	30 d
	In-hospital deaths	In-hospital deaths excluded	In-hospital deaths excluded	/
	excluded			-
	+ /	+ -	+ -	-
	Hospitals submit clinical information using an online, interactive case-report form	The data were abstracted from copies of hospital medical records by trained nurses and/or medical record specialists	Data abstraction was conducted by trained doctors/medical record specialists. In 2 hospitals the medical records were available for data abstraction. In the third, only the electronic medical records were available	
	Excluded transfers and excluded extremely low-volume hospitals with b15 cases	Interrater reliability was assessed in one hospital by a random replicate sample of 100 medical records which were reabstracted; Data were abstracted from the medical records of the index hospitalization†	Excluded patients: who were not expected to have follow-up visits within the system, such as residents of other states, or patients who belonged to a different health care system and were in the institution on an emergency or temporary basis	A trained nurse abstractor prospectively collected data with an integrated electronic medical record system
	Variation across percentiles of readmission/ adherence	OR	OR	OR

			Author, year, country	
			Patterson et al, 2010, USA,[26]	
Descriptive information	Study type		Retrospective cohort study	
	Data source		OPTIMIZE HF registry	
	Centers		150	
	Patients		22,750	
	Level of definition of process measures		Hospital	
	Included patient population		Primary diagnosis of HF: DRG codes[§]	
Methodological aspects	Indicator definition	Type of readmission	Time to cardiovascular readmission	
		Time frame	At 1 y	
	Effect of competing outcomes	Association with mortality	In-hospital deaths excluded	
	Data reliability	Case-mix adjustment	+	
		Readmissions to other hospitals	/	
		Coding	Automatic electronic data checks prevented out-of-range entries and duplicates	
		Completeness and accuracy of data source	In addition, an audit of the database based on predetermined criteria verified data against source documents for a 5% random sample of the first 10000 patients; Excluded transfers or subsequent admissions for rehabilitations	
	Reported data		HR	

	Author, year, country		
	Schopfer et al, 2012, USA[27]	Shahian et a., 2012, USA[28]	Stefan et al, 2012, USA[29]
	Retrospective cohort study	Observational retrospective cohort study	Cross-sectional analyses
	Hospital Compare database	Hospital data linked to administrative data with unique patient identifier	Quality Improvement Organization Clinical Data Warehouse
	3655	1	Varied between 374 and 3020?
	/	2773	328,830
	Hospital‡	Patient	Hospital
	Primary diagnosis of HF	Discharge diagnosis of HF; Eligible for at least one National Hospital Quality Measure	Discharge diagnosis of HF; Enrolled in fee-for-service Medicare; Age, ≥66 y; Discharged from an acute care; Hospital that reported data to the Hospital IQR Program
	/	Time to all-cause readmission	Time to any admission
	30 d	90 d and 1 y	Within 30 d
		In-hospital deaths excluded	In-hospital deaths excluded
	+	+	+
	/	+	-
	Hospitals with fewer than 25 patients with HF or incomplete data were excluded		Used discharge and quality of care assessment records that were submitted by hospitals that participated in the Hospital Inpatient Quality Reporting Program
		For the HF discharge instruction measure, we limited the study cohort to patients who were discharged to home with or without services, thus mitigating any potential confounding by the care rendered at extended care facilities; Excluded patients who left hospital against medical advice	Transfers to/from another acute care hospital were assigned to the hospital that discharged the patient; Extreme low volume hospitals further excluded
	OR	OR	Correlation coefficient

			Author, year, country	
			Sueta et al, 2000, USA[30]	
Descriptive information	Study type		Retrospective cohort study	
	Data source		Patient medical record, North Carolina Medicare Beneficiary File/North Carolina Medicare Claims History File	
	Centers		48	
	Patients		1195	
	Level of definition of process measures		Patient	
	Included patient population		Discharge diagnosis *ICD-9* codes: 402, 404, 425, 428; Age, ≥65 y	
Methodological aspects	Indicator definition	Type of readmission	Readmission or mortality	
		Time frame	30 d	
	Effect of competing outcomes	Association with mortality	In-hospital deaths excluded	
	Data reliability	Case-mix adjustment	+	
		Readmissions to other hospitals	+	
		Coding	Hospitals with fewer than 50 patients with HF or incomplete data were excluded. Limitation: most patients did not have documentation of left ventricular function, although most had radiographic evidence of cardiomegaly or CHF	
		Completeness and accuracy of data source	Patients admitted with an acute or recent myocardial infarction were excluded	
	Reported data		HR	

	Author, year, country		
	VanSuch et al, 2006, USA[32]	Whittaker et al, 2013, USA[31]	Yoo et al, 2014, Korea[35]
	Retrospective cohort study	Retrospective cohort study	Retrospective observational study
	Administrative/clinical data	Electronic medical record review, data acquired from the hospital billing system	Survey data
	1	1	23
	782	239	1297
	Patient	Patient	Patient
	Primary diagnosis of HF: 428, 402, 404; Age, ≥18 y	Primary diagnosis of HF (based on 24 *ICD*-9 codes); Age >18 y	Admission to hospital with systolic HF or dyspnea and verification of HF by clinical findings; Age, ≥20 y
	Time to readmissions, due to HF related and unrelated causes (2 analysis)	All-cause	Time to admission by aggravated HF
	Up to 90 d	30 d	Within 1 y
	In-hospital deaths censored	In-hospital deaths excluded	In-hospital deaths excluded
	+	+	+
	-	-	/
	Registered nurses reviewed the medical records to determine the extent of compliance with the documentation of written discharge instructions		Data were collected and managed by the Control of Data Committee of the study
	Patients included when they were discharged to home, home care or home care with intravenous treatment; Patients discharged to skilled nursing facilities or other acute care hospitals were excluded; Could not distinguish between planned and unplanned readmissions	Patients were excluded if they had a diagnosis of HF and received a LV assist device or a heart transplant and/or had a length of stay greater than 120 d; Furthermore, patients were excluded if they were transferred to another hospital, or left against medical advice	Excluded patients with inadequate echocardiographic and clinical data
	HR	OR	HR

			Author, year, country
			Youn et al, 2012, Korea[34]
Descriptive information	Study type		Nationwide, prospective observational, multicenter study
	Data source		KorHF registry
	Centers		24
	Patients		1527
	Level of definition of process measures		Patient
	Included patient population		HF symptoms as the primary reason for admission and signs of congestion such as pulmonary congestion or systematic edema
Methodological aspects	Indicator definition	Type of readmission	Diagnose-specific
		Time frame	60 d, 1 y
	Effect of competing outcomes	Association with mortality	Deaths within 1 y excluded
	Data reliability	Case-mix adjustment	+
		Readmissions to other hospitals	-
		Coding	Data were collected and managed by the Control of Data Committee of the Korean Heart Failure (KorHF) Registry
		Completeness and accuracy of data source	Excluded 17 patients with no follow up data; No distinction between planned/unplanned readmissions
	Reported data		OR

Abbreviations: GWTG-HF, GetWith the Guidelines–Heart Failure; CMS, Centres for Medicare and Medicaid Service; HQA, Hospital Quality Alliance; HF, heart failure, CHF, congestive heart failure

*Codes 402.01, 402.11, 402.91, 404.01, 404.03, 404.11, 404.13, 404.91, 404.93, 428.0, 428.20 to 428.33, 428.40 to 428.43, and 428.9.

†Three centers included in the studymention that in the third hospital, the quality of the administrative datawas not so good and therefore just a minor number of patients could be included; excluded patients, if they had left the hospital againstmedical advice, were transferred to another acute care facility, were discharged in 1998, or had secondary HF due to valvular heart disease, acute myocardial infarction, cor pulmonale, chronic renal failure, hyperthyroidosis, thiamine deficiency, amyloidosis, or chronic obstructive lung disease treated with oxygen.

‡Readmission rate defined on hospital level as well .§Codes 104, 112, 115-118, 121-125, 127-145, 476, 514-518, 525-527, 535-536, 547-558.

Figure 2. Meta-analysis of studies investigating the relationship between heart failure process indicators and rate of readmission (ORs).

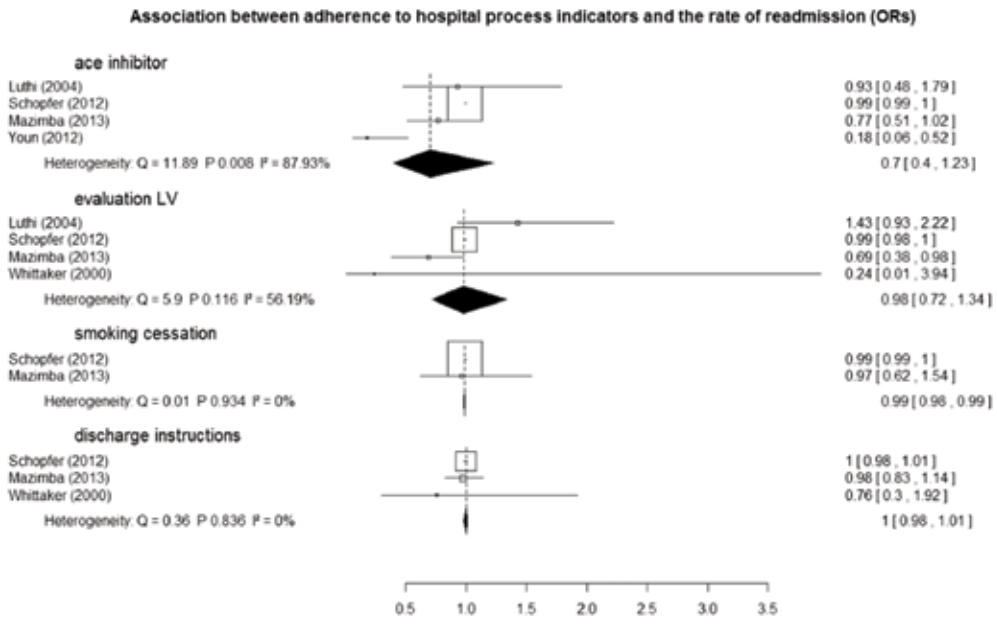


Figure 3. Meta-analysis of studies investigating the relationship between heart failure process indicators and time to readmission (HRs).

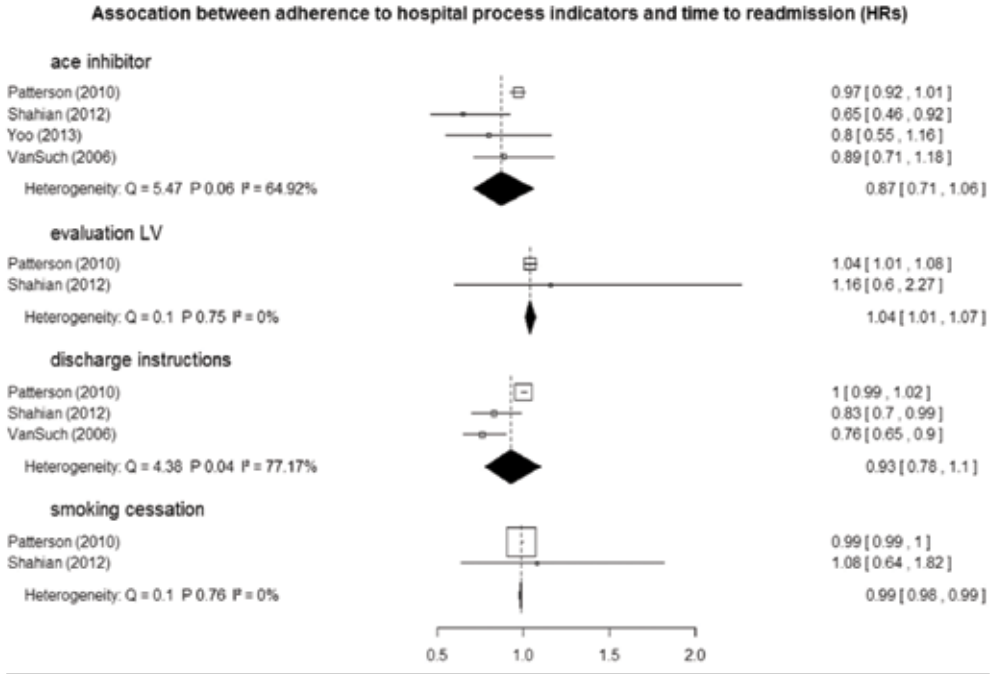


Table III. Association between adherence to hospital process indicators and the rate of readmission

	HF process indicators					
	ACEI, or ARB, for LVSD	Evaluation of LVS function	Smoking cessation counseling	Discharge instructions	AC at discharge for HF patients with AF	Composite
ORs						
Luthi et al[36] Mazimba et al[25] Schopfer et al[27] Youn et al[34] Whittaker et al[31]	0.93 (0.48-1.79) 0.77 (0.51-1.02) 0.99 (0.99-1.00) 0.18 (0.06-0.52) 0.24 (0.01-3.94)	1.43 (0.93-2.22) 0.69 (0.38-0.98) 0.99 (0.98-1.00) 0.76 (0.30-1.92)	 0.97 (0.62-1.54) 0.99 (0.99-1.00) 	 0.98 (0.83-1.14) 1.00 (0.98-1.01) 	1.61 (0.59-4.35) 	
HRs						
Hernandez 2010 Patterson 2010 Shahian 2012 VanSuch 2006 Yoo 2014	 0.97 (0.92-1.01) 0.65 (0.46-0.92) 0.89 (0.71-1.18) 0.80 (0.55-1.16)	 1.04 (1.01-1.08) 1.16 (0.6-2.27) 	 0.99 (0.94-1.04) 1.08 (0.64-1.82) 	 1.0 (0.99-1.02) 0.83 (0.7-0.99) 0.76 (0.65-0.9) 	1.02 (0.9-1.06) 	
ORs combined outcomes*						
Fonarow et al[33] Sueta et al[30]	0.51 (0.34-0.78)	1.06 (0.81-1.38) 0.70 (0.53-0.93)	0.74 (0.50-1.09)	1.07 (0.89-1.28)	0.83 (0.64-1.09)	
Other outcome measures						
Hernandez et al[21] Jha et al[22]	0.03 (0.68)	0.25 (0.002)†	0.06 (0.45)	0.04 (0.63) Highest compliance hospitals had nearly same rrs‡ as ones with lowest (23.7% vs 23.5%; P = .54)	−0.16 (0.05)	
Composite score						
Stefan et al[29] Chung et al[19] Mansi et al[24] Kociol et al[23]						−0.02 (0.22)†,§ HR 0.74 (0.57-0.97)§ OR 2.82 (1.46-5.44)§ Hospitals in highest rr quartile had best compliance: 75% in longest rr quartile vs 70% in shortest rr quartile (P b .0001)§
	4/10	3/9	0/6	2/9	0/4	1/4

* Combined readmission and mortality.

† Correlation coefficient (P value).

‡ Readmission rates.

§ Including indicators 1, 2, 4, and 5,

Warfarin for atrial fibrillation

No evidence was found for an association between administrating warfarin for atrial fibrillation and readmission. Of 4 studies testing the association, none could be included in the meta-analysis as they used different outcome measures. None of the studies found a significant association between administrating warfarin and readmissions.[20,21,33,36]

Composite measures

Four studies did not test single indicators, but instead included several of the process measures in a composite score. Of these 4, Chung et al[19] was the only study that found a significantly lower hazard of readmission when adhering to the process indicators (HR 0.74, 95% CI 0.57-0.97). In contrast, Kociol et al[23] and Mansi et al[24] both found a significant increase in readmission with greater adherence to composites of process indicators (Table III).

DISCUSSION

In this systematic review, we assessed the scientific evidence regarding the relation between readmission rates and quality of hospital care for patients with heart failure as measured by the ACC/AHA in-hospital process indicators. We found that readmission rates are mostly not related to these evidence-based process indicators.

There are several possible explanations for the fact that the studies included in our review did not find an association between adherence to process indicators and readmission rates:

Indicator definition
In most studies, no distinction between planned and unplanned readmissions was made, whereas only unplanned admissions are possibly related to quality of care.[37] In addition, the process measures themselves may not adequately capture the quality of care that was actually received by the patient. For example, "providing smoking cessation counseling" and "providing discharge instructions" do not assess the quality of the process but merely whether it has been executed. Furthermore, even when the smoking cessation counseling was executed perfectly and the patient stopped smoking, it is doubtful this could decrease the risk of readmissionwithin 30 days. Some process indicators require patient compliance in order to be effective. For example, the process indicator "ACEI/ARB blocker prescription" does not measure whether the patient actually takes the medication in the prescribed dosages. Furthermore, some process indicators are only applicable to part of the population, such as warfarin for patients with atrial fibrillation.

Insufficient case-mix correction
Case-mix correction is of major importance when investigating any outcome indicator as patients are not randomly distributed over hospitals, and process measures are not randomly assigned to patients within hospitals. Two patient aspects seem to increase the risk of being readmitted: being very sick and being very poor.[37–40] Nevertheless, disease severity and socioeconomic status are seldom taken into account because information regarding these aspects is not available in the administrative data that were used in most studies.

Some studies defined the process measures on the patient level and some on the hospital level, as percentage adherence. Measuring on the patient

level may result in confounding by indication because the sicker patients get the treatment, and without sufficient case-mix adjustment, we see a negative treatment effect.

Disregard competing outcomes
There is a relation between readmission rates and mortality rates per hospital,[41] as patients who died during their index admission cannot be readmitted anymore. Therefore, patients who have died should be excluded from the denominator for the readmission rate. This was done in most, but not all, of the included studies.

In summary, we found that readmission rates are mostly not related to the evidence-based ACC/AHA in-hospital process indicators. This may be partly due to shortcomings of the studies included in our review, but there are also suggestions that the quality of the in-hospital processes is just not—or weakly—reflected in the readmission rate.

Implications
Because heart failure readmissions are a heavy burden on the health care system, it is important to evaluate the quality of the care related to such readmissions. Readmission rates are used in many countries as a quality indicator for hospital care, which assumes that they are determined by quality of hospital care, that is, hospital processes and structures. Currently, it is, however, not known which hospital processes and structures exactly determine readmissions. Furthermore, although readmissions may, to some extent, be influenced by the quality of care received in the hospital, this is certainly not the only (nor perhaps the most significant) contributing factor. Patient characteristics, patients' life circumstances, and the nature of posthospital care have also been shown to affect readmissions.[42–44,1,45] Because the influence of these factors may vary between hospitals, they may be the underlying reason for outlier readmissions in hospitals.

External reporting and including readmissions in pay for performance schemes without knowing their cause have unintended consequences. When hospitals, which seem to perform badly in regard to their readmission rate, are punished without knowing where the high rates stem from, they may face unjustified punished, for example, for treating high-risk patients. Furthermore, when hospitals do not know which processes they have to improve in order to reduce their high readmission rates but at the same time fear the financial consequences of having high readmission rates, gaming may be stimulated. Therefore, as long as readmissions related to quality of

hospital care cannot be identified and robust scientific evidence underpinning the association between quality of hospital care and readmissions is lacking, readmission rates should be used for pay for performance purposes only with great caution.

A limitation of our study is that for some of the meta-analyses, only few studies could be included. This caused findings for some indicators to be determined by one large study. However, the limited number of studies itself represents the limited evidence base for the use of readmission rates as an indicator of quality of hospital care. Although we have only studied ACC/AHA process indicators, it is possible that other processes, such as β-blockers, spironolactone or adequate diuresis, and scheduling and executing postdischarge follow-ups, are more strongly related to readmissions.

In conclusion, readmission rates after heart failure are mostly not related to the evidence-based ACC/AHA in-hospital process indicators for heart failure. It is unclear whether in-hospital quality of care is the key determinate of the readmission rate or whether readmissions are influenced more by postdischarge care. Further research is needed to clarify whether the readmission rate is a reflection of quality of hospital care or quality of care in general. Such research is especially important because this rate is used in pay-for-performance schemes to measure quality of hospital care.

REFERENCES

1. Jencks SF, Williams MV, Coleman EA. Rehospitalizations among patients in the Medicare fee-for-service program. N Engl J Med 2009;360(14):1418-28.
2. Ashton CM, Kuykendall DH, Johnson ML, et al. The association between the quality of inpatient care and early readmission. Ann Intern Med 1995;122(6):415-21.
3. Kocher RP, Adashi EY. Hospital readmissions and the Affordable Care Act: paying for coordinated quality care. JAMA 2011;306(16): 1794-5.
4. CMS Centers for Medicare & Medicaid Services. Readmissions Reduction Program. Available from: http://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Readmissions-Reduction-Program.html. [Accessed 26.04.2015].
5. Joynt KE, Jha AK. A path forward on Medicare readmissions. N Engl J Med 2013;368(13):1175-7.
6. Axon RN, Williams MV. Hospital readmission as an accountability measure. JAMA 2011;305(5):504-5.
7. Lloyd-Jones D, Adams R, Carnethon M, et al. Heart disease and stroke statistics—2009 update a report from the American Heart Association Statistics Committee and Stroke Statistics Subcommittee. Circulation 2009;119(3):480-6.
8. Jong P, Vowinckel E, Liu PP, et al. Prognosis and determinants of survival in patients newly hospitalized for heart failure: a population-based study. Arch Intern Med 2002;162(15):1689-94.
9. Lloyd-Jones D, Adams RJ, Brown TM, et al. Heart disease and stroke statistics—2010 update. A report from the American Heart Association. Circulation 2010;121(7):e46-215.
10. Hernandez AF, Curtis LH. Minding the gap between efforts to reduce readmissions and disparities. JAMA 2011;305(7):715-6.
11. Davidson G, Moscovice I, Remus D. Hospital size, uncertainty, and pay-for-performance. Health Care Financ Rev 2007;29(1):45-57. [Fall].
12. Donabedian A. Evaluating the quality of medical care. Milbank Mem Fund Q 1966;44(3):166-206.
13. Bonow RO, Bennett S, Casey DE, et al. ACC/AHA clinical performance measures for adults with chronic heart failure. A report of the American College of Cardiology/American Heart Association Task Force on Performance Measures (writing committee to develop heart failure clinical performance measures): endorsed by the Heart Failure Society of America. Circulation 2005;112(12):1853-87.
14. Bonow RO, Bennett S, Casey Jr DE, et al. ACC/AHA clinical performance measures for adults with chronic heart failure: a report of the American College of Cardiology/American Heart Association Task Force on Performance Measures (writing committee to develop heart failure clinical performance measures) endorsed by the Heart Failure Society of America. J Am Coll Cardiol 2005;46(6):1144-78.
15. Flather MD, Yusuf S, Kober L, et al. Long-term ACE-inhibitor therapy in patients with heart failure or left-ventricular dysfunction: a systematic overview of data from individual patients. ACE-Inhibitor Myocardial Infarction Collaborative Group. Lancet 2000;355(9215):1575-81.
16. Garg R, Yusuf S. Overview of randomized trials of angiotensin-converting enzyme inhibitors on mortality and morbidity in patients with heart failure. Collaborative Group on ACE Inhibitor Trials. JAMA 1995;273(18):1450-6.
17. Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann Intern Med 2009;151(4):264-9.
18. Higgins JP, Thompson SG, Deeks JJ, et al. Measuring inconsistency in meta-analyses. BMJ 2003;327(7414):557-60.
19. Chung ES, Guo L, Casey Jr DE, et al. Relationship of a quality measure composite to clinical outcomes for patients with heart failure. Am J Med Qual 2008;23(3):168-75.
20. Hernandez AF, Hammill BG, Peterson ED, et al. Relationships between emerging measures of heart failure processes of care and clinical outcomes. Am Heart J 2010;159(3):406-13.
21. Hernandez AF, Fonarow GC, Liang L, et al. The need for multiple measures of hospital quality: results from the Get With the Guidelines– Heart Failure Registry of the American Heart Association. Circulation 2011;124(6):712-9.
22. Jha AK, Orav EJ, Epstein AM. Public reporting of discharge planning and rates of readmissions. N Engl J Med 2009;361(27):2637-45.

23. Kociol RD, Liang L, HernandezAF, et al.Are we targeting the rightmetric for heart failure? Comparison of hospital 30-day readmission rates and total episode of care inpatient days. Am Heart J 2013;165(6):987-994.e1.
24. Mansi IA, Shi R, Khan M, et al. Effect of compliance with quality performance measures for heart failure on clinical outcomes in high-risk patients. J Natl Med Assoc 2010;102(10): 898-905.
25. Mazimba S, Grant N, Parikh A, et al. Heart failure performance measures: do they have an impact on 30-day readmission rates? Am J Med Qual 2013;28(4):324-9.
26. Patterson ME, Hernandez AF, Hammill BG, et al. Process of care performance measures and long-term outcomes in patients hospitalized with heart failure. Med Care 2010;48(3):210-6.
27. Schopfer DW, Whooley MA, Stamos TD. Hospital compliance with performance measures and 30-day outcomes in patients with hear failure. Am Heart J 2012;164(1):80-6.
28. Shahian DM, Meyer GS, Mort E, et al. Association of National Hospital Quality Measure adherence with long-term mortality and readmissions. BMJ Qual Saf 2012;21(4):325-36.
29. Stefan MS, Pekow PS, Nsa W, et al. Hospital performance measures and 30-day readmission rates. J Gen Intern Med 2012:1-9.
30. Sueta CA, Schenck A, Chowdhury M, et al. Effect of angiotensin-converting enzyme inhibitor therapy on 30-day outcome in patient N or =65 years of age with chronic congestive heart failure. Am J Cardiol 2000;86(10):1151-3. [A9].
31. Whittaker BD, Soine LA, Errico KM. Patient and process factors associated with all-cause 30-day readmission among patients with heart failure. J Am Assoc Nurse Pract 2015;27(2):105-13.
32. VanSuch M, Naessens JM, Stroebel RJ, et al. Effect of discharge instructions on readmission of hospitalised patients with heart failure: do all of the Joint Commission on Accreditation of Healthcare Organizations heart failure core measures reflect better care? Qual Saf Health Care 2006;15(6):414-7.
33. Fonarow GC, Abraham WT, Albert NM, et al. Association between performance measures and clinical outcomes for patients hospitalized with heart failure. J Am Med Assoc 2007;297(1):61-70.
34. Youn YJ, Yoo BS, Lee JW, et al. Treatment performance measures affect clinical outcomes in patients with acute systolic heart failure: report from the Korean Heart Failure Registry. Circ J 2012;76(5): 1151-8.
35. Yoo BS, Oh J, Hong BK, et al. SUrvey of Guideline Adherence for Treatment of Systolic Heart Failure in Real World (SUGAR): a multi-center, retrospective, observational study. PLoS One 2014;9(1): e86596.
36. Luthi JC, Burnand B, McClellan WM, et al. Is readmission to hospital an indicator of poor process of care for patients with heart failure? Qual Saf Health Care 2004;13(1):46-51.
37. Fischer C, Lingsma HF, Marang-van de Mheen PJ, et al. Is the readmission rate a valid quality indicator? A review of the evidence. PLoS One 2014;9(11):e112282.
38. van Walraven C, Dhalla IA, Bell C, et al. Derivation and validation of an index to predict early death or unplanned readmission after discharge from hospital to the community. CMAJ 2010;182(6): 551-7.
39. Donze J, Aujesky D, Williams D, et al. Potentially avoidable 30-day hospital readmissions in medical patients: derivation and validation of a prediction model. JAMA Intern Med 2013;173(8): 632-8.
40. Heidenreich PA, Sahay A, Kapoor JR, et al. Divergent trends in survival and readmission following a hospitalization for heart failure in the Veterans Affairs health care system 2002 to 2006. J Am Coll Cardiol 2010;56(5):362-8.
41. Gorodeski EZ, Starling RC, Blackstone EH. Are all readmissions bad readmissions? N Engl J Med 2010;363(3):297-8.
42. Forster AJ, van Walraven C. The use of quality indicators to promote accountability in health care: the good, the bad, and the ugly. Open Medicine 2012;6(2):e75-9.
43. Kangovi S, Grande D. Hospital readmissions—not just a measure of quality. JAMA 2011;306(16):1796-7.
44. Berry JG, Hall DE, Kuo DZ, et al. Hospital utilization and characteristics of patients experiencing recurrent readmissions within children's hospitals. JAMA 2011;305(7):682-90.
45. Allaudeen N, Vidyarthi A, Maselli J, et al. Redefining readmission risk factors for general medicine patients. J Hosp Med 2011;6(2):54-60.

APPENDIX

Appendix A. PRISMA checklist

Section/Topic	No.	Checklist item	Reported on page no.
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	99
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	101
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	102
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	103
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (eg, Web address), and, if available, provide registration information including registration number.	/
Eligibility criteria	6	Specify study characteristics (eg, PICOS, length of follow-up) and report characteristics (eg, years considered, language, publication status) used as criteria for eligibility, giving rationale.	105, 110-118
Information sources	7	Describe all information sources (eg, databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	104, 105
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	129
Study selection	9	State the process for selecting studies (ie, screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	104, 105
Data collection process	10	Describe method of data extraction from reports (eg, piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	104
Data items	11	List and define all variables for which data were sought (eg, PICOS, funding sources) and any assumptions and simplifications made.	104-106
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	/

Summary measures	13	State the principal summary measures (eg, risk ratio, difference in means).	105, 106
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (eg, I^2) for each meta-analysis.	105, 106
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (eg, publication bias, selective reporting within studies).	/
Additional analyses	16	Describe methods of additional analyses (eg, sensitivity or subgroup analyses, meta-regression), if done, indicating which were prespecified.	/
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	Figure 1
Study characteristics	18	For each study, present characteristics for which data were extracted (eg, study size, PICOS, follow-up period) and provide the citations.	Tables II and III
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item12).	/
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group and (b) effect estimates and confidence intervals, ideally with a forest plot.	Table III or plots
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	Plots
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see item 15).	NA
Additional analysis	23	Give results of additional analyses, if done (eg, sensitivity or subgroup analyses, meta-regression (see item 16).	/
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (eg, healthcare providers, users, and policy makers).	119
Limitations	25	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (eg, healthcare providers, users, and policy makers).	124
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	123, 124
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (eg, supply of data); role of funders for the systematic review.	NA

Appendix B. Search strategy

("clinical indicator"/de OR "performance measurement system"/exp OR "quality control procedures"/de OR "quality control"/de OR "medical audit"/de OR (((qualit* OR perform* OR safet* OR governance) NEAR/3 (indicat* OR measure* OR assessment* OR control* OR marker* OR metric*)) OR ((clinical OR medical) NEAR/3 (indicator* OR audit*))):ab,ti) AND ("hospital readmission"/de OR (readmiss* OR rehospital* OR ((re OR return) NEAR/3 (hospital* OR admiss*))):ab,ti).

Part III

Empirical studies on the reliability and validity of quality indicators for hospital care

Chapter 5.

The influence of data quality on the reliability of quality indicators for hospital care

Helen A Anema, Job Kievit, Claudia Fischer, Ewout W Steyerberg, Niek S Klazinga

This article was published as:
Influences of hospital information systems, indicator data collection and computation on reported Dutch hospital performance indicator scores. *BMC Health Services Research. 2013; 13:212.*

ABSTRACT

Background: For health care performance indicators (PIs) to be reliable, data underlying the PIs are required to be complete, accurate, consistent and reproducible. Given the lack of regulation of the data-systems used in the Netherlands, and the self-report based indicator scores, one would expect heterogeneity with respect to the data collection and the ways indicators are computed. This might affect the reliability and plausibility of the nationally reported scores.

Methods: We aimed to investigate the extent to which local hospital data collection and indicator computation strategies differ and how this affects the plausibility of self-reported indicator scores, using survey results of 42 hospitals and data of the Dutch national quality database.

Results: The data collection and indicator computation strategies of the hospitals were substantially heterogenic. Moreover, the Hip and Knee replacement PI scores can be regarded as largely implausible, which was, to a great extent, related to a limited (computerized) data registry. In contrast, Breast Cancer PI scores were more plausible, despite the incomplete data registry and limited data access. This might be explained by the role of the regional cancer centers that collect most of the indicator data for the national cancer registry, in a standardized manner. Hospitals can use cancer registry indicator scores to report to the government, instead of their own locally collected indicator scores.

Conclusions: Indicator developers, users and the scientific field need to focus more on the underlying (heterogenic) ways of data collection and conditional data infrastructures. Countries that have a liberal software market and are aiming to implement a self-report based performance indicator system to obtain health care transparency, should secure the accuracy and precision of the heath care data from which the PIs are calculated. Moreover, ongoing research and development of PIs and profound insight in the clinical practice of data registration is warranted.

INTRODUCTION

Monitoring the quality of health care by means of performance indicator scores is part and parcel of national health care systems. Performance indicators (PIs) are used to monitor and improve quality and patient safety and to stimulate accountability and market processes in countries worldwide (e.g. USA (www.ahrq.com), UK (www.hqip.org.uk), and Denmark (www.ikas.dk). To play this role effectively, performance indicators need to be reliable and valid measures of health care quality[1-3] particularly when hospitals' performances are ranked and published in the lay press[4] and/or used to link reimbursement to indicator results.[4,5]

National hospital performance indicator programs commonly use PIs that are selected on basis of expert judgment (e.g. medical doctors, patient organizations) and existing scientific evidence[6] about valid relations between health care processes and outcome indicator (e.g.[7]). These PIs have often been successfully implemented in other countries. Due to differences in national healthcare and local hospital organization, copying a performance indicator into another health system does not automatically imply a valid reflection of the underlying health care process that it is intended to measure. Although small pilot studies to assess data collection, indicator computation, data sending and data analyses are commonly executed prior to national PI implementation [6], thorough evaluations of reliability and validity of the PIs after implementation, are scarce.[8,9] Therefore it remains unclear whether such selected PIs can be validly used in the national health care system.

For a PI to be valid, it needs to be composed with the least possible measurement error. Most indicators consist of a numerator (e.g. number of patients that timely received antibiotics prophylaxis) and a denominator specifying the population at risk[10,11] (e.g. the number of patients that should receive the prophylaxis). To compose the numerator and denominator, patients that fit the inclusion criteria need to be identified in the data systems. This selection process is explained in instruction manuals that describe the specific steps that need to be taken. Each step yields a data element, for instance the date of surgery, or a secondary diagnosis (comorbidity). For some steps it is required to select several data elements and as such consist of a set of rules that combine several data elements, the latter increasing the complexity of the process.[12]

In many quality indicator programs, (e.g. in the USA: AHRQ, Kaiser Permanente, Veteran Affairs Quality program; Denmark: IKAS; Germany: BQS) the coordinating organizations are responsible for PI data collection and computation, as opposed to programs that rely on self-report of the participating hospitals. They abstract the indicator data from digital administrative (hospital information system) or financial databases using computerized data abstraction algorithms. This approach however, is effective only when the data-systems are identical for all participating hospitals. When hospitals do not have an integral electronic patient record, patient information is stored in several information systems. Moreover, when a country has a liberal software market (US, The Netherlands) and the PIs are based on self-report, coupling of these independent information systems in an attempt to automatically collect the data might be difficult and prone to error due to the various software environments. Although manually selecting the data from all the systems and paper records seems less error-prone, it is very time consuming. It could be assumed, therefore, that hospitals obtain their own, unique, strategy to compute the PI score, which makes comparison of the PI scores difficult. Therefore, we question whether the construction of Dutch hospital information systems that are used to compute PI scores and the extent to which data elements are available and accessible affect the accuracy and precision of the PI scores in a negative way.

To investigate this, we study the Dutch local hospital data-infrastructure, here defined as the availability and accessibility of PI data elements. In 2008, the Dutch government implemented hospital PIs for various disease specific groups (Dutch Health Care Transparency Program; DHTPa). All Dutch hospitals are required to report these PIs to the government on a yearly basis, to pursue public disclosure of health care performance, data benchmarking, selective contracting by insurance companies, and decision making processes of patients looking for a healthcare provider. The selection of the indicators was performed by disease specific expert groups of medical doctors, patient organizations, and health care insurers, on the basis of published guidelines and expert opinion. Attempts were made to standardize the structure, process and intermediate outcome indicator definitions, as well as the data collection and indicator computation instructions. These instructions are principally code based, that is, based upon diagnose and procedure codes such as the ICD-9 classification or DRG codes (Diagnosis Related Group codes). Dutch hospitals are independent organizations and free to choose information technology systems for clinical and administrative data. Therefore, developing instructions that are specified for the information system that is used to handle the data is not feasible.

Since Dutch hospitals are solely responsible for reporting the required indicator scores, reliability (precision and accuracy) of the self-reported PIs might be particularly at risk. Thus, given the lack of regulation of the data-infrastructure (information systems that are used) in the Netherlands, and the self-report based indicator scores, one would expect heterogeneity with respect to the data collection and the ways indicators are computed. Moreover, it can be expected that hospitals that have a low level of automation and difficulties to connect all the individual information systems, make use of data that are registered in external databases such as the National Cancer Registry of the Comprehensive Cancer Centers (in Dutch: IKNL; further referred to as CCC).

Together, we aimed to obtain insight in the precision and accuracy of publically available PI scores and the impact that the local hospital data-infrastructure has on these aspect using Dutch indicator scores of two sets of surgical PIs (Table 1 and Appendix 1): Hip and Knee Replacements (further referred to as HR or KR) and Breast Cancer (further referred to as BC). First, to obtain a general idea of the accuracy and precision and second, to investigate the long term pattern of a hospital's performance, we evaluated the plausibility of the available PI scores of health care delivered between 2008 and 2010. With plausibility we mean the extent to which the indicator score is in line with what we can expect from the health care procedure, on basis of the literature, guideline compliance and audit studies. For example, a process indicator score of 100% that measures the number of patients that timely received surgically related antibiotic prophylaxis might be implausible, as relevant literature on guideline compliance reveals average scores as low as 50%.[13-15] Second, using data obtained through a questionnaire among a sample of Dutch hospitals in 2010, we checked whether the data elements that are required to compute the indicator scores were registered at all.

Further, if available, we wanted to know whether these data elements were easy to access or only after time consuming actions, whether hospitals calculated the indicator on basis of the entire population at risk, or whether the PI score was merely estimated. And finally, we investigated the relation between the data-infrastructure (data availability and accessibility), the way hospitals calculated the indicator scores and the plausibility of the submitted indicator score.

METHODS

Study design

A cross-sectional mixed methods design, using both qualitative and quantitative data from three different sources was used to explore the effect of the data-infrastructure on the accuracy and the precision of the PI scores.

Study population

The study population consisted of national hospitals (in 2010) of which 24 were small hospitals (< 320 beds), 48 were intermediate (320–627 beds), and 28 were large hospitals (> 627 beds), 27 were teaching hospitals and

Table 1. Overview of process and outcome indicators hip and knee replacements and breast cancer

National Performance Indicators				
	Total Hip and Knee replacements *	S	P	O
2b	% of patients that was administered thrombosis prophylaxis for 6 weeks to 3 months post-surgery, in case of total hip or knee surgery		X	
4b	% of patients that did not (2008 & 2009)/ did (2010) receive a homologue blood transfusion, in case of total hip or knee surgery		X	
5b	% of patients that was administered antibiotics perioperatively		X	
5c	% of patients that was administered antibiotics 15 to 60 min. prior to surgery or to blood emptiness		X	
5d	% of patients with a deep wound infection after a total hip or knee replacement			X
	Breast Cancer **	S	P	O
1	% of patients who were seen by a breast cancer nurse specialist preoperatively		X	
2	% of patients that was reviewed preoperatively in a multi-disciplinary team meeting		X	
3	% of patients with a non-radical primary tumor resection			X
4	% of surgeons in the surgery department that perform surgical treatments of breast tumors	X		
5	% of patients that are operated within 4 weeks after the final lab results are known		X	
6a	% of patients with local recurrences within 5 years after breast-conserving surgery			X
6b	% of patients that have local recurrences within 5 years after ablative breast surgery			X
7	% of patients with a breast tumor that was postoperatively reviewed in a documented multi-disciplinary team meeting	X		

* Note: 5 yes/no 'Hip/Knee structure indicators' are omitted from the table as they were not included in the current study; ** Indicators 1, 2 and 7 were removed from the indicator set in 2009, 4 in 2011; S structure, P process, O intermediate outcome; The PIs consist of numerators and denominators that each are composed of several variables according to combinatory logic that is described in instruction manuals. See for details of numerators and denominators Appendix 1.

seven were university hospitals. A teaching hospital is a large hospital that is approved of training medical doctors, without being affiliated to a university.

In total 42 national hospitals (42%) gave informed consent to participate in the study and returned the questionnaire. This representative sample included five small (< 320 beds), 25 intermediate (320–627 beds) and nine large hospitals (> 627), 13 were teaching hospitals and two were university hospitals.

Data sources

DHTP performance indicator database (A)

In 2009, 2010 and 2011, approximately 100 Dutch hospitals submitted performance indicator scores (web-based entry tool), indicative of care delivered the year before (2008, 2009 and 2010 respectively) at the DHTP.

Additional reliability data DHTP (B)

Besides the indicator scores, the DHTP requires hospitals to upload information (self-report) regarding the reliability of the submitted indicator scores. One dichotomous item specifically targets how the hospital computed the indicator score, i.e. by means of an integral calculation of the total population at risk (further referred to as 'calculation'), or by merely estimating it (further referred to as 'estimation'). 'Estimating' implies either an extrapolation of scores based on a small sample or based on locally implemented protocols. This item was used for our question about the strategy that hospitals used to compose the indicator and was only available for the hospitals that were enrolled in the qualitative part of our study in 2010.

Web-based questionnaire (C)

The short web-based questionnaire targeted the hospital's local data infrastructure (collection of computer programs and databases) and was setup to answer our questions about the data infrastructure (what is the data availability and data accessibility?). Data availability was dichotomously assessed; a hospital confirmed whether the required information was registered somewhere in the hospital or not. Data accessibility was divided into three categories:

1)Automatically (Aut) accessible, 2) partly automatically accessible (Partly Aut), 3) or manually accessible (Man). Automatically accessible refers to data elements that are stored within a computer information system, can be easily reviewed ('only a few mouse clicks away') and can be abstracted

by means of computerized search algorithms (Queries). Partly automatically accessible refers to data elements that are available in electronic systems, can be reviewed easily, but cannot be abstracted by means of a computerized search algorithm as some manual actions are required. Manually accessible refers to data elements that are available but only through labor intense data handlings such as medical chart reviews.

The BC questionnaire differed from that of the hip and knee questionnaire on the accessibility items. In the Netherlands, the accessibility of performance indicators can be dependent upon external organizations. For example, the data of BC care is simultaneously collected by the Dutch CCC. Data-managers of these centers retrospectively visit a hospital and collect and register cancer specific information (tumor type, health status etc.) in the national cancer database (NKR). When uploading the DHTP indicators, hospitals can decide whether to use their own calculated indicator scores or those calculated by the CCC. As part of our interest in indicator composition strategies, we additionally added an item in the BC questionnaire about whether hospitals collected and calculated their own BC indicator scores, or whether indicator scores from the CCCs was used.

Analyses

In the current study we used the following variables of interest:

1. The plausibility of the PI score: plausible score (PS): a score larger than 0% and smaller than 100%; implausible score (IS): a score of 0% or 100%.
2. Data-availability scores: all required data elements available (A); not available (NA).
3. Data-accessibility scores: data elements automatically available, partly automatically available, manual available; and easy accessible (EA), difficult to access (DA).
4. Indicator computation scores: PI score based on integral calculation of population (Cal); estimation of PI score (Est); CCC calculation: indicator score calculated by the CCC (BC only).

To get an overall idea of the plausibility of the submitted national PI scores we first provide an overview of the characteristics (means and SDs) of the PI scores that are available in the DHTP database (Data source A) and judge the plausibility (that is, perfect performance of 100% or 0%) and compare the indicator scores with what could be expected on basis of the literature (qualitative approach). Secondly, to answer our question regarding the data infrastructure we analyses the web-questionnaire items (Data Source

C) and present frequency results of data-availability and data accessibility scores. Finally, to obtain information about the relation between the data infrastructure, the procedures that hospitals use to compose the PIs (obtained from Data Source B) and the resulting PI scores, we calculated 2 × 2 and 2 × 3 Chi-square tests. Results are presented separately for the Hip replacements indicator set and the Breast Cancer indicator set. As Hip and Knee replacements yield fairly similar scores (see Table 2 and Figure 1B), we present the data of HR only.

Table 2. Results of the descriptive statistics of hip and knee replacements and Mammacare

		2008				2009				2010				
	PI	N	M	SD	Range	N	M	SD	Range	N	M	SD	Range	IS
HR	2b	64	99.92	0.602	95 –100	95	99.8	0.949	93 – 100	94	99.9	0.30	98 – 100	53
	4b	52	91.27	22.79	0 – 100	91	90.6	15.24	0 - 100	93	16.13	26.58	0 – 100	6
	5b	65	100	0.000	100 - 100	96	99.7	1.402	93 - 100	94	101.0	15.65	70 – 100	53
	5c	59	97.38	14.00	0 – 100	94	98.0	5.827	66 - 100	93	98.9	16.98	64 – 100	37
	5d	60	0.816	0.740	0 – 2.7	93	0.719	0.674	0 – 2.75	93	0.754	0.804	0 – 4	5
GM		60	78	8	/	94	78	5	/	93	63	12	/	37
KR	2b	63	99.92	0.663	95 - 100	94	99.8	0.834	93 - 100	93	100	0.246	98 – 100	52
	4b	54	91.17	25.72	0 - 100	89	95.6	10.99	0 - 100	92	11.65	27.15	0 – 100	7
	5b	64	100	0.00	100 - 100	95	99.8	1.101	92 - 100	93	99.6	2.342	78 – 100	52
	5c	59	96.84	15.71	0 - 100	93	97.8	6.52	60 - 100	92	96.8	8.979	49 – 100	39
	5d	59	0.50	0.649	0 – 3	92	0.554	0.631	0 – 3.2	92	0.544	0.616	0 – 3.3	6
GM		60	78	9	/	93	79	4	/	92	62	8	/	37
BC	1	68	100	5.055	75 – 100	/	/	/	/	/	/	/	/	/
	2	68	100	3.200	85 – 100	/	/	/	/	/	/	/	/	/
	3	66	9.675	5.464	0 – 24	95	9.215	4.733	0 – 29	94	7.279	4.026	0.95 - 23	0
	4	68	41.4	12.68	10 - 75	95	38.5	11.53	10 - 60	/	/	/	/	/
	5	63	90.48	14.92	17 - 100	95	89.2	10.30	51 - 100	94	88.9	11.85	34 – 100	0
	6a	57	2.130	2.247	0 -11	89	1.748	1.945	0 - 9	93	1.490	1.703	0 – 8	0
	6b	57	2.700	2.838	0 - 11	90	2.581	2.522	0 - 11	93	2.455	2.351	0 - 10	0
	7	65	100	5.568	74 - 100	/	/	/	/	/	/	/	/	/
GM		64	56	6	/	93	28	6	/	94	25	5	/	0

Note: PI Performance Indicator, N number of hospitals, M mean, SD standard deviation, GM grand mean, IS Implausible Score (100% or 0% score three consecutive years), HR Hip replacements, KR Knee replacements, BC Breast Cancer; Number of hospitals vary slightly throughout the text due to differences in how hospitals are enrolled in the study (concerns vs separate hospitals). Note 2: No perfect score defined for Performance indicator BC 4.

RESULTS

Hip replacement indicators

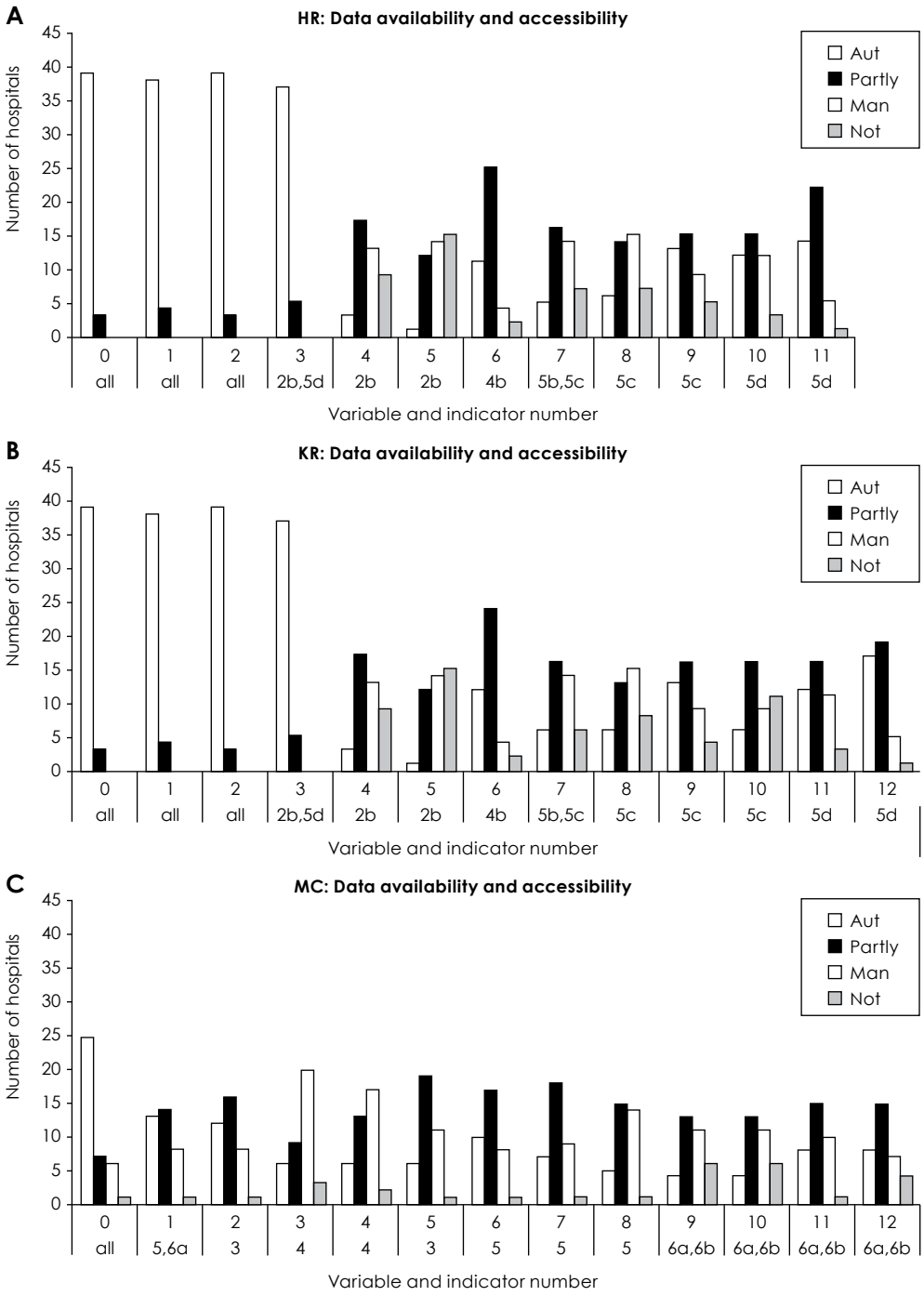
Data source A: Plausibility of the reported scores

All PIs have high averages (ceiling effect; see Table 2) and PIs 4b (2008, 2009, 2010) and 5c (2008) have large ranges, from 0% to 100%. A 0% compliance to blood transfusion guidelines can be regarded as implausible, being most likely an error. This might be explained by a change in the indicator definition of 4b in 2010; from 'no homologue blood transfusions' to 'homologue blood transfusions'. Homologue blood transfusions increase the risk of blood borne infections and thus need to be reduced (Dutch Institute for Health Care improvement CBO guideline hip and knee arthrosis 2007).[16] As described in the introduction, high averages of the PIs can be regarded as unrealistic, particularly when performance is consistently high for several consecutive years. We determined the number of 100% scores (and 0% scores for outcome indicator 5d; further referred to as 'implausible score') for each year and calculated the number of hospitals that maintained a perfect score for three consecutive years. Implausible scores were rather consistent for the indicators 2b, 5b and 5c (53, 53 and 37 times; see Table 2). Also, five hospitals scored an implausible 0% of post-operative deep wound infections in three consecutive years (outcome indicator 5d).

Data source C: Heterogeneity of reported local data infrastructure

The questionnaire revealed that, averaged over all indicators, 26 hospitals (62%) reported to have all the required data elements that are necessary for the calculation of the PI available at the time of study (Figure 1A). When at least one data element was missing (data element unavailable; n = 16), eight hospitals indicated to miss one or two data elements, four hospitals missed three or four data elements and four hospitals missed five or six data elements. Most of the HR data elements were on average automatically accessible (43%) or partly automatically accessible (30%), whereas 17% was only manually accessible (10% of the information was not available). The data elements that were most frequently indicated to be fully automatic accessible were necessary for computing the denominator (category ALL: elements: 0 patient identification number, 1 financial code hip replacement, 2 procedure code hip replacement, 3 date of hip replacement surgery), whereas the data elements that are less easily accessible are necessary for the numerators of the various indicators.

Figure 1. Reported data infrastructure of the orthopedic and oncology sets' as short and concise descriptive title of Figure 1ABC



Data source ABC: Relation between computation methods, data collection and PI score

To investigate the relation between the data availability and computation methods we first divided the hospitals in two separate categories; those who indicated to have at least one required data element for a certain indicator unavailable (NA = not available; in total 15 hospitals) and those who had all required data elements for an indicator available (A = available; in total 27 hospitals). Chi square tests revealed that the computation method that hospitals choose is significantly associated with the data availability. That is, when data were unavailable, indicator scores were on average more often estimated as compared to calculated (Est = 91%; Cal = 9%). When data were available this pattern was reversed (Est = 42%; Cal = 58%; $\chi 2$ (1) = 51.71, $p < 0.01$. However, nearly half of the indicator scores were estimated even when all the required data elements were available.

ABC: Reported data infrastructure of the orthopedic and oncology sets. AUT = fully automatic accessible, Partly = partly automatic, partly manually accessible, Man = manually accessible, NOT = not available; HR = Hip replacement, KR = Knee replacement, BC = Mammacarcinoma; Numbers 1, 2, 3 etc. = numbers that indicate the indicator variable which is part of the indicator set; 2b 4b 5b etc. = the unique number of the indicator.

A similar association might be found between data accessibility and the composition method. To test this association we divided the data set of our sub sample with all data elements Available (N = 27) in an 'Easy Access' (EA) and 'Difficult Access' (DA) category for each indicator separately. We assigned 1, 2 or 3 points to the accessibility scores Automatic, Partly Automatic and Manual and averaged the scores per indicator. An EA score for a certain indicator was obtained when the average accessibility score was below 1.5, or else it was given a DA score.

Table 3 revealed that the 27 hospitals reported to have all data available of in total 186 indicators (76 + 73 + 32 + 5). Most indicators were classified as 'Difficult to access' (76 + 73 = 149) as compared with the 'Easy access' category (32 + 5 = 37). Again Chi square tests revealed a significant relation between composition strategy and data accessibility as reported for the data availability $\chi 2$ (1) = 42.35, $p < 0.01$), but this relation was different as compared to the NA category. That is, when data was difficult to access an almost equal number of indicators were based upon integral calculations and estimations (Cal = 51%; Est = 49%). When data was easy to access, the percentage of scores that were based upon an estimation decreased considerably (Cal = 86%; Est = 5%).

Next, we tested the association between the data infrastructure of available data and the plausibility of the indicator scores. Overall more implausible (IV) than plausible scores (PV) were observed (A: IS = 65%, PS = 35%). However, easy accessible data yielded more plausible indicator scores (PS = 64%; IS = 38%) as compared to the difficult access indicators (PS = 27%; DA, IS = 73%; $\chi 2$ (1) = 24.64, $p < 0.01$).

Finally, to see whether the plausibility of the indicator scores was more associated to the composition strategy as compared to the data infrastructure we summated across all data-infrastructure categories (NA + DA + EA) all the indicators that were calculated and checked whether they were implausible or plausible scores and did the same for the estimated

Table 3. Number of Hip Replacement indicators with plausible, implausible and missing values for the separate indicator computation and data collection strategies

				2b	4b	5b	5c	5d	Total
NA		**CAL**		**2**	**0**	**0**	**0**	**1**	**3**
		EST		**12**	**2**	**7**	**9**	**0**	**30**
A	DA	CAL	Total	12	22	12	10	20	76
			IS	11	4	11	7	7	40
			PS	1	18	1	3	12	35
			MV	0	0	0	0	0	0
		EST	**Total**	**24**	**5**	**21**	**22**	**1**	**73**
			IS	23	3	20	21	1	68
			PS	1	2	1	1	0	5
			MV	0	0	0	0	0	0
	EA	**CAL**	**Total**	**1**	**11**	**4**	**3**	**13**	**32**
			IS	1	1	4	2	4	12
			PS	0	10	1	1	11	23
			MV	0	0	0	0	0	0
		EST	**Total**	**1**	**0**	**1**	**2**	**1**	**5**
			IS	1	0	0	1	0	2
			PS	0	0	0	1	1	2
			MV	0	0	0	0	0	0

Note: IS Implausible (100% or 0%) Score, PS Plausible Score (<100%), MV missing value, NA Not available, EA easy access category, DA Difficult Access category, Cal calculation, EST Estimation; All data in frequencies. Scores are frequencies at the level of indicators, and not at hospital level, to avoid excluding hospitals with missing values from the analysis.

indicators. The Chi square test revealed a strong significant association ($\chi 2$ (1) = 59.05, $p < 0.01$). Implausible indicator scores were more often estimated (91%) as compared to calculated (38%) whereas plausible scores (scores < 100%) were more often calculated (62%) as compared to estimated (9%).

Breast cancer indicators

Data source A: Plausibility of the reported scores

In contrast to the orthopedic data, implausible perfect scores in three consecutive years could not be identified.

Data source C: Heterogeneity of reported local data infrastructure

The questionnaire revealed that, averaged over all indicators, one third of the hospitals (13 out of 41) have all the required data elements available (Figure 1C). When having registered some of the required data elements (27 hospitals), eight hospitals indicated to miss one or two data elements, one hospital missed three data elements and another hospital reported to miss ten of the required twelve data elements. Most of the information was on average partly automatically accessible (39%), whereas 24% was automatically and 30% only manually accessible (7% was not available). The data

Table 4. Number of breast cancer care indicators for the separate indicator computation and data collection strategies

			1	2	3	4a	4b	Total
NA		**Total**	1	3	3	8	8	**23**
		OWN	0	3	0	0	0	**3**
		CCC	1	0	3	7	7	**18**
		MV	0	0	0	1	1	**2**
A	DA	**Total**	28	32	28	29	28	**145**
		OWN	16	28	14	3	4	**65**
		CCC	12	4	13	25	23	**77**
		MV	0	0	1	1	1	**3**
	EA	**Total**	8	4	7	5	5	**29**
		OWN	4	2	3	1	1	**11**
		CCC	4	1	4	4	4	**17**
		MV	0	1	0	0	0	**1**

Note: OWN Self-calculated, CCC calculated by Collective Cancer Center, MV missing value, NA Not available, EA easy access category, DA Difficult Access category, Cal calculation, Est Estimation; Scores are frequencies at the level of indicators, and not at hospital level, to avoid excluding hospital with missing values from the analysis.

element that was indicated as automatically accessible the most frequent (element 0: patient identification number) was necessary for computing the denominator.

Data source ABC: Relation between computation methods, data collection and PI score

In contrast to the results of the orthopedic indicator sets, only one hospital indicated to have estimated the indicator score. Therefore we focused on the hospitals choice to use the CCC to compose the indicator score, or to use own scores. As can be observed in Table 4 (total column) most indicator scores were, on average, based upon CCC data when data was not available (NA; CCC = 18 indicators, OWN= 3 indicators). To test whether the choice to use the CCC data was associated with the accessibility of available data we determined for each indicator and access category separately how often a hospital used the CCC score and tested the association with a 2 × 2 (DA/EA vs. OWN, CCC) chi square test. The chi square test revealed that the accessibility did not seem to influence the choice for the CCC (DA: OWN= 46%, CCC = 54%; EA: OWN= 39%, CCC =61%; $\chi 2$ (1) = 0.74, p = 0.39). As was described above, implausible perfect scores on the breast cancer indicators are scarce. Therefore we did not further investigate the effect of the data infrastructure and composition strategy on the indicator score.

DISCUSSION

Summary

The current study revealed that the Dutch PI hospital data infrastructure is heterogeneous, and the reported performance data under investigation can be regarded as largely implausible, particularly those of the Hip and Knee replacement indicators. Moreover, in both cases, only few data elements were 'one mouse click away' (poorly accessible), indicating a large amount of labor to extract all the required data from stand-alone computers and (paper) medical records. In case of automatic availability, manual collection was still necessary to complete the computation. Together with the overall under reporting of the required data elements, this leads to more implausible, estimated scores in the HKR case. That is, when HKR data was unavailable or difficult to access, hospitals did not withdraw from submitting indicator scores but estimated their indicator score to be 100%. In contrast, the CCC employees, who have access to items that are unavailable for the hospitals, (for instance for the data elements of 'percentage of patients with recurrences within 5 years after surgery') covered most of this labor intensive work for the BC PIs, and therefore, that data is less implausible. It has to be noted here, that many hospitals preferred to use their own indicator scores, when available, instead of that of the CCC even if the data was poorly accessible. As hospitals are free to choose which scores to be uploaded, we conclude that for both indicator sets the heterogeneous data collection and indicator computation largely affects the comparability of hospital performance.

Policy implications

A governance model that increasingly relies on performance information as the basis for policy decisions (e.g. directly through selective contracting and indirectly through transparency of performance of care providers in the media), assumes the existence of high quality, reliable and valid performance information. Interestingly, this study has shown that the accuracy and precision of the PIs is questionable and further improvement of the current local hospital data-infrastructure in the Netherlands is necessary. There are several bottlenecks that need to be dealt with, ranging from the patchwork of hospital information systems, to the lack of a data-quality feedback loop back to the government.

Our results suggest that a nationally organized registry (in the case of breast cancer) led to more plausible indicator scores. Having one entity responsible for the data collection and indicator computation increases the

comparability of the hospitals performance scores. Within the hip and knee replacement care, a fully operational and nationally coordinated medical registry does not yet exist. Therefore, hospitals are entirely dependent on their own local data infrastructure. As a result, many hospitals choose to upload indicator scores that are estimated on basis of locally implemented treatment protocols and not on basis of empirical observations, neither of the entire population, nor of a representative sample. In the latter case, one would expect lower indicator scores (score < 100%), as the timely administration of antibiotic prophylaxis does not often exceed 70%.[13-15] Estimating the indicator score instead of withdrawing from reporting altogether might be explained by the experienced external pressure caused by for example ranking lists that are published by the lay press. Hospitals end up at the bottom of the ranking when no indicator score is available, hence reporting 100% might be considered a favorable strategy to prevent reputation damage. Nevertheless, other factors such as a lack of priority or understaffing might additionally be at work as hospitals estimated their indicator score even though the data was available and reasonably accessible. The reporting of estimated data needs to be prevented as it results in an unrepresentative reflection for the quality of care delivered. Clinicians could for example set up a systematic peer review and consensus conference to discuss the PI scores before submitting them to the public database.

Despite the positive effect of the National Cancer Registry (NCR) on the Breast Cancer indicator plausibility, more profound standardization of these processes remains warranted. Particularly as governmental (DHTP) regulation regarding the data sources and the software systems that should be used for data collection and indicator computation is still lacking. This allows hospitals to choose their own strategies, which decreases the comparability of performance between hospitals. An alternative solution might be provided by disease specific registries that appear to be effective in improving health care quality and reducing costs, through publically available outcomes of health care.[17]

Recently, several diagnoses specific medical registries have been set up in the Netherlands (e.g. Dutch Surgical Colorectal Audit).[18] Such a system avoids problems which arise when combining different data sources such as administrative data that can be easily calculated, with those based on other specific internal sources that can be easily manipulated. A drawback, however, is that a registry is set up from a unilateral, health care professional approach. The Netherlands has chosen to develop performance indicators according to a consensus driven perspective, implying that e.g. patient orga-

nizations and health care insurers are involved in the indicator selection and development process. Nevertheless this perspective might have led to a situation that methodological arguments for indicator selection and refinement lost too easily from political arguments to reach consensus on indicators that are supported by a broad selection of stakeholders. A consensus approach might benefit from more regulation with respect to the data quality, for instance by developing a data quality control framework that encompasses the most crucial steps of prevention, detection and actions to be taken with respect to insufficient data quality,[19] particularly when using administrative data systems. It has been suggested that administrative data alone is not always appropriate for the valid computation of performance measures.[10,20] Moreover, a language formalization process of all the relevant items during the indicator development phase seems vital as in the Netherlands every hospital collects its own data, has its own local data infrastructure, and DHTP has no insight in the underlying data that hospitals submit.[21,22] After such a formalization process it could additionally be suggested to improve already existing national databases such as the Dutch Hospital Discharge registry or financial databases that are hosted by health care insurance companies.

Finally, the consensus approach entails that the indicators are used for several goals such as benchmarking performance, pay for performance schemes, selective contracting by insurance companies, and decision-making processes of patients looking for a healthcare provider. Particularly in the case of self-reported data, it should be made clear which indicators can be used for which specific goal.

Conclusion

Our study provided insight in how performance indicator scores can be affected by heterogeneity of hospital information systems, data collection and data computation methods; factors that influence the reliability. Therefore, indicator developers, users and the scientific field need to focus more on the complexity of health care measurement instruments and conditional data infrastructures. Countries that have a liberal software market and are aiming to implement a self-report based performance indicator system to obtain health care transparency, should secure the accuracy and precision of the heath care data from which the PIs are calculated from. Moreover, ongoing research and development of PIs and profound insight in the clinical practice of data registration is warranted.

Endnotes

a.In Dutch: Zichtbare Zorg.

b.According to the CCMO (Central Committee on Research involving Human Subjects), no medical-ethical approval of the study was necessary.

Abbreviations

PI: Performance indicator; HKR: Hip and knee replacements; HR: Hip replacements; KR: Knee replacements; DHTP: Dutch health care transparency program; CCC: Comprehensive cancer centers; CCMO: Central committee on research involving human subjects.

REFERENCES

1. Mainz J: Defining and classifying clinical indicators for quality improvement. Int J Qual Health Care 2003, 15(6):523–530.
2. Williams SC, Watt A, Schmaltz SP, Koss RG, Loeb JM: Assessing the reliability of standardized performance indicators. Int J Qual Health Care 2006, 18(3):246–255.
3. Wollersheim H, Hermens R, Hulscher M, Braspenning J, Ouwens M, Schouten J, Marres H, Dijkstra R, Grol R, et al: Clinical indicators: development and applications. Neth J Med 2007, 65(1):15–22.
4. van Dishoeck AM, Lingsma HF, Mackenbach JP, Steyerberg EW: Random variation and rankability of hospitals using outcome indicators. BMJ Qual Saf 2011, 20(10):869–874.
5. Egol A, Shander A, Kirkland L, Wall MH, Dorman T, Dasta J, Bagwell S, Kaufman D, Matthews P Jr, Greenwald BM, Herr DL, Stavish C, Thompson C, Fahy BG, Society of Critical Care Medicine, et al: Pay for performance in critical care: an executive summary of the position paper by the Society of Critical Care Medicine. Crit Care Med 2009, 37(9):2625–2631.
6. Groene O, Skau JK, Frolich A: An international review of projects on hospital performance assessment. Int J Qual Health Care 2008, 20(3):162–171.
7. Ogbu UC, Westert GP, Slobbe LC, Stronks K, Arah OA: A multifaceted look at time of admission and its impact on case-fatality among a cohort of ischaemic stroke patients. J Neurol Neurosurg Psychiatry 2011, 82(1):8–13.
8. Fischer C, Anema HA, Klazinga NS: The validity of indicators for assessing quality of care: a review of the European literature on hospital readmission rate. Eur J Public Health 2011, 22(4):484–491.
9. Pronovost PJ, Goeschel CA: Viewing health care delivery as science: challenges, benefits, and policy implications. Health Serv Res 2010, 45(5 Pt 2):1508–1522.
10. Booth JL, Collopy BT: A national clinical indicator database: issues of reliability and validity. Aust Health Rev 1997, 20(4):84–95.
11. Pringle M, Wilson T, Grol R: Measuring "goodness" in individuals and healthcare systems. BMJ 2002, 325(7366):704–707.
12. Huff ED: Comprehensive reliability assessment and comparison of quality indicators and their components. J Clin Epidemiol 1997, 50(12):1395–1404.
13. Lundine KM, Nelson S, Buckley R, Putnis S, Duffy PJ: Adherence to perioperative antibiotic prophylaxis among orthopedic trauma patients. Can J Surg 2010, 53(6):367–372.
14. van Kasteren ME, Kullberg BJ, de Boer AS, Mintjes-de GJ, Gyssens IC: Adherence to local hospital guidelines for surgical antimicrobial prophylaxis: a multicentre audit in Dutch hospitals. J Antimicrob Chemother 2003, 51(6):1389–1396.
15. Stefansdottir A, Robertsson O, Dahl A, Kiernan S, Gustafson P, Lidgren L: Inadequate timing of prophylactic antibiotics in orthopedic surgery. We can do better. Acta Orthop 2009, 80(6):633–638.
16. Dutch Orthopedic Society: Diagnosis and treatment of hip and knee arthrosis; 2007 (http://www.cbo.nl/Downloads/363/)
17. Larsson S, Lawyer P, Garellick G, Lindahl B, Lundström M: Use of 13 disease registries in 5 countries demonstrates the potential to use outcome data to improve health care's value. Health Affairs 2012(31):220–227.
18. van GW, Krijnen P, Lemmens VE, den DM, Putter H, van de Velde CJ: Quality assurance in rectal cancer treatment in the Netherlands: a catch up compared to colon cancer treatment. Eur J Surg Oncol 2010, 36(4):340–344
19. Arts DG, De Keizer NF, Scheffer GJ: Defining and improving data quality in medical registries: a literature review, case study, and generic framework. J Am Med Inform Assoc 2002, 9(6):600–611.
20. Iezzoni LI: Using administrative data to study persons with disabilities. Milbank Q 2002, 80(2):347–379.
21. Dentler K, Cornet R, Ten Teije A, De Keizer NF: Comparison of Reasoners for large Ontologies in the OWL 2 EL Profile. SEMANTIC WEB 2012, 2(2):71–87.
22. Medlock S, Opondo D, Eslami S, Askari M, Wierenga P, de Rooij SE, Abu-Hanna A, et al: LERM (Logical Elements Rule Method): a method for assessing and formalizing clinical rules for decision support. Int J Med Inform 2011, 80(4):286–295.

APPENDIX

Appendix 1. National Performance Indicators: numerators and denominators

	Total Hip and Knee replacements *	**S**	**P**	**O**
2b	% of patients that was administered thrombosis prophylaxis for 6 weeks to 3 months post-surgery, in case of total hip or knee surgery		X	
Num	Number of patients that received thrombosis prophylaxis for 6 weeks to 3 months post-total hip or knee surgery			
Den	Number of patients that underwent total knee or total hip replacement in a certain calendar year			
4b	% of patients that **did not** (2008 & 2009)/ **did** (2010) receive a homologue blood transfusion, in case of total hip or knee surgery		X	
Num	Number of patients that received (2008 & 2009)/ did not receive (2010) a homologue blood transfusion			
Den	Number of patients that underwent total knee or total hip replacement in a certain calendar year			
5b	% of patients that was administered antibiotics perioperatively		X	
Num	Number of patients that was administered antibiotic prophylaxis, peri-operatively.			
Den	Number of patients that underwent total knee or total hip replacement in a certain calendar year			
5c	% of patients that was administered antibiotics 15 to 60 min. prior to surgery or to blood emptiness		X	
Num	Number of patients that was administered antibiotics 15 to 60 min. prior to total hip or knee surgery or to blood emptiness			
Den	Number of patients that underwent total knee or total hip replacement in a certain calendar year			
5d	% of patients with a deep wound infection after a total hip or knee replacement			X
Num	Number of patients with a deep wound infection after a total hip or knee replacement			
Den	Number of patients that underwent a total knee or total hip replacement in a certain calendar year			

	Breast Cancer **	S	P	O
1	% patients who were seen by a breast cancer nurse specialist preoperatively		X	
Num	Number of patients with a breast tumor that had at least one preoperative meeting with a breast cancer nurse specialist			
Den	Total number of patients with primary surgery of a breast tumor			
2	% patients that was reviewed preoperatively in a multi-disciplinary team meeting		X	
Num	Number of patients with a breast tumor that was reviewed in a documented, multi-disciplinary team meeting, prior to any treatment			
Den	Total number of patients that was diagnosed with a breast tumor			
3	% patients with a non-radical primary tumor resection			X
Num	Number of patients with a non-radical, primary tumor resection (breast saving surgery)			
Den	Total number of patients with a primary tumor resection (breast saving surgery)			
4	% surgeons in the surgery department that perform surgical treatments of breast tumors	X		
Num	Number of surgeons in department that perform surgical treatments of breast tumors			
Den	Total number of surgeons in department			
5	% patients that are operated within 4 weeks after the final lab results are known		X	
Num	Number of patients that are operated within 4 weeks after the final lab results are known			
Den	Total number of patients with a primary tumor resection			
6a	% patients with local recurrences within 5 years after breast-conserving surgery			X
Num	Number of patients with local recurrences within 5 years after breast-conserving surgery, primarily treated in own center (no referral)			
Den	Total number of patients with breast conserving therapy, primarily treated in own center (no referral)			
6b	% patients that have local recurrences within 5 years after ablative breast surgery			X
Num	Number of patients with local recurrences within 5 years after ablative breast surgery, primarily treated in own center (no referral)			
Den	Total number of patients with ablative breast surgery, primarily treated in own center (no referral)			
7	% of patients with a breast tumor that was postoperatively reviewed in a documented multi-disciplinary team meeting	X		
Num	Number of patients that was postoperatively reviewed in a documented multi-disciplinary team meeting			
Den	Total number of patients with breast surgery			

* Note: 5 yes/no 'Hip/Knee structure indicators' are omitted from the table as they were not included in the current study; ** Indicators 1,2 and 7 were removed from the indicator set in 2009, 4 in 2011; S = structure, P = process, O = intermediate outcome; The PIs consist of numerators and denominators that each are composed of several variables according to combinatory logic that is described in instruction manuals.

Chapter 6.

The influence of data quality on the correlations between quality indicators for hospital care

Claudia Fischer, Hester F Lingsma, Helen A Anema, Job Kievit, Ewout W Steyerberg, Niek S Klazinga

This article is submitted.
Testing the construct validity of hospital care quality indicators: A case-study on hip replacement.

ABSTRACT

Background: Quality indicators are increasingly used to measure quality of care and compare quality across hospitals. In the Netherlands over the past few years numerous hospital quality indicators have been developed and reported. Dutch indicators are mainly based on expert consensus and face validity and little is known about their construct validity. Therefore, we aim to study the construct validity of a set of national hospital quality indicators for hip replacements.

Methods: We used the scores of 100 Dutch hospitals on national hospital quality indicators looking at care delivered over a two year period. We assessed construct validity by correlating structure, process and outcome indicators using chi-square statistics, bootstrapped Spearman correlations, and independent sample t-tests. We studied indicators that are expected to correlate as they measure the same clinical construct.

Results: Among the 28 hypothesized correlations, three correlations were significant in the direction hypothesized. Hospitals with low scores on wound infections had high scores on scheduling postoperative appointments (meeting scheduled = 0.01; no meeting scheduled = 0.02; p-value = 0.001) and high scores on not transfusing homologous blood (correlation coefficient = -0.28; p-value = 0.05). Hospitals with high scores on scheduling complication meetings also had high scores on providing thrombosis prophylaxis (correlation coefficient = 0.21; p-value = 0.04).

Conclusions: Despite the face validity of hospital quality indicators for hip replacement, construct validity seems to be limited. Although the individual indicators might be valid and actionable, drawing overall conclusions based on the whole indicator set should be done carefully, as construct validity could not be established. Factors that may explain the lack of construct validity are poor data quality, no adjustment for case-mix and statistical uncertainty.

Key words: hip replacement, database, health care quality, quality indicators, validity

BACKGROUND

As quality improvement becomes a central tenet of health care, quality indicators (QIs) are becoming increasingly important. Quality is monitored and publicly reported in order to provide patients and health insurers with information regarding choices and to improve the quality of the underlying complex and resource-intensive care procedures [1].

For such purposes QIs need to be based on reliable data [2,3], and they must cover quality aspects on a structural, process, and outcome level [4]. The underlying assumption is that good structures of care increase the likelihood of good processes and good processes increase the likelihood of good outcomes (the Donabedian framework) [4]. Another important prerequisite for the external usefulness of the indicators and fair comparison of hospitals is that QIs are valid [5] and actionable. QIs need to provide insight into which factors determine the occurrence of an outcome, so that hospitals are able to act on the process to improve the outcome.

Total hip replacements are very interesting for quality of care research because hip replacements are common, elective procedures that are being performed more and more frequently [6]. Although the clinical and economic effectiveness of hip replacements is proven [7], it is still possible to observe variation in performance between providers [8,9]. As a result, these orthopaedic procedures have for instance been included in pay-for-performance schemes by social insurance programs such as Medicare and Medicaid [10]. In such a program hospitals are rewarded for meeting pre-defined performance targets related to the health care that is delivered [11]. In the pay-for-performance scheme of Medicare and Medicaid, the so-called 'Premier Quality Initiative Demonstration', a composite score was created from three measures of surgical process quality and three measures of surgical outcome. A performance bonus consisting of two percent of diagnosis-related group payments for total hip and knee arthroplasty was given to hospitals that scored in the top 10 percent on the composite measure [10]. For such external use (as well as for internal use such as in local hospital quality improvement), it is critical that indicators present a valid picture of the quality of the health care that is provided by a hospital [5]. However, empirical evaluations of the relation between outcome indicators and process and structure indicators that measure the same construct are scarce in Europe[12]. Even if quality indicators are tested in different health care systems, an evaluation in the health care system in which the indicator is used is essential. Differences in national health care and local hospital organization may influence the indi-

cator's validity [1]. Insight into the validity of QIs is particularly important when data reliability is at stake, for instance when there are no national standards that hospitals or database software providers should follow when setting up their in-hospital quality registries in which the quality data is entered [1,2]. This is the case in the Netherlands, where QIs were developed by the Dutch Health Care Transparency Program (DHTP) through a combination of expert consensus and available scientific literature. They were tested in only a few hospitals. Employees of the hospitals are required to calculate and report these QIs annually to the DHTP; public reporting and publication of these QIs has occurred for several subsequent years [13].

Therefore we aimed to evaluate several publicly available indicators of quality of hospital care in the Netherlands related to hip replacements (15 indicators) with regard to their construct validity, or the "degree to which an indicator measures what it claims to be measuring" [14]. In this study construct validity is operationalized by a significant correlation, in the expected direction, between two quality indicators that measure the same underlying construct.

METHODS

We conducted a cross-sectional data analysis, using quantitative data from two registration years (2008 and 2009) as reported by the hospitals.

QIs under investigation

The QIs we evaluated are all related to pre-operative and post-operative health care for hip replacements. We used data from two consecutive years. Table1 shows an overview of the definitions, numerators (i.e. number of patients who underwent a certain care process) and denominators (i.e. total number of patients) of the structure, process and outcome (S-P-O) QIs evaluated in this study. Moreover, it can be seen that the structure QIs in the hip replacement set are dichotomous (yes/no), whereas the majority of the process and outcome indicators are continuous measures (a proportion of patients with particular treatment or outcome).

Data source

Dutch Health Care Transparency Program Data (DHTP)

The QI data originate from a national database hosted by the DHTP [15]. Each year, Dutch hospital staff collect and submit to DHTP hospital-specific performance scores (numerators and denominators) for various diseases and interventions based on health care delivered in the preceding calendar year. In 2009 and 2010 hospitals were not obliged to report performance information about health care delivered in 2008 and 2009; as a result, not all hospitals participated. From 2010 onwards hospitals were obliged to submit the indicator scores. The hip replacement dataset included a total of ten structure indicators, four process indicators and one outcome indicator (Table 1). For our study we selected the available numerators and denominators for each hospital and indicator. All QI scores were aggregated on the hospital level.

Table 1. Included DHTP total hip replacement quality indicators

Total hip replacement			
Qi number	**Qi name**	**Indicator type****	**QI definition**
qi1	Preoperative patient information	S*	**Definition:** hospitals provide written or audio-visual preoperative patient information (yes/no)

Qi number	Qi name	Indicator type**	QI definition
qi2a	Guideline thrombosis prophylaxis	S	**Definition:** hospitals have a guideline or protocol on thrombosis prophylaxis for cases of hip replacement (yes/no)
qi2b	Thrombosis prophylaxis	P	**Definition and Numerator:** in hip replacement cases, the number of operations in which patients received medical thrombosis prophylaxis within 6 weeks and no more than 3 months after the operation
qi3a	Complication register	S	**Definition:** an automated information system is available to provide insight into the occurrence of complications (e.g. wound infection, lung emboli) within 6 weeks of HR (yes/no)
qi3b	Appointment within 6 weeks	S	**Definition:** to detect complications, a postoperative appointment is held within 6 weeks of a hip replacement (yes/no)
qi3c	Orthopaedic registration form	S	**Definition:** in hip replacement cases, an orthopaedic registration form is used to register complications (yes/no)
qi3d	Complications meeting	S	**Definition:** minuted meetings are held to discuss hip-replacement complications (number of meetings per year)
qi3e	Improvement plan	S	**Definition:** minuted meetings are held to discuss hip-replacement complications, if necessary an improvement plan with the person in charge is assigned (yes/no)
qi4a	Blood management guideline	S	**Definition:** a blood-management guideline or protocol to reduce perioperative administered in case of hip-replacement is present (yes/no)
qi4b	Transfusion of homologue blood	P	**Definition and Numerator:** in hip replacement cases, the number of operations in which patients did not receive transfusion of homologue blood
qi5a	Guideline for antibiotic prophylaxis	S	**Definition:** a guideline/protocol is available for antibiotic prophylaxis in the event of hip replacement (yes/no)
qi5b	Perioperative antibiotics	P	**Definition and Numerator:** in hip replacement cases, the number of operations in which perioperative antibiotics were administered
qi5c	Antibiotics 60 - 15 minutes	P	**Definition and Numerator:** in hip replacement cases, the number of operations in which patients received antibiotics 60 to 15 minutes before incision
qi5d	Wound infection	O	**Definition and Numerator:** in hip replacement cases, the number of patients with deep wound infections within 6 weeks of the operation
qi6	National prosthetic register	S	**Definition:** the hospital participates in the national arthroplasty register (yes/no)

* According to number in DHTP hip and knee replacement indicator set

** S = structure, P = process, O = outcome;

Table 2. Hypothesized indicator correlation and direction of association

	hypothesized indicator correlations	Evidence for expected indicator correlation	expected correlation direction	p-value indicator correlation strength 2008	2009
hc* 1	having a thrombosis prophylaxis management guideline (qi2a**) and the percentage of patients who accurately receive a thrombosis prophylaxis (qi2b)	(25)	positive	/	/
hc 2	having a blood management guideline (qi4a) and the percentage of patients who do not receive a blood transfusion (qi4b)	(26) (25)	positive	/	/
hc 3	having a guideline for antibiotic prophylaxis (qi5a) and the percentage of patients who receive antibiotic prophylaxis perioperative (qi5b)[25]	(25)	positive	/	/
hc 4	having a guideline for antibiotic prophylaxis (qi5a) and the percentage of patients who receive antibiotic prophylaxis 60-15 minutes before incision (qi5c)[25]	(25)	positive	/	/
hc 5	the percentage of patients who receive their perioperative antibiotic prophylaxis in a timely manner (qi5b) and the percentage of patients with deep wound infection (qi5d)[26-28]	(26-28)	negative	/	0.74
hc 6	the percentage of patients that receive antibiotic prophylaxis 60-15 minutes before incision (qi5c)and the percentage of patients with deep wound infection (qi5d)[26-28]	(26-28)	negative	0.14	0.74
hc 7	the percentage of patients who receive no blood transfusion (qi4b) and the percentage of patients with deep wound infection (qi5d) [29,30]	(29, 30)	negative	**0.05**	0.07
hc 8	having a timely postoperative appointment (q3b) and the percentage of patients with deep wound infections (qi5d)		negative	/	**0.001**
hc 9	having a complication register (qi3a) and providing a thrombosis prophylaxis (qi2b)		positive	0.73	0.19
hc 10	having a complication register (qi3a) and the percentage of patients receiving no blood transfusion (qi4b)		positive	0.09	0.57
hc 11	having a complication register (qi3a) and the percentage of patients receiving perioperative antibiotic prophylaxis (qi5b)		positive	/	0.60
hc 12	having a complication register (qi3a) and the percentage of patients receiving antibiotic prophylaxis 60-15 minutes before incision (qi5c)		positive	0.29	0.57
hc 13	having a complication register (qi3a) and the percentage of patients with deep wound infection (qi5d)		negative	0.74	0.43
hc 14	having an orthopaedic registration form (qi3c) and the percentage of patients receiving thrombosis prophylaxis (qi2b)	(31)	positive	0.80	0.89

hc 15	having an orthopaedic registration form (qi3c) and the percentage of patients receiving no blood transfusion (qi4b)	(31)	positive	0.98	0.26
hc 16	having an orthopaedic registration form (qi3c) and the percentage of patients receiving perioperative antibiotic prophylaxis (qi5b)	(31)	positive	/	0.06
hc 17	having an orthopaedic registration form (qi3c) and the percentage of patients receiving antibiotic prophylaxis 60-15 minutes before incision (qi5c) [31]	(31)	positive	/	0.28
hc 18	having an orthopaedic registration form (qi3c) and the percentage of patients with deep wound infections (qi5d) [31]	(31)	positive	0.60	0.42
hc 19	having complication meetings (qi3d) and the percentage of patients receiving thrombosis prophylaxis (qi2b)		positive	0.50	**0.04**
hc 20	having complication meetings (qi3d) and the percentage of patients receiving no blood transfusion (qi4b)		positive	0.26	0.91
hc 21	having complication meetings (qi3d) and the percentage of patients receiving perioperative antibiotic prophylaxis (qi5b)		positive	/	0.16
hc 22	having complication meetings (qi3d) and the percentage of patients receiving antibiotic prophylaxis 60-15 minutes before incision (qi5c)		positive	0.26	0.32
hc 23	having complication meetings (qi3d) and the percentage of patients with deep wound infections (qi5d)		negative	0.39	0.91
hc 24	having an improvement plan to avoid complications (qi3e) and the percentage of patients receiving thrombosis prophylaxis (qi2b)		positive	0.86	0.52
hc 25	having an improvement plan to avoid complications (qi3e) and the percentage of patients receiving no blood transfusion (qi4b)		positive	0.09	0.17
hc 26	having an improvement plan to avoid complications (qi3e) and the percentage of patients receiving perioperative antibiotic prophylaxis (qi5b)		positive	/	0.39
hc 27	having an improvement plan to avoid complications (qi3e) and the percentage of patients receiving antibiotic prophylaxis 60-15 minutes before incision (qi5c)		positive	0.51	0.05
hc 28	having an improvement plan to avoid complications (qi3e) and the percentage of patients with deep wound infections (qi5d)		negative	0.26	0.72

* hypothesized correlation (hc), ** quality indicator (qi)

Analysis

To describe the range in scores across hospitals we calculated the mean and interquartile range (IQR) of all indicator scores and denominators on the hospital level.

Based on the description in the indicator manual, the literature and medical expert opinion, we hypothesized 28 correlations between hip replacement indicators that measure the same underlying construct. Table 2 shows an overview of the hypothesized indicator correlations and their direction of association.

First, to investigate the relationship between continuous structure, process and outcome indicators we used non-parametric Spearman correlations. To assess the uncertainty in the estimated correlation coefficient we calculated 95% confidence intervals. To give a more robust estimation, these intervals were additionally estimated (bootstrapped) using 1000 random replicas (fictitious hospitals) that were constructed from the original dataset. The relationships between the dichotomous structure indicators were analysed by means of chi-square tests. Finally, to examine the relationship between dichotomous structure and continuous process/outcome indicators independent sample t-tests were applied. Here we also bootstrapped using 1000 random replicas. Analyses were conducted in the statistical programs SPSS version 21. Significance was set at $a < 0.05$. P-values below 0.1 were regarded as marginally significant.

RESULTS

Of the total number of 100 hospitals in the Netherlands, on average, 64 hospitals provided data to calculate indicator scores in year 2008. The participation increased in subsequent year, in which on average 95% of the hospitals provided data. Many indicator scores improved from 2008 to 2009. For example, the percentage of wound infections ranged from 0 to 3% across hospitals in 2008, while in 2009 the range was from 0 to 0.03%. (Table 3).

Based on their face validity and on the literature, we hypothesized 28 correlations (hypothesized correlations, hc) to be significant. We found three of these correlations to be significant in the direction hypothesized, of which one was found in the data from 2008 and two were found in the data from 2009 (hc 7, hc 8, hc 19).

As expected, hospitals that reported planning appointments within six weeks after surgery scored lower on the number of deep wound infections than hospitals that did not report scheduling postoperative meetings (hc 8: yes appointment = 0.01% infections, no appointment = 0.02% infections, p-value = 0.001). Further, our analysis showed that hospitals with a higher percentage of patients who did not receive a homologue blood transfusion had a lower percentage of wound infections, although this correlation was only marginally significant (hc 7: $r = -0.28$, p-value = 0.05). Hospitals that reported to have more complication meetings were also more likely to report to provide thrombosis prophylaxis (hc 19: $r = 0.21$, p-value = 0.04).

We found several indicator correlations, which were not a priori expected. We found two significant structure-structure correlations. We observed that hospitals that maintained a complication registration were more likely to score high on planning a postoperative appointment with the patient within six weeks after surgery ($\chi 2$: 19.97, p-value < 0.01). Further, hospitals that reported holding complication meetings also reported using an improvement plan (yes improvement plan = 11%; no improvement plan 0%, p-value = 0.01). We also observed several process-process correlations. First of all, hospitals that reported to administer thrombosis prophylaxis were more likely to report to administer perioperative and preoperative antibiotic prophylaxis to their patients (Spearman R = 0.27, p-value < 0.05) (Spearman R = 0.28, p-value < 0.05). We additionally observed a significant correlation between the administration of antibiotic prophylaxis and the timely administration of antibiotic prophylaxis (Spearman R = 0.46, p-value < 0.01). The timely administra-

Table 3. Hospital-level variation in total hip replacement scores in year 2008 and 2009

		2008						2009					
			Indicator scores on hospital level			Denominators on hospital level			Indicator scores on hospital level			Denominators on hospital level	
		N**	mean	IQR	min-max	median	IQR	N**	mean	IQR	min-max	mean	IQR
qi*1	preoperative patient information	68	1	1-1	1-1	/	/	97	1	1-1	0-1	/	/
qi2a	guideline thrombosis prophylaxis	68	1	1-1	1-1	/	/	68	1	1-1	1-1	/	/
qi2b	thrombosis prophylaxis	64	100	100-100	95 -100	245	49-745	95	100	100-100	93-100	226	56-647
qi3a	complication register	68	1	1-1	0-1	/	/	97	1	1-1	0-1	/	/
qi3b	appointment within 6 weeks	68	1	1-1	1-1	/	/	97	1	1-1	0-1	/	/
qi3c	orthopaedic register form	68	1	1-1	0-1	/	/	97	1	1-1	0-1	/	/
qi3d	complication meeting	63	11	4-12	0-52	/	/	96	11	4-12	0-260	/	/
qi3e	improvement plan	65	1	1-1	0-1	/	/	96	1	1-1	0-1	/	/
qi4a	blood management guideline	68	1	1-1	1-1	/	/	68	1	1-1	1-1	/	/
qi4b	transfusion of homologous blood	52	91	94-100	0-100	241	49-745	90	91	88-100	11-100	222	56-647
5a	guideline for antibiotic prophylaxis	68	1	1-1	1-1	/	/	68	1	1-1	1-1	/	/
5b	perioperative antibiotics	65	100	100-100	100-100	245	49-745	65	100	100-100	100-100	226	56-647
qi5c	antibiotics 60-15 minutes	59	97	100-100	0-100	237	49-745	94	98	100-100	66-100	226	56-647
qi5d	wound infections	60	1	0-1	0-3	245	49-745	93	0	0-0	0-0	213	52-647
qi6	countrywide implemen-tation	68	1	1-1	0-1	/	/	97	1	1-1	0-1	/	/
	average	64	X	X	X			95	X	X	X		

*Quality indicator (qi) **Number of hospitals that delivered the indicator score

tion of antibiotics was negatively associated with having an improvement plan (having an improvement plan = 98; having no improvement plan = 100, p-value = 0.03), taking part in the countrywide implementation registration (taking part = 97.9; not taking part = 100, p-value = 0.03) and having postoperative appointments with patients (having postoperative appointments = 98; having no postoperative appointments = 100, p-value = 0.04).

Having an improvement plan was related to the percentage of patients who received their antibiotic prophylaxis in a timely manner; however, they were related differently than what was expected (hc 27: yes improvement plan = 98%; no improvement plan = 100%, p-value = 0.03) (Table 4).

Table 4. Associations among total hip replacement indicators within the years 2008 and 2009

			quality indicator number										
			qi3a[a]	qi3b	qi3c	qi3e	qi6	qi3d	qi2b	qi4b	qi5b	qi5c	qi 5d
CHI-SQUARE TEST													
Quality indicator number	Quality Indicator Name (indicator type)	Year											
qi3a	Complication register (S[b])	2008											
		2009											
qi3b	Appointment within 6 weeks (S)	2008	\										
		2009	**19.97** (0.00)[c]										
qi3c	Orthopaedic register form (S)	2008	0.57 (0.45)	\									
		2009	0.43 (0.51)	0.09 (0.77)									
qi3e	Improvement plan (S)	2008	0.29 (0.59)	\	0.13 (0.71)								
		2009	0.28 (0.60)	4.38 (1.00)	0.13 (0.71)								
qi6	Countrywide implementation (S)	2008	1.41 (0.24)	\	0.20 (0.66)	0.10 (0.75)							
		2009	0.43 (0.51)	0.09 (0.77)	0.18 (0.67)	0.13 (0.71)							

INDEPENDENT T-TEST								SPEARMAN CORRELATION COEFFICIENT					
qi3d	Complication meeting (S)	2008	yes 11 no 8.2 (0.41) [d]	\	yes 10.5 no 13 (0.81)	**yes 11 no 0 (0.01)**	yes 10.8 no 8.0 (0.73)						
		2009	yes 11.3 no 13.1 (0.86)	yes 11.4 no 12 (0.98)	yes 11.7 no 4.5 (0.19)	yes 11.7 no 4.7 (0.19)	yes 11.6 no 7 (0.74)						
qi2b	Thrombosis prophylaxis (P)	2008	yes 99.9 no 100 (0.73)	\	yes 60 no 100 (0.80)	yes 99.9 no 100 (0.86)	yes 99.9 no 100 (0.83)	0.09 (0.50)[e]					
		2009	yes 99.9 no 98.9 (0.19)	yes 99.8 no 98.9 (0.18)	yes 99.8 no 99.8 (0.89)	yes 99.8 no 99.4 (0.52)	yes 99.8 no 100 (0.65)	**0.21 (0.04)**					
qi4b	Transfusion of homologous blood (P)	2008	yes 90.7 no 98.4 (0.09)	\	yes 91 no 90.7 (0.98)	yes 99.1 no 100 (0.09)	yes 90.9 no 97.4 (0.12)	-0.16 (0.26)	0.08 (0.58)				
		2009	yes 91.8 no 89.3 (0.57)	yes 91.5 no 94.1 (0.70)	yes 92 no 82.9 (0.26)	yes 91.8 no 75.8 (0.17)	yes 91.5 no 93.6 (0.55)	0.01 (0.91)	0.17 (0.12)				
qi5b	Perioperative antibiotics	2008	\	\	\	\	\	\	\	\			
		2009	yes 99.7 no 99.2 (0.60)	yes 99.6 no 100 (0.07)	yes 99.6 no 100 (0.06)	yes 99.7 no 97.8 (0.39)	yes 99.6 no 100 (0.06)	0.15 (0.16)	**0.27 (0.01)**	0.15 (0.17)			
qi5c	Antibiotics 60-15 minutes (P)	2008	yes 97.1 no 100 (0.29)	\	\	yes 99 no 50 (0.51)	yes 97.2 no 100 (0.74)	0.15 (0.26)	-0.04 (0.79)	-0.11 (0.45)	\		
		2009	yes 98 no 99 (0.57)	**yes 98 no 100 (0.04)**	yes 98.4 no 88.5 (0.28)	**yes 98 no 100 (0.03)**	**yes 97.9 no 100 (0.03)**	0.10 (0.32)	**0.28 (0.01)**	0.12 (0.25)	**0.46 (0.00)**		
qi5d	Wound infections (O)	2008	yes 0.8 no 1.0 (0.74)	\	yes 0.8 no 1.0 (0.60)	yes 0.84 no 0.44 (0.26)	yes 0.8 no 0.3 (0.17)	0.12 (0.39)	-0.03 (0.84)	**-0.28 (0.05)**	\	0.20 (0.14)	
		2009	yes 0.01 no 0.01 (0.43)	**yes 0.01 no 0.02 (0.001)**	yes 0.01 no 0.0 (0.42)	yes 0.01 no 0.01 (0.72)	yes 0.01 no 0.01 (0.89)	0.01 (0.91)	-0.04 (0.71)	-0.19 (0.07)	-0.03 (0.74)	0.04 (0.74)	

[a] numbers indicate indicator numbers according to vertical indicator numbering
[b] S= structure, P= process, O=outcome
[c] x2 test (p-value)
[d] t-test: mean group 1, 2 (p-value)
[e] Spearman correlation coefficient (p-value)
bold numbers indicate significance

DISCUSSION

By correlating structure, process, and outcome indicators we measured the construct validity of national quality indicators for hip replacement. Of the 28 a priori expected correlations (per year) only three were observed to be significant in the direction hypothesized. Additionally eight correlations that were not a priori expected were also found to be significant. None of the correlations were consistent over the two-year time period, despite the scientific foundation of the quality indicators and overall expert consensus regarding their validity. Therefore, the construct validity of the quality indicator set under evaluation seems limited. We only found three of the a priori expected correlations to be significant. For example, we observed that in hospitals that scheduled an appointment with a patient within six weeks after the patient's hip replacement, the number of relevant wound infections after hip replacement was lower compared to hospitals that did not plan such an appointment. This is consistent with the international literature and with the widely held opinion that an appointment within this period helps to detect postoperative complications at an early stage, and thereby to prevent advanced, severe wound infections [16]. We additionally observed several process-process correlations, which, in retrospect, might indicate an overall quality awareness culture on the hospital level. For example, hospitals that had high scores on the administration of perioperative antibiotics also had high scores on the administration of antibiotics prior the incision.

An important factor that can explain the lack of construct validity that was observed in our study is the low data reliability. Although the data registration showed signs of improvement in 2009 compared to 2008, data reliability remained an issue. In previous studies it was found that differences in data collection and reporting methods used by hospital employees, such as the use of different indicator definitions, most likely influenced the comparability of the DHTP data [2]. Moreover, many of the indicators are not very specific. For instance, 9 of the 15 hip replacement indicators are dichotomous indicators (yes/no). But for example, although guidelines are available (e.g. qi4a, qi5b), this does not ensure the quality of the guideline or the adherence to the guideline.

The lack of correlation we found among the indicators may be explained by the limited variation and the small numbers observed among many of the included quality indicators. For example, in 2008 the average event rate for patients developing wound infections was merely 1%. When there are few observations and event rates are that low, indicator scores will ran-

domly fluctuate over time, even if the underlying quality of care remains constant [17].

Furthermore, an important factor influencing construct validity is the lack of case-mix correction, as case-mix factors make up a large part of observed outcome variation [18]. Lack of adjustment for patient characteristics, which are not related to quality of hospital care but influence the patients' risk for an outcome, may lead to a biased reflection of quality of care and an unfair comparison between hospitals. As aggregated hospital-level data currently does not include information on the underlying patient characteristics, a valid and fair analysis between the hospitals cannot be guaranteed.

As quality improvement has become a central tenet of health care, QIs are becoming increasingly important. Many countries have already started their own QI program and many more are preparing to start QI programs soon. Despite the increasing number of countries implementing QI programs, the number of studies testing the validity of indicators is limited. While a number of studies have tested the construct validity of indicators in the U.S. [19, 20, 21-24], a limited number of such studies have been conducted in the European health care setting [12]. However, given the differences in national health care and local hospital organization, indicators should be evaluated before they are adopted from another health system. The validity of quality of care indicators cannot be assumed for a health care setting outside of the one where the indicator was developed and tested [1]. Therefore, further research on the validity of the currently used indicators in the health care setting in which they are used is warranted. Several methodological lessons can be learned from our observations. In order for a QI to be valid, it must be reliable [2]. An indicator's reliability is determined by the accuracy of the underlying data and the unambiguousness of the definition of the indicator [2]. Moreover, when hospital employees are responsible for collecting the data and computing the QIs, there needs to be some central control over these processes. Moreover, to increase data reliability the software market should be regulated and standards should be set for the development of automatic data extraction software. In order to find relations between indicators, it is crucial to take into account the influence of low event rates and case-mix differences. Failing to adjust for these factors may confound the relationship between quality indicators.

In our study we could not adjust for patient characteristics and low event rates, as we only had hospital-level data, which is a limitation of our study. There are other limitations of our study that need to be noted. First, we op-

erationalized construct validity by the correlation between two test scores. Usually, in psychometric research, a person's score on for example a new psychological test is correlated with a score on a more established test measuring the same underlying construct [14]. In our study both test scores were derived from the same database and were both the subject of study. Merely the presence of a significant correlation that was expected based on the literature was considered to be an indication of the construct validity of both indicators. One could argue therefore that the method of validity assessment in our study is not very strong. A better way to assess the construct validity is to correlate the indicator scores of interest with measures derived from other clinical databases. However, for countries in which reliable health care databases are scarce ours is the only approach possible. In our study we used a significant correlation in the expected direction as a sign for construct validity; however, most of the significant correlations were weak. Third, when assessing multiple correlations one typically corrects for multiple testing, for instance with a Bonferoni correction. As we a priori planned our correlations based on the available scientific evidence, we did not correct for multiple testing. However, we do realize that we have to treat the observed significant correlations with caution. Further research and trend data is needed to test construct validity over a longer time period in order to be able to identify systematic indicator associations.

Conclusion

In all it can be concluded that despite the face validity of hospital quality indicators for hip replacement, construct validity seems to be limited. Although the individual indicators might be valid and actionable, drawing overall conclusions based on the whole indicator set should be done carefully, as construct validity could not be established. Factors that may explain the lack of construct validity are poor data quality, no adjustment for case-mix and statistical uncertainty. Before any action can be taken based on the indicator scores, these limitations must be addressed.

REFERENCES

1. Anema HA, Kievit J, Fischer C, Steyerberg EW, Klazinga NS. Influences of hospital information systems, indicator data collection and computation on reported Dutch hospital performance indicator scores. BMC Health Serv Res. 2013;13:212. PubMed PMID: 23758921. Pubmed Central PMCID: 3698115. Epub 2013/06/14. eng.
2. Anema HA, van der Veer SN, Kievit J, Krol-Warmerdam E, Fischer C, Steyerberg E, et al. Influences of definition ambiguity on hospital performance indicator scores: examples from The Netherlands. Eur J Public Health. 2013 Apr 18. PubMed PMID: 23543677.
3. Adeyemo D, Radley S. Unplanned general surgical re-admissions - How many, which patients and why? Ann R Coll Surg Engl. 2007;89(4):363-7.
4. Donabedian A. The quality of care. How can it be assessed? JAMA. 1988 Sep 23-30;260(12):1743-8. PubMed PMID: 3045356. Epub 1988/09/23. eng.
5. Mainz J. Defining and classifying clinical indicators for quality improvement. Int J Qual Health Care. 2003 Dec;15(6):523-30. PubMed PMID: 14660535. Epub 2003/12/09. eng.
6. Torjesen I. NHS is unlikely to meet Nicholson challenge to deliver pound20bn in efficiency savings, says King's Fund. Bmj. 2012;345:e6496. PubMed PMID: 23015359.
7. Jenkins PJ, Clement ND, Hamilton DF, Gaston P, Patton JT, Howie CR. Predicting the cost-effectiveness of total hip and knee replacement: a health economic analysis. The bone & joint journal. 2013 Jan;95-B(1):115-21. PubMed PMID: 23307684.
8. SooHoo NF LJ, Ko CY, Zingmond DS. Provider volume of total knee arthroplasties and patient outcomes in the HCUP-nationwide inpatient sample. The Journal of bone and joint surgery American volume. 2003;85(9):12.
9. Mahomed NN, Barrett JA, Katz JN, Phillips CB, Losina E, Lew RA, et al. Rates and outcomes of primary and revision total hip replacement in the United States medicare population. The Journal of bone and joint surgery American volume. 2003 Jan;85-A(1):27-32. PubMed PMID: 12533568.
10. Bhattacharyya T, Freiberg AA, Mehta P, Katz JN, Ferris T. Measuring the report card: the validity of pay-for-performance metrics in orthopedic surgery. Health Aff (Millwood). 2009 Mar-Apr;28(2):526-32. PubMed PMID: 19276012. Pubmed Central PMCID: 3004748.
11. Desai AS, Stevenson LW. Rehospitalization for heart failure: predict or prevent? Circulation. 2012 Jul 24;126(4):501-6. PubMed PMID: 22825412. Epub 2012/07/25. eng.
12. Fischer C, Anema HA, Klazinga NS. The validity of indicators for assessing quality of care: a review of the European literature on hospital readmission rate. Eur J Public Health. 2012 Aug;22(4):484-91. PubMed PMID: 22140251. Epub 2011/12/06. eng.
13. Heiden-van der Loo M; Ho VKY DR, et al. Weinig lokaal recidieven na mammachirurgie: goede kwaliteit van de Nederalndse borstkankerzorg. Nederlands tijdschrift voor geneeskunde. 2010;154:A1984:1.
14. Cronbach LJ, Meehl PE. Construct validity in psychological tests. Psychol Bull. 1955 Jul;52(4):281-302. PubMed PMID: 13245896. Epub 1955/07/01. eng.
15. Kallewaard M BN, van Everdingen JJE, et al. . Kwaliteit van Zorg in de Etalage, Eindrapportage 2007 [cited 2013 15.05.2013]. Available from: https://zichtbarezorg.dmdelivery.com/mailings/FILES/htmlcontent/Ziekenhuizen/Eindrapportage%20%27Kwaliteit%20van%20zorg%20in%20de%20etalage%27.pdf
16. Saleh K OM, Resig S, et al. Predictors of wound infection in hip and knee joint replacement: results from a 20 year surveillance program. J Orthop Res. 2000;20(3):10.
17. Walker K, Neuburger J, Groene O, Cromwell DA, van der Meulen J. Public reporting of surgeon outcomes: low numbers of procedures lead to false complacency. Lancet. 2013 Nov 16;382(9905):1674-7. PubMed PMID: 23831144. Epub 2013/07/09. eng.
18. van Gestel YRBM, Lemmens VEPP, Lingsma HF, de Hingh IHJT, Rutten HJT, Coebergh JWW. The hospital standardized mortality ratio fallacy: a narrative review. Medical Care. 2012;50(8):662-7.
19. Peterson ED, Roe MT, Mulgund J, DeLong ER, Lytle BL, Brindis RG, et al. Association between hospital process performance and outcomes among patients with acute coronary syndromes. JAMA. 2006 Apr 26;295(16):1912-20. PubMed PMID: 16639050. Epub 2006/04/28. eng.

20. Bradley EH, Herrin J, Elbel B, McNamara RL, Magid DJ, Nallamothu BK, et al. Hospital quality for acute myocardial infarction: correlation among process measures and relationship with short-term mortality. JAMA. 2006 Jul 5;296(1):72-8. PubMed PMID: 16820549. Epub 2006/07/06. eng.
21. Silber JH, Williams SV, Krakauer H, Schwartz JS. Hospital and patient characteristics associated with death after surgery. A study of adverse occurrence and failure to rescue. Med Care. 1992 Jul;30(7):615-29. PubMed PMID: 1614231. Epub 1992/07/01. eng.
22. Tsai TC, Joynt KE, Orav EJ, Gawande AA, Jha AK. Variation in surgical-readmission rates and quality of hospital care. N Engl J Med. 2013 Sep 19;369(12):1134-42. PubMed PMID: 24047062. Pubmed Central PMCID: 4107655. Epub 2013/09/21. eng.
23. Isaac T, Jha AK. Are patient safety indicators related to widely used measures of hospital quality? J Gen Intern Med. 2008 Sep;23(9):1373-8. PubMed PMID: 18574640. Pubmed Central PMCID: 2518036. Epub 2008/06/25. eng.
24. Werner RM, Bradlow ET. Relationship between Medicare's hospital compare performance measures and mortality rates. JAMA. 2006 Dec 13;296(22):2694-702. PubMed PMID: 17164455. Epub 2006/12/14. eng.
25. Woolf SH, Grol R, Hutchinson A, Eccles M, Grimshaw J. Clinical guidelines: potential benefits, limitations, and harms of clinical guidelines. Bmj. 1999 Feb 20;318(7182):527-30. PubMed PMID: 10024268. Pubmed Central PMCID: 1114973. Epub 1999/02/19. eng.
26. Grimshaw JM, Russell IT. Effect of clinical guidelines on medical practice: a systematic review of rigorous evaluations. Lancet. 1993 Nov 27;342(8883):1317-22. PubMed PMID: 7901634. Epub 1993/11/27. eng.
27. Engesaeter LB, Lie SA, Espehaug B, Furnes O, Vollset SE, Havelin LI. Antibiotic prophylaxis in total hip arthroplasty: effects of antibiotic prophylaxis systemically and in bone cement on the revision rate of 22,170 primary hip replacements followed 0-14 years in the Norwegian Arthroplasty Register. Acta Orthop Scand. 2003 Dec;74(6):644-51. PubMed PMID: 14763692. Epub 2004/02/07. eng.
28. Southwell-Keely JP, Russo RR, March L, Cumming R, Cameron I, Brnabic AJ. Antibiotic prophylaxis in hip fracture surgery: a metaanalysis. Clin Orthop Relat Res. 2004 Feb(419):179-84. PubMed PMID: 15021151. Epub 2004/03/17. eng.
29. Slappendel R, Dirksen R, Weber EW, van der Schaaf DB. An algorithm to reduce allogenic red blood cell transfusions for major orthopedic surgery. Acta Orthop Scand. 2003 Oct;74(5):569-75. PubMed PMID: 14620978. Epub 2003/11/19. eng.
30. Sculco TP, Baldini A, Keating EM. Blood management in total joint arthroplasty. Instr Course Lect. 2005;54:51-66. PubMed PMID: 15948435. Epub 2005/06/14. eng.
31. Pedersen A, Johnsen S, Overgaard S, Soballe K, Sorensen HT, Lucht U. Registration in the danish hip arthroplasty registry: completeness of total hip arthroplasties and positive predictive value of registered diagnosis and postoperative complications. Acta Orthop Scand. 2004 Aug;75(4):434-41. PubMed PMID: 15370588. Epub 2004/09/17. eng.

Chapter 7.

The impact of statistical uncertainty and case-mix correction on the reliability and validity of quality indicators for colon cancer surgery

**Claudia Fischer, Hester F Lingsma, Nicoline van Leersum ,
Rob AEM Tollenaar, Michel W Wouters, Ewout W Steyerberg**

This article was published as:
Comparing colon cancer outcomes: the impact of low hospital case volume and case-mix adjustment.
European Journal of Surgical Oncology. 2015; 41(8):1045-53.

ABSTRACT

Background: When comparing performance across hospitals it is essential to consider the noise caused by low hospital case volume and to perform adequate case-mix correction. We aimed to quantify the role of noise and case-mix correction on standardized postoperative mortality and anastomotic leakage (AL) rates.

Methods: We studied 13 120 patients who underwent colon cancer resection in 85 Dutch hospitals. We addressed differences between hospitals in postoperative mortality and AL, using fixed (ignoring noise) and random effects (incorporating noise) logistic regression models with general and additional, disease specific, case-mix correction.

Results: Adding disease specific variables improved the performance of the case-mix correction models for postoperative mortality (c-statistic increased from 0.77 to 0.81). The overall variation in standardized mortality ratios was similar, but some individual hospitals changed considerably. For the standardized AL rates the performance of the adjustment models was poor (c-statistic 0.59 and 0.60) and overall variation was small. Most of the observed variation between hospitals was actually noise.

Conclusions: Noise had a larger effect on hospital performance than extended case-mix correction, although estimates for some individual hospitals were affected by more detailed case-mix correction. To compare outcomes between hospitals it is crucial to consider noise due to low hospital case volume with a random effects model.

INTRODUCTION

Accountability to the public has become a major topic in health care in the past decade; consequently outcome measures increasingly receive attention.[1,2] Colorectal cancer is the third most common cancer,[3] in 2012 an incidence of 13 408 new cases is registered in the Netherlands.[4] Surgeries for colorectal cancer cause considerable morbidity and mortality.[5] In cancer surgery mainly short-term outcome measures are used to measures quality of care, as they are little influenced by non-modifiable disease-related factors.[6] Short-term outcome measures for colorectal cancer surgery include postoperative mortality and anastomotic leakage (AL). AL is one of the most serious complications in colon cancer surgery, as it is associated with increased morbidity and mortality.[7] Moreover, variations between providers can be observed.[8-10]

However, variation in outcome across centres is not made up exclusively by quality of care differences, which it is intended to display.[11,12] Case-mix factors are thought to explain a large part of the observed outcome variation,[13] but case-mix models are usually fairly generic.[14] Especially when they are based on administrative data, not intended for quality assessment.[15-17] For colon cancer, research has shown that surgical high risk profile patients are not equally distributed across centres.[18,19]

Next, partial variation across centres can be attributed to 'noise' in the comparison (variation by chance). The noise in the comparison is especially influenced when the hospital case volume, so the number of patients with an event per hospital, is low. For example, an average number of 10 deaths per year in a hospital will fluctuate over the years because of chance, also if no changes in quality occur. With larger numbers of outcomes per hospital (e.g. >50) the effect of chance variation decreases. Random effects regression models are able to take the variation by chance into account, in contrast to the mainly used fixed effects models.[20,21]

In this study we aimed to quantify the role of noise and case-mix correction on standardized postoperative mortality and AL rates.

METHODS

Study population

For this study data was derived from the Dutch Surgical Colorectal Audit (DSCA), a national, web-based and interactive database.[22] The database, in which all Dutch hospitals participate, includes detailed information on patient-and tumour characteristics, diagnostics, procedures and outcomes of patients undergoing a resection of a primary colorectal carcinoma. In each hospital a surgeon is appointed for the data-entry. For participating hospitals it is possible to review their uploaded data as well as benchmarking information on a protected web page. Further, data quality reports are sent to the hospitals.[18] Approximately 97 percent of all patients with a primary colorectal carcinoma resection in the Netherlands are captured in this database.[23] The dataset is based on evidence-based guidelines and annually verified with the Netherlands Cancer Registry (NCR) data. Further information on the data collection and methodology of the DSCA can be found elsewhere[18,22] [www.dsca.clinicalaudit.nl].

We used data of all patients who underwent a primary colon carcinoma resection in the Netherlands between 1st of January 2011 and 31st of December 2012 (92 hospitals, 13 672 patients). Seven hospitals (with 552 patients) were excluded because they had zero outcome events (no deaths or no anastomotic leakage (AL)), resulting in 85 hospitals and 13120 patients included in the analyses.

Predictors and outcome measures

Demographic variables included age and gender. Clinical variables included American Society of Anesthesiologists (ASA) score, Charlson comorbidity index, International Union Against Cancer tumour node metastasis (TNM) classification of malignant tumours (6th edition), histological tumour type and tumour number and distance, in case of double tumour (distant tumours are located in a segment not adjacent to other (hemi left, hemi right and rectum)) and preoperative complications (perforation with fecal peritonitis, abscess, bowel obstruction, blood loss/anemia). The postoperative short-term outcomes assessed were postoperative mortality and anastomotic leakage. Postoperative mortality was defined as dead during the index admission in which the surgery took place or within 30 days after surgery. AL was defined as 'a clinically relevant anastomotic leak requiring a reintervention'. Both radiological and surgical reinterventions were considered.

Statistical analysis

To describe between-hospital differences in operated patients we calculated medians and interquartile ranges for the baseline demographic and clinical characteristics on hospital level. We constructed six different models for each of the two outcomes (postoperative mortality and AL):

1. A crude fixed effects logistic model with taking no patient characteristics or chance into account.
2. A fixed effects logistic multivariable model including generic case-mix factors: age, gender, urgency of surgery, Charlson comorbidity index and year of operation.
3. The extended case-mix adjusted fixed effects logistic model, which consists of model 2 plus additional disease specific factors. The variables considered were age, gender, urgency of surgery, Charlson comorbidity index, year of operation, ASA-score, TNM stage, preoperative tumour complications, histological tumour type, number and distance of tumours.

These three models were also fitted as logistic random effects models with hospital as a random intercept. Such a random effects model takes out the noise in the comparison. To evaluate the effect of additional case-mix correction, the c-statistic (equivalent to the area under the ROC curve) was used. The c-statistic quantifies discrimination, i.e. the models ability to separate patients with different outcomes.[24] A c-statistic of 0.5 indicates no discriminative power at all, whereas a model with c-statistic of 1.0 indicates perfect discrimination.

Standardized postoperative mortality and AL rates were used to quantify the effect of case-mix correction and noise on the overall between-hospital variation and the estimated outcomes for individual hospitals. The Standardized Mortality (or AL) Rate (SMR) is a ratio between the observed number of deaths in a hospital and the number of expected deaths in a hospital. The expected number of deaths per hospital was estimated with the case-mix correction models. Being either the overall mortality rate (crude model) or the sum of the predicted probabilities per hospital (general and extended model). The observed number of deaths in the fixed effects model is the counted number of deaths. In the random effects models the observed number of deaths is the overall intercept plus the hospital specific random intercept from the crude random effects model, transformed into a probability. Multiplying this probability by the total number of patients in that hospital gives the 'observed' number, adjusted for variation by chance. An

SMR above 100 indicates "excess deaths" whereas a number below 100 indicates that a hospital performed better as one would have predicted on basis of the hospitals expected deaths.

The effect of the different models on the overall between-hospital variation was quantified by calculating interquartile ranges of the SMRs. The effect on individual hospital SMRs was assessed with bi-plots.

Missing data in the predictors were imputed with multiple imputation using all relevant prognostic factors and outcome.

Statistical analysis were carried out with the statistical software packages SPSS version 18.0 (SPSS, Chicago, IL, USA) and R statistical software 3.0.1 (R Foundation for Statistical Computing, Vienna, Austria).

RESULTS

Between 1st January 2011 and 31st December 2012 13 120 patients underwent colon cancer surgery and were registered in the DSCA data set. For the outcome postoperative mortality we analyzed 13 120 patients. Information on the outcome anastomotic leakage (AL) was missing in 1756 patients and we thus analyzed 11 364 patients. Three baseline variables had missing values: ASA-score (3%), urgency (<1%) and histological tumour type (2%). Across the 85 participating hospitals considerable variation in patient characteristics was observed. For example, the mean age varied by two years across hospitals (interquartile range 70-72), but larger variation was observed in the patients with preoperative complications (26%-56%) and the number of urgent operations (16%-24%) (Table 1).

Case-mix correction models

In the univariate analysis it appeared that the odds of postoperative mortality was higher in patients who were older (OR, 2.3; 95% CI, 2.0-2.5), had a higher Charlson score (OR, 1.3; 95% CI, 1.3-1.4), higher ASA-score (e.g. ASA-score 2 vs. ASA-score 1: OR, 2.3; 95% CI, 1.4-3.6), an urgent operation (OR, 2.7; 95% CI, 2.2-3.3), higher TNM stage (e.g. TNM stage 2 vs. TNM stage1: OR, 2.3; 95% CI, 1.4-3.6) and preoperative complications (OR, 2.2; 95% CI, 1.4-3.6). Predictors that significantly increased the odds of developing an AL were: male gender (OR, 1.5; 95% CI, 1.3-1.7), higher Charlson score (OR. 1.0; 95% CI, 1.0-1.1), an urgent operation (OR, 1.5; 95% CI, 1.3-1.7), higher ASA-score (e.g. ASA-score 2 vs.1: OR, 1.2; 95% CI, 1-1.4), a distant tumour (OR, 1.6; 95% CI, 1.0-2.6) and preoperative complications (OR, 1.4; 95% CI, 1.2-1.6) (Table 2).

In the multivariable case-mix models significant predictors for postoperative mortality were an urgent operation (mortality: OR, 2.8; 95% CI, 2.3-3.4; AL: OR, 1.5; 95% CI, 1.3-1.9) and a higher Charlson score (mortality: OR, 1.3; 95% CI, 1.2-1.3; AL: OR, 1.1; 95% CI, 1.0-1.1). The presence of preoperative complications and ASA-score were important predictors for both outcomes. For instance, a higher ASA-score increased the likelihood of both events significantly, with a patient with ASA-score four or five had an almost 12-fold risk of dying compared to a patient with an ASA-score of 1 (OR, 11.6; 95%CI, 6.6-20.4) and a nearly twofold risk for the occurrence of an AL (OR, 1.7; 95%CI, 0.93-3.2). Older age was only a statistically significant predictor for postoperative mortality (OR 2.1, 95% CI 1.9-2.4), a 10 years older age implied a twofold risk of postoperative mortality (Table 2). Gender, the number and location of the tumour were significant predictors for the outcome AL. Patients with a distant tumour had a 62% higher chance of an AL compared to a patient with a single tumour (OR, 1.6; 95% CI, 1.0-2.6).

Table 1. Between hospital variation of patients undergoing colon cancer resection in Dutch hospitals

	Number of patients		Between hospital comparison	
			Median	IQR[c]
Patients	13 120		137	101-194
Age (in years)[a]	13 120		71	70-72
Gender [a]	13 120	Female	52%	49%-55%
Year of operation[a]	13 120	2012	51%	47%-55%
Charlson score[a]	13 120		1	0.8-1.1
ASA-score[b]	13 081	Normal healthy patient (1)	17%	12%-22%
		Mild systemic disease (2)	55%	51%-60%
		Severe systemic disease (3)	24%	18%-30%
		Life-threatening disease/ Moribund patient (4/5)	1%	1%-3%
Urgency[b]	13 090	Elective	81%	76%-84%
		Urgent/Acute	19%	16%-24%
TNM stage[b]	13 120	0	1%	0%-2%
		I	17%	14%-19%
		II	36%	31%-40%
		III	32%	29%-35%
		IV	13%	10%-16%
		X	0 %	0 %-1%
Histological tumour type[b]	12 930	Adenocarcinoma	92%	90%-96%
		Mucinous adenocarcinoma	5 %	1%-8%
		Signet ring cell carcinoma	0%	0%-1%
		Others	0%	0%-1%
Number of tumour and location[b]	13 120	Single tumour	96%	94%-97%
		Close double tumour	2%	1%-3%
		Synchronous colorectal tumour	2%	1%-3%
Preoperative tumour complication[b]	13 120	Yes	37%	26%-56%
Anastomotic leakage	11 364	Yes	6.7% (9)[d]	4.7%-8.8% (5-12)
Postoperative mortality	13 120	Yes	3.2% (5)[d]	2.3%-4.3% (3-7)

[a] generic case-mix factor
[b] extended case-mix factor
[c] interquartile range
[d] percentage (total numbers)

Table 2. Multivariable models with outcomes postoperative mortality and anastomotic leakage with different case-mix correction

	Postoperative mortality						Anastomotic leakage					
	Univariate analysis		Generic case-mix		Extended case-mix		Univariate analysis		Generic case-mix		Extended case-mix	
	OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI
Age per decade[a]	2.3	2.0-2.5	2.1	1.9-2.4	1.9	1.7-2.2	0.9	0.8-1	0.9	0.8-0.9	0.8	0.8-0.9
Year of operation[a]												
2012	1		1		1		1		1		1	
2011	1.1	0.9-1.3	1.1	0.9-1.3	1.1	0.9-1.3	1.0	0.9-1.2	1.0	0.8-1.1	1.0	0.9-1.2
Charlson score[a]	1.3	1.3-1.4	1.3	1.2-1.3	1.1	1.1-1.2	1.0	1.0-1.1	1.1	1.0-1.1	1.0	1.0-1.1
Urgency[a]												
Elective	1		1		1		1		1		1	
Urgent/acute	2.7	2.2-3.3	2.8	2.3-3.4	1.8	1.4-2.2	1.5	1.3-1.8	1.5	1.3-1.9	1.3	1.0-1.6
Gender[a]												
Female	1		1		1		1		1		1	
Male	1.1	0.9-1.3	1.2	1.0 -1.4	1.1	0.9-1.4	1.5	1.3-1.7	1.5	1.3-1.7	1.4	1.2-1.7
Histological type[b]												
Adenocarcinoma	1				1		1				1	
mucinous adenocarcinoma	0.8	0.5-1.3			0.7	0.4-1.2	1.0	0.7-1.4			1.0	0.7-1.4
Signet ring cell carcinoma	1.1	0.4-3.1			0.7	0.3-2.2	1.5	0.8-3.1			1.3	0.7-2.7
Others	1.0	0.5-2.1			0.9	0.4-2.1	0.5	0.2-1.1			0.5	0.2-1.2
ASA-score[b]												
Normal healthy patient (1)	1				1		1				1	
Mild systemic disease (2)	2.3	1.4-3.6			1.3	0.8-2.1	1.2	1-1.4			1.3	1.0-1.6
Severe systemic disease (3)	9.5	6.0-15.0			3.5	2.2-5.8	1.5	1.2-1.9			1.6	1.3-2.1
Life-threatening disease/ Moribund patient (4/5)	37.5	22.1-63.7			11.6	6.6-20.4	1.7	0.9-3.1			1.7	0.9-3.2
TNM Stage[b]												
0	1				1		1				1	
I	1.8	0.5-5.7			1.3	0.4-4.3	1.1	0.5-2.2			0.9	0.4-1.8
II	2.0	0.6-6.2			1.2	0.3-3.9	1.4	0.7-2.7			1.1	0.6-2.3
III	1.8	0.6-5.6			1.2	0.3-4.0	1.4	0.7-2.7			1.1	0.5-2.2
IV	3.6	1.1-11.4			2.3	0.7-7.9	1.6	0.8-3.2			1.1	0.6-2.4
X	4.0	1.0-15.7			2.8	0.7-11.5	1.5	0.5-4.2			1.3	0.4-3.5
Preoperative complication[b]												
No	1				1		1				1	
Yes	2.2	1.8-2.6			1.3	1.1-1.7	1.4	1.2-1.6			1.2	1.1-1.5

Number & location tumour[b] Single tumour	1				1		1				1	
Close synchronous colorectal tumour	0.7	0.3-1.4			0.7	0.3-1.4	1.1	0.7-1.7			1.1	0.7-1.7
Distant synchronous colorectal tumour	1.6	0.9-2.9			1.5	0.8-2.7	1.6	1.0-2.6			1.6	1.0-2.6

[a]generic case-mix factor
[b]extended case-mix factor

With regard to the overall model performance, the c-statistic increased from 0.765 (95%CI 0.744-0.787) for the generic mortality case-mix correction model to 0.816 (95%CI 0.796-0.836) for the extended case-mix correction mortality model. For AL the c-statistic with generic case-mix and extended case-mix were 0.585 (95% CI: 0.565-0.606) and 0.605 (95% CI: 0.585-0.625) respectively.

Effect of case-mix and noise adjustment

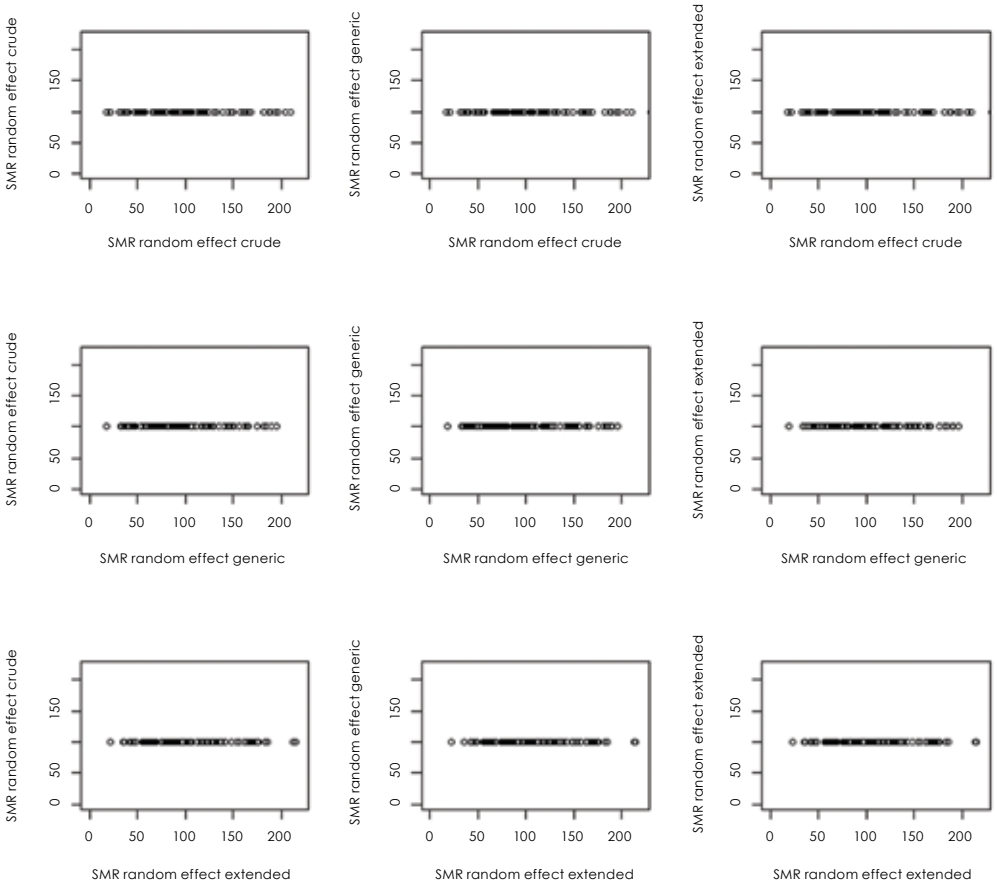
Without any adjustment (crude fixed effects model) the between-hospital differences in standardized postoperative mortality appeared substantial (interquartile range 70-129). When taking into account generic patient characteristics, the differences even increased (interquartile range 67-141). Using

Table 3. Standardized mortality/anastomotic leakage ratios and model performance of fixed and random effects models

	Standardized postoperative mortality ratio		Standardized anastomotic leakage ratio	
	IQR[a]		IQR[a]	
Fixed effects model				
Crude	70	129	69	127
Generic case-mix correction	67	141	69	133
Extended case-mix correction	67	133	68	133
Random effects model				
Crude	100	100	86	109
Generic case-mix correction	100	100	83	110
Extended case-mix correction	100	100	84	110

[a] Interquartile range

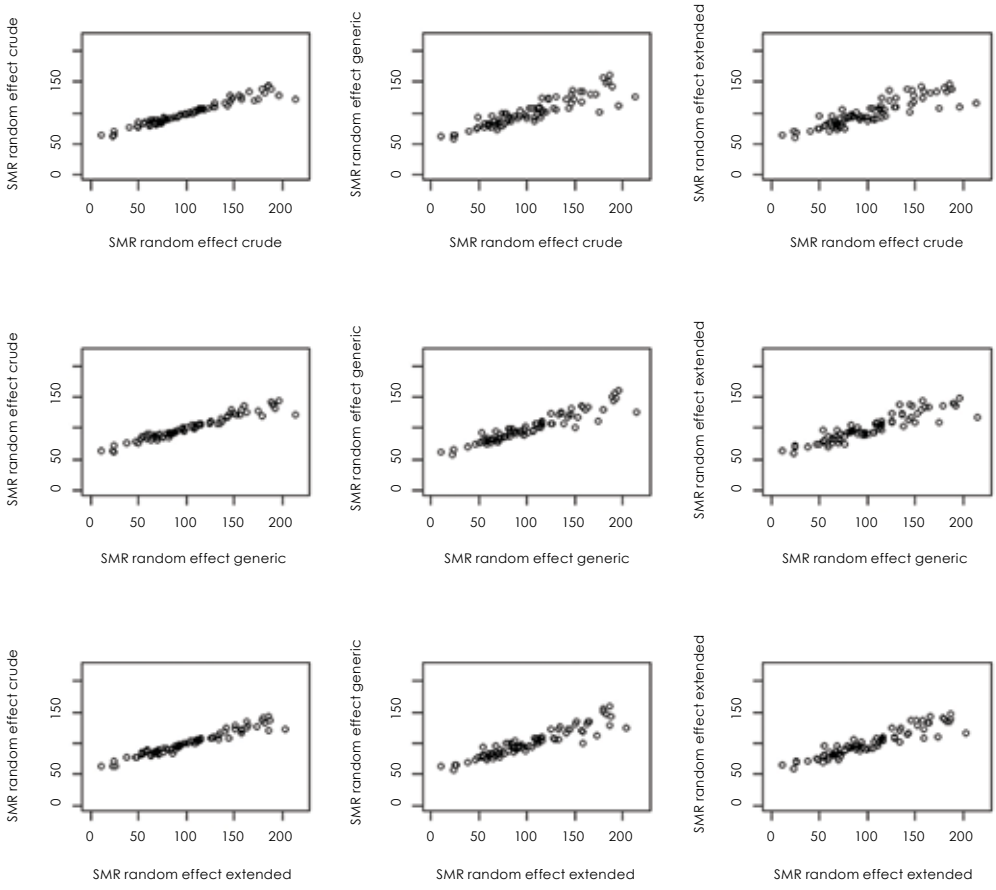
Figure 1. Between–hospital variation for standardized postoperative mortality ratio (SMR) based on generic and extended case-mix correction in fixed and random effects models



the extended case-mix model revealed, however, that a part of the differences could be explained by the extra disease specific variables; the SMRs converged again to an interquartile range of 67-133. Additionally, for some individual hospitals, the SMR changed substantially with additional case-mix correction.

Taking into account the noise in the comparison with a random effects model diminished remaining differences in standardized mortality rates between the hospitals totally (Table 3) (Figure 1).

Figure 2. Between-hospital variation for standardized anastomotic leakage ratio (SMR) based on generic and extended case-mix correction in fixed and random effects models



For the outcome AL the crude fixed effects model indicated similarly large outcome differences (interquartile range 69-127). The generic case-mix correction and the extended case-mix correction had very limited effect on the overall variation and on the standardized AL rates of individual hospitals (Figure 2). The random effects models revealed that also for AL most of the between-hospital variation was attributable to noise in the comparison (inter-quartile range, crude random effects model: 86-109). This can be also seen from Figure 2: the spread in AL ratios becomes much smaller when taking into account noise.

DISCUSSION

In this study we quantified the role of noise and case-mix correction on standardized postoperative mortality and anastomotic leakage (AL) rates. We found that additional disease specific case-mix correction had some effect on standardized postoperative mortality rates compared to generic case-mix correction. However, all variation between hospitals in standardized postoperative mortality rates was explained by noise in the comparison. AL rates were not affected by case-mix correction, but a large part of the between hospital variation was again explained by noise.

Case-mix correction

It is generally recognized that between-hospital comparisons in outcome should be adjusted for potential differences in case-mix. Yet, many current case-mix correction models include only a limited number of generic case-mix factors and lack disease specific information.

We found that several of the additional disease specific case-mix correction variables were significant predictors of postoperative mortality. Adding those predictors to the models resulted in an improvement of the discriminative ability of the model to a c-statistic of 0.81. In contrast, the models for the outcome AL performed rather poor with a c-statistic of around 0.6. These findings are in line with results of earlier research showing that while postoperative mortality is highly sensitive to (disease specific) case-mix factors,[25-27] postoperative complications are especially influenced by intraoperative risk factors.[28] A recent study concluded that disease specific case-mix variables did result in noticeable model improvement in terms of c-statistic.[29] These authors started however with an extensive prediction model and added only two cancer-specific variables.

When comparing outcomes across centres, it is not only relevant whether a case-mix variable is predictive of the outcome, but also whether it shows variation across hospitals.[30] For instance, a strong predictor of mortality that is equally distributed across all hospitals would not affect the SMRs when added to the case-mix correction model. In our data the distribution of generic case-mix correction factors, such as age, varied little between hospital, compared to disease specific factors, such as preoperative complications and ASA score. This explains that although the generic case-mix correction model had a high c-statistic, the model had a limited effect on the SMRs. While the extended adjustment model, with only a somewhat higher c-statistic affected the SMRs

much more. Consequently, our findings confirm that the c-statistic has limited usefulness for quantifying the credibility and accuracy of comparisons across hospitals.[31] The discriminative ability only reflects between-patient outcome variation and not differences in the distribution of patient characteristics across hospitals. Therefore, the c-statistic will remain unchanged if the added risk factor is equally distributed across hospitals. For AL the case-mix factors varied similarly across hospitals but had only a limited predictive strength. Hence, the standardized AL rates were minor affected by case-mix correction. Clinical expertise needs to be used to develop risk adjustment models specifically for the clinical condition or procedure of interest.[31]

Noise

In the cases in which the effect of additional case-mix correction was limited, the consideration of noise was essential for the outcome comparison between hospitals. Taking the noise into account with the random effects models diluted all observed variation in postoperative mortality across the hospitals, as well as a vast amount of variation in anastomotic leakage. Our findings are in line with previous work showing that the number of patients, and specifically the number of patients who got the event per hospital (hospital case volume), have a much greater impact on the accuracy of the hospital comparison than the c-statistic of the risk-adjustment model.[31] Even when applying extensive case-mix correction, a number of hospitals will be misclassified when the effect of noise caused by a low hospital case volume is not taken into account.[32,33] The amount of noise directly depends on the number of observed outcomes. If the total patient number is large, a rare outcome will still lead to a low hospital case volume and thus a lot of random variability. The AL rates were less affected by noise than the postoperative mortality rates. This is, as the number of events per hospital was almost twice as high (median 9) as the number of deaths (median 5), which is crucial from a statistical point of view. In contrast to the fixed effects models, random effects models allow to draw inferences on the overall patient population, which enables the estimation of expected outcome in future patients.

Choice for an outcome indicator

Death is the most undesirable outcome for a patient. Therefore postoperative mortality forms a highly relevant outcome indicator and needs to be monitored. Methodologically, however, mortality as a measure for quality of care has several disadvantages: even when the absolute patient numbers are high, with the steady improvement of medical care and centralization, the number of deaths per hospital is decreasing, leading to a higher impact of noise when comparing mortality rates between hospitals.[34] This is especially the case when

comparing disease specific outcomes,[35] as in our study, where on average just 5 patients per hospital died in the two years study period. To increase the number of patient volume and event rates disease specific outcomes can be rather monitored over a longer time period or alternatively aggregated on hospital level. Doing so, however, complicates the interpretation and improvement of the outcome, as the direct link to care processes cannot be made,[36] feedback loops get longer and therefore results become less relevant and actionable for care givers.

Another problem, also shown in this research, is the sensitivity of mortality to case-mix factors.[36] Given the limitations of mortality rates as an indicator of surgical quality, other indicators are worthwhile to monitor in addition. Anastomotic leaks are severe post-operative complications causing significant burden to patients[37] and closely related to the surgical process and therefore likely be attributable to quality of care. As our study confirmed, their incidence rate is higher than the ones of postoperative mortality, which is crucial from a methodological point of view. A further advantage of this indicator is its robustness towards patient characteristics.[38] We could show that general patient and disease specific characteristics had a minor predictive ability for ALs, suggesting that observed outcome differences might be due to treatment and hospital factors. Nevertheless, to date no standard definition of AL exists in the literature.[39] When aiming for a valid benchmark of this outcome indicator, a standardized definition forms a pre-requisite.

Strengths and limitations

A major strength of this study is its large, national cohort. The use of audit data enabled the analysis of reliable, clinical case-mix correction information. Although we tested for a range of important clinical variables, patients' behavioural factors could not be taken into account in this study.

Implications

When outcome indicators are used to compare hospital performance, high-quality prospectively set up data registries are a basic pre-requisite. As our study shows, it is important that clinical data are available to perform adequate case-mix correction. High quality data demands active monitoring and data validation.[40] In addition to postoperative mortality, AL seems to be suitable as an indicator of surgical performance, as they occur frequent and are less sensitive to case-mix. However, a single indicator never can reflect a whole quality spectrum. Hospitals performing poor on one indicator do not necessarily perform poor on other related indicators.[41] This implies that multiple (related) indicators, such as postoperative mortality and surgical complications, should

be measured and compared. Novel statistical methods might reduce noise by jointly analyzing multiple surgical indicators, for example, by summarizing them as complicated course or textbook outcome.[42]

In conclusion, noise, due to low hospital case volume, had a larger effect on hospital performance than extended case-mix correction. Although some individual hospital outcome rates were affected by more detailed case-mix correction. To compare outcome rates between hospitals it is crucial to take into account the noise due to low hospital case volume with a random effects model.

Conflict of interest statement
All authors of the manuscript 'Comparing colon cancer outcomes: the impact of low hospital case volume and case-mix correction' declare that there is no conflict of interest.

Acknowledgements
We thank Daan Nierboer, Dept of Public Health, Erasmus MC, for the technical support in creating the graphical illustrations presented in this article.

REFERENCES

1. Iezzoni LI, Shwartz M, Ash AS, Hughes JS, Daley J, Mackiernan YD. Severity measurement methods and judging hospital death rates for pneumonia. Med Care 1996;34(1):11-28.
2. Forster AJ, van Walraven C. The use of quality indicators to promote accountability in health care: the good, the bad, and the ugly. Open Med 2012;6(2):e75.
3. Haggar FA, Boushey RP. Colorectal cancer epidemiology: incidence, mortality, survival, and risk factors. Clin colon rectal Surg 2009;22(4):191.
4. Comprehensive cancer centre the Netherlands (integraal kankercentrum Nederland). Incidentie darmkanker. comprehensive cancer centre the Netherlands (integraal kankercentrum Nederland); Available from: http://www.cijfersoverkanker.nl/selecties/Incidentie_darmkanker/img534e81c492d09 (Accessed on 20.08.2014).
5. Schilling PL, Dimick JB, Birkmeyer JD. Prioritizing quality improvement in general surgery. J Am Coll Surg 2008;207(5):698-704.
6. Kirchhoff P, Clavien P-A, Hahnloser D. Complications in colorectal surgery: risk factors and preventive strategies. Patient Saf Surg 2010;4(5).
7. Krarup PM, Jorgensen LN, Andreasen AH, Harling H, Danish Colorectal Cancer G. A nationwide study on anastomotic leakage after colonic cancer surgery. Colorectal Dis 2012;14(10):e661-7.
8. Morris EJA, Taylor EF, Thomas JD, Quirke P, Finan PJ, Coleman MP, et al. Thirty-day postoperative mortality after colorectal cancer surgery in England. Gut 2011;60(6):806-13.
9. Birkmeyer JD, Sun Y, Wong SL, Stukel TA. Hospital volume and late survival after cancer surgery. Ann Surg 2007;245(5):777.
10. Nickelsen TN, Jørgensen T, Kronborg O. Thirty□day mortality after surgery for colorectal cancer in Denmark. Colorectal Dis 2005;7(5):500-6.
11. Friese CR, Earle CC, Silber JH, Aiken LH. Hospital characteristics, clinical severity, and outcomes for surgical oncology patients. Surgery 2010;147(5):602-9.
12. Lingsma HF, Dippel DWJ, 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 Psychiatr 2008;79(8):888-94.
13. van Gestel YRBM, Lemmens VEPP, Lingsma HF, de Hingh IHJT, Rutten HJT, Coebergh JWW. The hospital standardized mortality ratio fallacy: a narrative review. Med Care 2012;50(8):662-7.
14. Jarman B, Pieter D, van der Veen AA, Kool RB, Aylin P, Bottle A, et al. The hospital standardised mortality ratio: a powerful tool for Dutch hospitals to assess their quality of care? Qual Saf Health Care 2010;19(1):9-13.
15. Ayanian JZ. Using administrative data to assess health care outcomes. Eur heart J 1999;20(23):1689-91.
16. Glance LG, Osler TM, Mukamel DB, Dick AW. Impact of the present-on-admission indicator on hospital quality measurement: experience with the Agency for Healthcare Research and Quality (AHRQ) Inpatient Quality Indicators. Med Care 2008;46(2):112-9.
17. Iezzoni LI. Assessing quality using administrative data. Annals of internal medicine. 1997;127(8_Part_2):666-74.
18. Kolfschoten NE, Marang van de Mheen PJ, Gooiker GA, Eddes EH, Kievit J, Tollenaar R, et al. Variation in case-mix between hospitals treating colorectal cancer patients in the Netherlands. Eur J Surg Oncol 2011;37(11):956-63.
19. Khuri SF, Najjar SF, Daley J, Krasnicka B, Hossain M, Henderson WG, et al. Comparison of surgical outcomes between teaching and nonteaching hospitals in the Department of Veterans Affairs. Ann Surg 2001;234(3):370.
20. Glance LG, Dick A, Osler TM, Li Y, Mukamel DB. Impact of changing the statistical methodology on hospital and surgeon ranking: the case of the New York State cardiac surgery report card. Med Care 2006;44(4):311-9.
21. Steyerberg EW, Eijkemans MJ, Boersma E, Habbema JD. Applicability of clinical prediction models in acute myocardial infarction: a comparison of traditional and empirical Bayes adjustment methods. Am Heart J 2005 Nov;150(5):920.
22. Van Leersum NJ, Snijders HS, Henneman D, Kolfschoten NE, Gooiker GA, ten Berge MG, et al. The Dutch surgical colorectal audit. Eur J Surg Oncol 2013 Oct;39(10):1063-70.

23. Annual report. The dutch surgical colorectal audit 2013.
24. Steyerberg EW. Clinical prediction models. A practical approach to development, validation, and updating.: New York: Springer; 2009.
25. van der Sluis FJ, Espin E, Vallribera F, de Bock GH, Hoekstra HJ, van Leeuwen BL, et al. Predicting postoperative mortality after colorectal surgery: a novel clinical model. Colorectal Dis 2014;16(8):631-9.
26. Jorgensen ML, Young JM, Dobbins TA, Solomon MJ. Predictors of variation in colorectal cancer care and outcomes in New South Wales: a population-based health data linkage study. Med J Aust 2014;200(7):403-7.
27. van Walraven C, Wong J, Hawken S, Forster AJ. Comparing methods to calculate hospital-specific rates of early death or urgent readmission. Can Med Assoc J 2012;184(15):E810-E7.
28. Telem DA, Chin EH, Nguyen SQ, Divino CM. Risk factors for anastomotic leak following colorectal surgery: a case-control study. Arch Surg 2010;145(4):371-6.
29. Merkow RP, Kmiecik TE, Bentrem DJ, Winchester DP, Stewart AK, Ko CY, et al. Effect of including cancer-specific variables on models examining short-term outcomes. Cancer 2013;119(7):1412-9.
30. McNamee R. Regression modelling and other methods to control confounding. Occup Environ Med 2005;62(7):500-6.
31. Austin PC, Reeves MJ. The relationship between the C-statistic of a risk-adjustment model and the accuracy of hospital report cards: a Monte Carlo Study. Med Care 2013;51(3):275-84.
32. Austin PC, Reeves MJ. Effect of provider volume on the accuracy of hospital report cards: a Monte Carlo study. Circ Cardiovasc Qual Outcomes 2014;7(2):299-305.
33. Austin PC, Alter DA, Tu JV. The use of fixed- and random-effects models for classifying hospitals as mortality outliers: a Monte Carlo assessment. Med Decis Making 2003;23(6):526-39.
34. Dimick JB, Welch HG, Birkmeyer JD. Surgical mortality as an indicator of hospital quality: the problem with small sample size. JAMA 2004;292(7):847-51.
35. Bosch DJ, Pultrum BB, de Bock GH, Oosterhuis JK, Rodgers MGG, Plukker J. Comparison of different risk-adjustment models in assessing short-term surgical outcome after transthoracic esophagectomy in patients with esophageal cancer. Am J Surg 2011;202(3):303-9.
36. Mant J. Process versus outcome indicators in the assessment of quality of health care. Int J Qual Health Care 2001;13(6):475-80.
37. Matthiessen P, Hallbook O, Andersson M, Rutegard J, Sjodahl R. Risk factors for anastomotic leakage after anterior resection of the rectum. Colorectal Dis 2004;6(6):462-9.
38. Snijders HS, Henneman D, van Leersum NL, Ten Berge M, Fiocco M, Karsten TM, et al. Anastomotic leakage as an outcome measure for quality of colorectal cancer surgery. BMJ Qual Saf 2013;22(9):759-67.
39. Bruce J, Krukowski ZH, Al-Khairy G, Russell EM, Park KG. Systematic review of the definition and measurement of anastomotic leak after gastrointestinal surgery. Br J Surg 2001;88(9):1157-68.
40. Gunnarsson U, Seligsohn E, Jestin P, Påhlman L. Registration and validity of surgical complications in colorectal cancer surgery. Br J Surg 2003;90(4):454-9.
41. Almoudaris AM, Burns EM, Bottle A, Aylin P, Darzi A, Vincent C, et al. Single measures of performance do not reflect overall institutional quality in colorectal cancer surgery. Gut 2013;62(3):423-9.
42. Kolfschoten NE, Kievit J, Gooiker GA, van Leersum NJ, Snijders HS, Eddes EH, et al. Focusing on desired outcomes of care after colon cancer resections; hospital variations in 'textbook outcome'. Eur J Surg Oncol 2013;39(2):156-63.

Chapter 8.

The development of a case-mix correction model for comparing outcomes of oesophago-gastric cancer surgery

Claudia Fischer, Hester F Lingsma, Richard Hardwick, David Cromwell, Ewout W Steyerberg, Oliver Groene

This article is accepted.
Risk adjustment models for short-term outcomes after surgical resection for oesophago-gastric cancer patients.
British Journal of Surgery.

ABSTRACT

Background: Outcomes for oesophago-gastric cancer surgery are compared with the aim to benchmark quality of care. Adjusting for patient characteristics is crucial to avoid biased comparisons between providers. The study objective was to develop a case-mix adjustment model for comparing 30-, 90-day mortality and anastomotic leakage rates after oesophago-gastric (O-G) cancer resections.

Methods: The study reviewed existing models, considered expert opinion and examined audit data in order to select predictors that were consequently used to develop a case-mix adjustment model for the National Oesophago-Gastric Cancer Audit, covering England and Wales. Models were developed on patients undergoing surgical resection between April 2011 and March 2013 using logistic regression. Model calibration and discrimination was quantified using a bootstrap procedure.

Results: Most existing risk models for O-G resections were methodologically weak, out-dated or based on detailed laboratory data not generally available. In 4882 O-G cancer patients used for model development, 30-day mortality was 2.3%, 90-day mortality was 4.4% and 6.2% of patients developed an anastomotic leakage. The internally validated models, based on predictors selected from the literature, showed moderate discrimination (AUC 0.65 for 30-day mortality, 0.66 for 90-day mortality and 0.59 for anastomotic leakage) and good calibration.

Conclusions: Based on available data, three case mix adjustment models for postoperative outcomes in patients undergoing curative surgery for O-G cancer were developed. These models should be used for risk adjustment when assessing hospital performance in the NHS, and should be tested in other large health systems.

Keywords: oesophago-gastric cancer resection, case mix adjustment, 30-day mortality, 90-day mortality, anastomotic leakage

INTRODUCTION

As public interest in quality of hospital care is growing, outcome measures are increasingly used to benchmark hospital performance. When comparing outcomes between hospitals, risk adjustment for patient characteristics is crucial because when patient populations differ between hospitals, differences in outcome may represent differences in baseline risk rather than quality of care. Insufficient case-mix adjustment then leads to unfair comparisons. This is of particular relevance where surgery bears substantial risks, as in the case of O-G cancer resections.

The National Oesophago-Gastric Audit (NOGCA) was set up to monitor the quality of care provided to patients with O-G cancer in England and Wales, to evaluate care processes and patient outcomes [1]. A recent systematic review concluded, however, that current models for prediction of outcomes after oesophagectomy had numerous limitations in regarding methodology and clinical credibility [2]. Centralization of surgery, decision-making in multidisciplinary teams and improved care pathways have already been shown to contribute to a decrease in short-term mortality [3, 4], so that earlier prediction models might no longer be valid. The aim of the present study was to develop a case-mix adjustment model for comparisons of 30- and 90-day mortality, and anastomotic leak rates after resections for O-G cancer between NHS trusts, based on a review of existing prediction models, expert opinion and audit data.

METHODS

Data collection

The study used data submitted to the National Oesophago-Gastric (O-G) Cancer Auditfrom all 154 English NHS trusts that provide O-G cancer care and from all 13 Welsh NHS organisations contributing to the Welsh Cancer Information System (CANISC). The Audit included adults diagnosed with invasive, epithelial cancer of the oesophagus or stomach between 1 April 2011 and 31 March 2013, and captured information using a prospectively developed database on the patient (age at diagnosis, gender, comorbidities, Eastern Co-operative Oncology Group (ECOG) functional performance), cancer details (cancer site oesophagus including Siewert types I-III junctional tumours, or stomach), histology, TNM stage (Tumour, Node Metastasis) version 7 [5], American Society of Anesthesiologists (ASA) score, comorbidities and procedure (performance of neoadjuvant treatment, operation mode), as described previously[1]. All patients undergoing curative resection were included in the present study; those undergoing curative oncological treatment for squamous cell carcinoma and all palliative patients were excluded, (Appendix Figure A1).

Review of existing models

Potential prognostic factors for 30-day, 90-day mortality and anastomotic leak were selected on the basis of a review of the existing literature and clinical expert advice. Literature was searched for multivariable risk models of short-term mortality (30-, 90-days, or in-hospital mortality) or complications including anastomotic leaks following O-G cancer surgery. From studies meeting these inclusion criteria, risk factors included in the models that were available in routine clinical databases and not modifiable by the provider were selected (Appendix Table A1).

Outcome measures

The short-term outcomes were 30-day and 90-day all-cause postoperative mortality and anastomotic leak rates[6]. Date of death was obtained from the Office for National Statistics death certificate register. Anastomotic leak was defined as a severe disruption to the anastomosis (whether detected clinically or radiologically, and irrespective of whether it is managed conservatively or by re-operation) [7]. All leaks, including those from conduit staple lines away from the oesophago-gastric anastomosis, were included in the study based on self-reported data from the surgeon/surgical team.

Model development and statistical analysis

Potential predictors were tested initially in univariable logistic regression models. Variable categories containing small number were regrouped in advance (ASA score, co-morbidity count, predominant histology by cancer location, performance status, histology type). The linearity of the continuous independent variable age at diagnosis with 30-, 90-day mortality and anastomotic leakage was tested by adding quadratic terms. As this did not significantly improve the models, no quadratic terms were included in the model. To prevent exclusion of predictors with borderline significance, a p-value of 0.10 was used rather than 0.05 for inclusion of variables in the model. Decisions to include and exclude predictors were primarily based on the evidence gathered from the literature review and complemented with the information generated from the statistical analyses and expert clinical opinion. Odds ratios (OR) with 95% confidence intervals were used to express the strength of the predictive effects.

The model performance was assessed with respect to discrimination and calibration [8]. Discriminative ability represents how well the model was able to discriminate between patients with and without the outcome of interest, expressed as the area under the receiver operating curve (AUC, c-statistic) ranging from 0.5 – 1.0, where 0.5 indicates no discriminative power and 1.0 perfect discrimination. Calibration of the model was assessed by using scatter plots of observed versus predicted outcomes in deciles of predicted risk on the imputed data set.

The internal validity of the models was evaluated using a bootstrapping procedure [9]. With bootstrapping, multiple patient samples were drawn, considered as cases included under the same conditions as in the original data set. 800 bootstrap samples were used to re-estimate the multivariable logistic regression coefficients and consequently applied to the original dataset, resulting in 800 AUC statistics. The mean of these AUCs represented the optimism-corrected or internally validated AUC.

All analyses were performed in Stata and R. Missing data was assumed to be Missing at Random and was handled with the MICE (multiple imputation by chained equations) approach by White and Royston using Stata software (version 12 (StataCorp LP, College Station, Texas, USA)) [10]. Chained equations with 10 imputation sets were used. The outcome measures and the independent variables deprivation, age at diagnosis, ECOG performance status, ASA score, gender, tumour location, number of comorbidities, size and/or extent of the primary tumour (T stage from the TNM classification) and re-

gional lymph nodes (N stage of the TNM classification) were included in the imputation model. A sensitivity analysis comparing complete-case analysis with the one derived from the imputation model demonstrated no significant differences (Appendix Table A2, Table A3). The bootstrap procedure was performed with the *validate* function in the *rms* package in R statistical software, and the imputation with the MICE (multiple imputation by chained equations) approach by White and Royston using Stata software (version 12 (StataCorp LP, College Station, Texas, USA)) [10].

RESULTS

Published prognostic models

The literature search resulted in the identification of 41 prediction models for short-term outcomes after O-G cancer surgery. Some of the studies that we identified had a dual aim, i.e. providing insight in predictor effects and providing predictions based on the combination of predictors in a multivariable model. 34 models addressed postoperative mortality (12 studies used 30-day mortality, three used 90-day mortality, 16 used in-hospital mortality, 2 used major morbidity inclusive death, one postoperative mortality not further defined) and seven were predicting anastomotic leakages (AL) (Table 1). The majority of the studies considered outcomes after oesophagectomy and were designed as clinical prediction models as opposed to risk-adjustment models for provider comparisons. Numerous models were based on the POSSUM, O-POSSUM, P-POSSUM scores, a prediction score requiring detailed laboratory test values. These POSSUM scores are based on data not commonly available in audit data, such as white blood cell count or urea level [12, 17, 21, 26, 27, 31, 35-37]. In addition, the majority of the studies were based on single centre data that either pooled data over long periods of time [11, 13-17, 19-21, 23, 26, 31, 33, 34, 36, 37] or had a small sample size (e.g. N = 70, 121, 143, 204, 232) [11, 16, 20, 21, 37] and were performed in other countries than the UK. Event rates were typically far higher than those currently observed in the NOGCA, especially in the models developed in earlier years [19-22]. The predictive ability of most models was limited, at maximum, moderate [21, 22, 24, 30].

A detailed description of the predictors identified in the literature search and reasons for in-or exclusion is available in the appendix (Appendix Table A1).

Table 1. Descriptive information of currently available prediction models of O-G cancer short-term outcomes

AUTHOR (YEAR)	COUNTRY	PERIOD OF DATA COLLECTION OPERATION TYPE	NUMBER OF CENTRES	PATIENT NUMBER	REPORTED RATE OF OUTCOME EVENT RATE	DISCRIMINATION	
Law (1994) [33]	HKG	1982-92	Oesophageal	1	1105	In-hospital mortality	15.5%	/
Bartels (1998)[18]	GER	1982-5 (1996)	Oesophageal	1	432	30-day mortality	10% (1%)	/
Liu (2000) [19]	AUT	1994-7	Oesophageal	1	70	In-hospital mortality & complication	13%	/
Karl (2000) [10]	USA	1989-99	Ivor Lewis Gastro-Oesophageal	1	143	30-day mortality AL	2.1% 3.5%	/
Zafirellis (2002) [20]	UK	1990-9	Oesophageal	1	204	30-day mortality	12.7%	AUC=0.62 POSSUM
Bailey (2003) [21]	USA	1991-2000	Oesophageal	109	1777	30-day mortality	9.8%	c-index 0.69
McCulloch (2003) [11]	UK	1999-2002	Gastro-oesophageal	26	955	In-hospital mortality Surgical complications	12%	AUC=0.68 POSSUM AUC=0.71 POSSUM
Mariette (2004) [12]	FR	1982-93 (1994-2002)	Oesophageal	1	742	In-hospital mortality AL	5.4% (2.9%) 9.8% (2.2%)	/
Law (2004) [13]	HKG	1990-5	Oesophageal	1	421	In hospital mortality	1.1%	/
Atkins (2004) [22]	USA	1996-2002	Oesophageal	1	379	Operative mortality	5.8%	/
Tekkis (2004) [34]	UK	1994-2000	Gastro-oesophageal	36	1042	In-hospital mortality	12%	AUC=79.7 O-POSSUM AUC=74.6 P-POSSUM
Junemann-Ramirez (2004) [14]	UK	1992-9	Ivor Lewis gastro-oesophageal	1	276	AL	5.1%	/

Steyerberg (2006) [23]	USA/ NL	1991-1996	Oesophageal	Population database / clinical centre	3592	30-day mortality (in 4cohorts)	11% (10%, 7%,4%)	AUC=0.66
Viklund (2006) [24]	SWE	2001-3	Oesophageal	Nationwide study	275	30-day mortality & AL	3% 8%	/
Naga-bhushan (2007) [35]	UK	1990-2002	Gastro-oesophageal	1	313	30-day mortality	10.2%	AUC=0.61 O-POSSUM AUC=0.68 P-POSSUM
Lagarde (2007) [25]	NL	1993-2005	Oesophageal	1	663	In-hospital mortality	3.6%	AUC=0.60 O-POSSUM
Lai (2007) [26]	HKG	2001-5	Oesophageal	14	545	In-hospital mortality	5.5%	AUC=0.776 POSSUM AUC=0.776 P-POSSUM AUC=0.676 O-POSSUM
Ra (2008) [27]	USA	1997-2003	Oesophageal	Population database	1172	In-hospital mortality	14%	/
Wright (2009) [28]	USA	2002-7	Oesophageal	73 STS General Thoracic Data-base	2315	Major morbidity[c] (incl. death and AL)		/
Park (2009) [29]	UK	1995-2007	Oesophageal	ICNARC Case-mix Programme Database 181	7227	In-hospital mortality	11%	AUC=0.60 APACHE II [d] AUC=0.63 SAPSS II[e] AUC=0.65 ICNARC [f]
Dutta (2011) [36]	UK	2005-9	Gastro-oesophageal	1	121	30-day mortality	4%	AUC=0.759 POSSUM AUC=0.715 O-POSSUM
Bosch (2011) [31]	NL	1991-2007	Oesophageal	1	278	90-day mortality	5.4%	AUC=0.766 P-POSSUM AUC=0.756 O-POSSUM
Morita (2011) [32]	JPN	1964-79	Oesophageal	1	1106	In-hospital mortality	16.1%	/

Sunpaweravong (2012) [15]	THA	1998-2007	Oesophageal	1	232	30-day mortality AL	3.8% 15.9%	/
Noble (2012) [16]	UK	2005-10 2011	Oesophageal	1	258	AL AL major complication/ death	10%	AUC=0.801 Nun score[g] AUC=0.879 Nun score AUC=0.856 Nun score
Koppert (2012) [44]	NL	2005-9	Gastro-oesophageal	Eindhoven Cancer Registry	6223	30-day mortality	7.7%	/
Rutegard (2012) [31]	SW	2001-5	Oesophageal	Nationwide	559	90-day mortality	7.1%	/
Kassis (2013) [17]	USA	2001-11	Oesophageal	STS General Thoracic Database	7595	AL	10.6%	/

[b] Anastomotic leakage; [c] Including reoperation for bleeding, AL, pneumonia, re-intubation, ventilation beyond 48 hours, or death; [d] Acute Physiology and Chronic Health Evaluation; [e] Simplified Acute Physiology Score; [f]ICNARC physiology score; [g] Nun score calculated using the log-likelihood ratio of blood-borne variables of the systematic inflammatory response (albumin, WCC and CRP from POD4)

Patient characteristics

Of 22 766 patients identified the study included 4882 patients who had undergone O-G cancer resection in the period between April 2011 and March 2013 (Table 2). The patients had a mean age of 66 years (interquartile range=14 years) and the majority were male (74%). In 2135 (43.7%) patients, at least one comorbidity was present. Most patients had an adenocarcinoma histology (89%), while the most common location was the lower third of the oesophagus and Siewert type 1 tumour (39%). 30-day mortality was 2.3% (N=112) and 90-day mortality was 4.4% (N=216). 6.2% (N=305) of the patients developed an AL. Further descriptive information is shown in Table 2.

Table 2. Descriptive information on study population

Patient and prognostic information	No. of patients	%
Year of operation, 2013		
2012	2417	49.5
2013	2465	50.5
Age, years	4873	66.3*
Missing values	9	0.2
Comorbidity count		
No comorbidities	2747	56.3
One comorbidity	1311	26.8
Two comorbidities	566	11.6
Three or more comorbidities	258	5.3
Gender		
Male	3618	74.1
ECOG (WHO) performance status		
Carries out all normal activity	2519	51.6
Restricted but walks/does light work	1557	31.9
Walks, full self-care but no work	527	10.8
Limited self-care – fully disabled	120	2.5
Missing values	159	3.3
Size and/or extent of the primary tumour (T)		
No evidence of primary tumour (T0)	202	4.2
Tumour invades lamina propria or submucosa (T1)	929	19.0
Tumour invades muscularis propria (T2)	792	16.2
Tumour invades adventitia (T3)	2323	47.6
Tumour invades adjacent structures (T4)	490	10.0
Missing values	146	3.0
Regional lymph nodes (N)		
No regional lymph node metastasis (N0)	2143	43.9
Metastasis in 1 to 2 regional lymph nodes (N1)	1498	30.7
Metastasis in 3 to 6 (N2)	615	12.6
Metastasis in 7 or more (N3)	508	10.4
Missing values	118	2.4
ASA Scale		
Normal healthy patient	816	16.7
Mild systemic disease	2502	51.2
Severe systemic disease	1248	25.6
Life-threatening disease/ Moribund patient	60	1.2
Missing values	256	5.2

HISTOLOGY		
Adenocarcinoma	4336	88.8
Squamous cell carcinoma	420	8.6
Other carcinoma types	126	2.6
Predominant histology by cancer location		
Squamous cell carcinomas of the oesophagus	492	10.1
Adenocarcinomas of the upper and middle oesophagus	184	3.8
Adenocarcinomas of the lower third of the oesophagus and Siewert type 1 tumours	1906	39.0
Siewert type II and type III tumours	844	17.3
Tumours of the stomach	1456	29.8
Level of socio-economic deprivation (IMD quintile)		
1 Least deprived	840	17.2
2	860	17.6
3	846	17.3
4	800	16.4
5 Most deprived	746	15.3
Missing values	790	16.2
Patient outcomes		
Anastomotic leak	305	6.2
30-day postoperative mortality	112	2.3
90-day postoperative mortality	216	4.4

*Mean

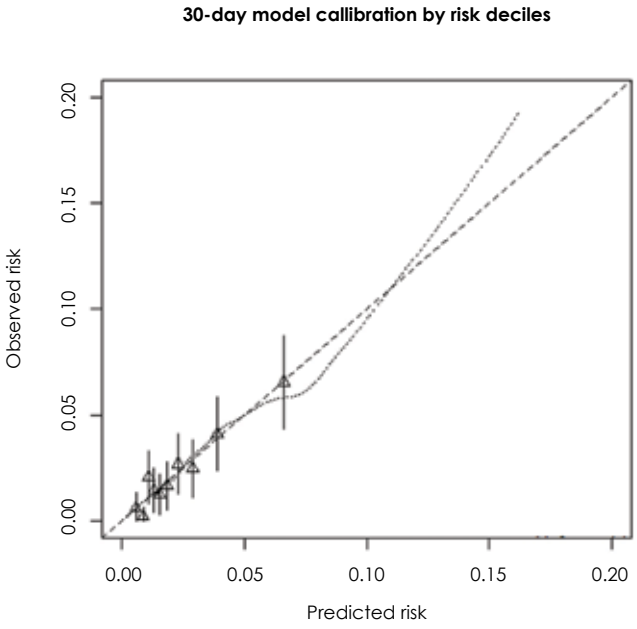
Model performance and validation

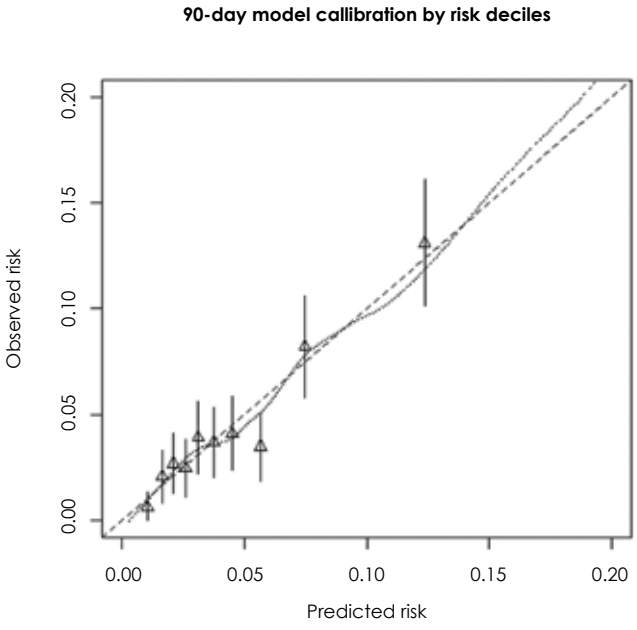
The models AUC was of our primary interest as they present the models predictive ability. The discriminative ability was moderate for the mortality models (AUC 0.70 for the 30-day mortality and AUC 0.69 for the 90-day mortality outcome) and somewhat lower for the AL model (AUC=0.63). Internally validated AUCs were 0.65 for the 30-day mortality model, 0.66 for the 90-day mortality model and 0.59 for the anastomotic leakage model, indicating some over fitting.

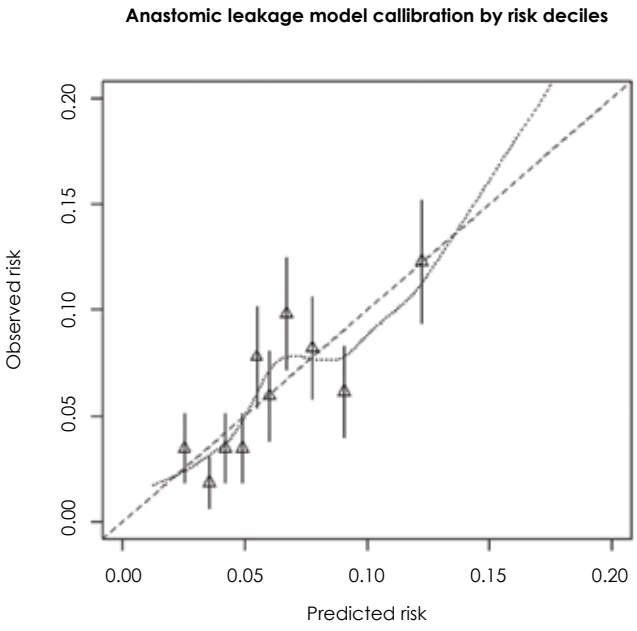
Model calibration

The scatter plots of predicted and observed probabilities showed that patients had an overall low risk for developing one of the three tested outcomes. For example, patients in the highest risk decile for developing an AL had a risk below 0.2 on average in the overall cohort. The difference between observed and predicted risk for developing an AL was smaller than 0.1.

Figure 1a,1b,1c. 30-day, 90-day mortality and anastomotic leakage model calibration by deciles of risk







Univariable analyses

In the following paragraphs odds ratio are presented to give an impression of the strength of the different predictors. However, our main aim is to give valid prediction and not valid estimates of the individual predictor effects.

The risk factor with the strongest association with all outcomes was the ASA grade (ASA grade 3 vs 1: 30-day mortality: OR=4.7 (95%CI=2.2-10); 90-day mortality: OR=5.0 (95%CI=2.8-8.8); AL: OR=1.4 (95%CI=1.0-2.0)). A greater number of comorbidities also increased the risk for all three outcomes (3 or more comorbidities vs.no comorbidities: 30-day mortality: OR 2.9 (95%CI: 1.5-5.6); 90-day mortality: OR 3.0 (95%CI: 1.8-4.8); AL: OR 1.7 (95%CI; 1.0-2.7)). Further, patients with an ECOG performance status of 3 or higher had a threefold risk of dying within 30- or 90-days compared to patients with an ECOG performance status of 0. In contrast, female gender and cancer located in the stomach compared to the oesophagus was associated with decreased risk of developing an AL (Table 3).

Table 3. Univariable logistic regression analyses for 30-day and 90-day mortality and anastomotic leakage

Predictor	30-day Mortality n=4882		90-day Mortality n=4882		Anastomotic leakage n=4882	
	OR	95% CI	OR	95% CI	OR	95% CI
Age per decade, years	**1.3***	**1.1-1.6**	**1.3**	**1.1-1.5**	1.0	0.9-1.1
Gender						
Male	1		1		1	
Female	0.8	0.5-1.2	0.7	0.5-1.0	**0.7**	**0.5-0.9**
Comorbidity count						
No comorbidities	1		1		1	
One comorbidity	**1.5**	**1.0-2.4**	**1.5**	**1.1-2.1**	**1.5**	**1.2-2.0**
Two comorbidities	**2.4**	**1.4-4.1**	**2.5**	**1.7-3.7**	**1.7**	**1.2-2.5**
Three or more comorbidities	**2.9**	**1.5-5.6**	**3.0**	**1.8-4.8**	**1.7**	**1.0-2.7**
ECOG (WHO) performance status						
Carries out all normal activity	1		1		1	
Restricted but walks/does light work	1.2	0.7-1.8	**1.3**	**1.0-1.9**	0.9	0.7-1.2
Walks, full self-care but no work	**1.7**	**1.0-3.0**	**2.1**	**1.4-3.1**	0.8	0.5-1.2
Limited self-care – fully disabled	**3.4**	**1.6-7.4**	**3.8**	**2.1-6.7**	1.1	0.5-2.2
ASA Scale						
Normal healthy patient	1		1		1	
Mild systemic disease	1.8	0.9-3.9	**2.3**	**1.3-4.0**	1.0	0.7-1.4
Severe systemic disease	**4.7**	**2.2-10.0**	**5.0**	**2.8-8.8**	**1.4**	**1.0-2.0**
Life-threatening disease/ Moribund patient	**7.1**	**2.1-24.4**	**8.7**	**3.5-21.6**	0.8	0.2-2.7
Predominant histology by cancer location						
Squamous cell carcinomas of the oesophagus	1		1		1	
Adenocarcinomas of the upper and middle oesophagus	0.9	0.3-2.7	0.7	0.3-1.7	0.5	0.3-1.1
Adenocarcinomas of the lower third of the oesophagus and Siewert type 1 tumours	0.9	0.5-1.6	1.0	0.6-1.6	**0.7**	**0.5-0.9**
Siewert type II and type III tumours	0.6	0.3-1.3	1.0	0.6-1.6	0.7	0.5-1.1
Tumours of the stomach	0.7	0.4-1.4	0.9	0.5-1.4	**0.4**	**0.3-0.6**
Histology						
Adenocarcinoma	1		1		1	
Squamous cell carcinoma	1.3	0.7-2.4	0.9	0.5-1.5	**1.5**	**1.0-2.2**
Other type	1.1	0.3-3.4	1.3	0.6-2.7	1.1	0.5-2.2
Size and extent of primary tumour (T)						
No evidence of primary tumour (T0)	1		1		1	
Tumour invades lamina propria or submucosa (T1)	0.6	0.2-1.4	0.7	0.3-1.5	1.0	0.5-1.8
Tumour invades muscularis propria (T2)	0.8	0.3-1.9	1.2	0.6-2.6	0.9	0.5-1.7
Tumour invades adventitia (T3)	0.6	0.3-1.4	0.9	0.5-1.9	0.9	0.5-1.6
Tumour invades adjacent structures (T4)	0.7	0.3-1.7	1.5	0.7-3.2	0.7	0.4-1.4
Regional lymph nodes (N)						
No regional lymph node metastasis (N0)	1		1		1	
Metastasis in 1 to 2 regional lymph nodes (N1)	1.2	0.8-1.8	**1.3**	**1.0-1.9**	1.0	0.8-1.3
Metastasis in 3 to 6 (N2)	0.8	0.4-1.5	1.4	0.9-2.2	0.8	0.6-1.2
Metastasis in 7 or more (N3)	1.0	0.5-2.0	**1.8**	**1.1-2.7**	0.9	0.6-1.3

Level of socio-economic deprivation (IMD quintile)						
1 least deprived	1		1		1	
2	0.8	0.5-1.4	0.7	0.4-1.1	0.9	0.6-1.3
3	0.6	0.3-1.1	0.8	0.5-1.2	0.8	0.6-1.2
4	0.8	0.4-1.4	0.8	0.5-1.3	**0.6**	**0.4-0.9**
5 most deprived	0.8	0.5-1.5	1.0	0.7-1.5	0.9	0.7-1.3

* Numbers in bold indicate significance

Multivariable analyses

Predictors with a p-value of <0.1 in the univariable data analysis for 30-day mortality were patient age at diagnosis, the number of comorbidities, ECOG performance status and ASA score. Furthermore, for 90-day mortality, outcome gender and regional lymph nodes (N) were identified as important predictors. For the anastomotic leakage model, the following predictors were chosen on basis of the univariable data analysis: gender, number of comorbidities, ASA score, histologic tumour type and tumour location. In consistency with previous studies and clinical expert opinion, the predictors gender, age, TNM stage, and ECOG performance status and predominant histology by cancer location and deprivation were entered into the multivariable models.

Table 4 presents the results for the multivariable case-mix adjustment models. For 30-day mortality, comorbidity count and ASA grade were the strongest predictors. A patient with an ASA grade of 4 or higher had an increase odds of 4.7 (95%CI 1.3-16.5) to die within 30-days compared to a patient with ASA grade 1. ASA grade was also the strongest predictor for the 90-day mortality outcome (ASA grade 4 or higher vs ASA grade 1 OR 5.1; 95% CI 2.0-13.3). Other predictors significantly associated with the mortality outcomes were: age at diagnosis, and the number of comorbidities.

The multivariable analysis for anastomotic leakage revealed that the number of comorbidities was strongly associated with the development of anastomotic leaks (3 or more comorbidities vs. no comorbidities OR=1.7; 95% CI 1.0-2.8). Further, patients with a tumour located in the stomach had a decreased of developing an AL (OR 0.4; 95% CI: 0.1-0.6).

The model equations are presented in table 5.

Table 4. Multivariable logistic regression for 30-day and 90-day mortality and anastomotic leakage

Predictor	30-day Mortality n=4882 ROC 0.698		90-day Mortality n=4882 ROC 0.694		Anastomotic leakage n=4882 ROC 0.631	
	Optimism corrected* ROC 0.646		Optimism corrected ROC 0.664		Optimism corrected ROC 0.587	
	OR	**95% CI**	**OR**	**95% CI**	**OR**	**95% CI**
Age per decade, years	1.2	1.0-1.5	1.2	1.0-1.4	0.9	0.8-1.1
Gender						
Male			1		1	
Female			0.7	0.5-1.1	**0.7**	**0.5-0.9**
Comorbidity count						
No comorbidities	1		1		1	
One comorbidity	1.3	0.8-2.1	1.3	0.9-1.9	**1.5**	**1.1-2.0**
Two comorbidities	**1.8****	**1.1-3.2**	**1.9**	**1.3-2.8**	**1.7**	**1.2-2.5**
Three or more comorbidities	**2.1**	**1.0-4.1**	**2.0**	**1.2-3.3**	**1.7**	**1.0-2.7**
ASA Grade						
I Normal healthy patient	1		1		1	
II Mild systemic disease	1.6	0.7-3.5	**1.9**	**1.1-3.4**	1.0	0.7-1.4
III Severe systemic disease	**3.5**	**1.6-7.8**	**3.5**	**1.9-6.3**	**1.4**	**1.0-2.1**
IV Life-threatening disease/Moribund patient	**4.7**	**1.3-16.5**	**5.1**	**2.0-13.3**	0.8	0.2-2.8
ECOG (WHO) performance status						
Carries out all normal activity	1		1		1	
Restricted but walks/does light work	0.9	0.6-1.4	1.1	0.8-1.5	0.9	0.7-1.2
Walks, full self-care but no work	1.3	0.7-2.2	**1.6**	**1.1-2.5**	0.8	0.5-1.2
Limited self-care – fully disabled	1.8	0.8-4.1	**2.3**	**1.3-4.3**	1.0	0.5-2.0
Size and/or extent of the primary tumour (T)						
No evidence of primary tumour (T0)	1		1		1	
Tumour invades lamina propria or submucosa (T1)	0.5	0.2-1.3	0.7	0.3-1.4	1.1	0.6-2.0
Tumour invades muscularis propria (T2)	0.7	0.3-1.8	1.1	0.5-2.3	1.0	0.6-2.0
Tumour invades adventitia (T3)	0.5	0.2-1.2	0.7	0.3-1.4	1.0	0.5-1.8
Tumour invades adjacent structures (T4)	0.6	0.2-1.9	1.1	0.5-2.6	1.0	0.5-2.1
Predominant histology by cancer location						
Squamous cell carcinomas of the oesophagus	1		1		1	
Adenocarcinomas of the upper and middle oesophagus	1.0	0.3-2.8	0.6	0.2-1.6	0.5	0.2-1.0
Adenocarcinomas of the lower third of the oesophagus and Siewert type 1 tumours	0.8	0.4-1.5	0.8	0.5-1.3	**0.6**	**0.4-0.8**
Siewert type II and type III tumours	0.5	0.2-1.1	0.7	0.4-1.2	**0.6**	**0.4-0.9**
Tumours of the stomach	0.5	0.3-1.0	**0.5**	**0.3-0.8**	**0.4**	**0.1-0.6**
Regional lymph nodes (N)						
No regional lymph node metastasis N(0)	1		1		1	
Metastasis in 1 to 2 regional lymph nodes N(1)	1.3	0.8-2.1	**1.4**	**1.0-2.0**	1.0	0.8-1.3
Metastasis in 3 to 6 N(3)	0.8	0.4-1.7	1.4	0.9-2.2	0.9	0.6-1.3
Metastasis in 7 or more N(4)	1.2	0.6-2.5	**1.8**	**1.1-2.9**	1.0	0.6-1.5

Deprivation						
1 Least deprived					1	
2					0.9	0.6-1.3
3					0.8	0.6-1.1
4					**0.6**	**0.4-0.9**
5 Most deprived					**0.9**	**0.6-1.3**

* ROC derived from bootstrapped sample (internal validation)
** Numbers in bold indicate significance

Table 5. Model equations for 30-day mortality, 90-day mortality and anastomotic leakage

Model	Equation
30-day mortality	Log(odds)= - 5.3205 + 0.0200 x (age) + 0.2984 x (one comorbidity) + 0.6168 x (two comorbidities) + 0.7318 x (three or more comorbidities) + 0.4760 x (ASA grade, mild systemic disease) + 1.2677 x (ASA grade, severe systemic disease) + 1.5399 x (ASA grade, life threatening disease/moribund patient) – 0.0971 x (ECOG performance status, restricted but walks/does light work) + 0.2315 x (ECOG performance status, walks, full self-care but no work) + 0.6159 x (ECOG performance status, limited self-care/fully disabled) - 0.6664 x (t, tumour invades lamina propria or submucosa) - 0.3077 x (t, tumour invades muscularis propria) - 0.6496 x (t, tumour invades adventitia) – 0.4202 x (t, tumour invades adjacent structures) + 0.2779 x (n, metastasis in 1 to 2 regional lymph nodes) – 0.1897 x (n, metastasis in 3 to 6) + 0.1920 x (n, metastasis in 7 or more) - 0.0238 x (tumour location, adenocarcinomas of the upper and middle oesophagus) - 0.1957 x (tumour location, adenocarcinomas of the lower third of the oesophagus and Siewert type I tumours) - 0.6097 x (tumour location, Siewert type II and type III tumours) - 0.6246 x (tumour location, tumours of the stomach)
90-day mortality	Log(odds)= - 4.8534 - 0.0152 x (age) - 0.2884 x (female gender) + 0.0963x (one comorbidity) + 0.6472 x (two comorbidities) + 0.7033 x (three or more comorbidities) + 0.6452 x (ASA grade, mild systemic disease) + 1.2431 x (ASA grade, severe systemic disease) + 1.6439 x (ASA grade, life-threatening disease/moribund patient) + 0.0963 x (ECOG performance status, restricted but walks/does light work) + 0.5003 x (ECOG performance status, walks, full self-care but no work) + 0.8491 x (ECOG performance status, limited self-care/fully disabled) – 0.4057 x (t, tumour invades lamina propria or submucosa) + 0.0802 x (t, tumour invades muscularis propria) – 0.3967 x (t, tumour invades adventitia) + 0.1470 x (t, tumour invades adjacent structures) + 0.3290 x (n, metastasis in 1 to 2 regional lymph nodes) + 0.3344 x (n, metastasis in 3 to 6) + 0.5829 x (n, metastasis in 7 or more) - 0.4601 x (tumour location, adenocarcinomas of the upper and middle oesophagus) – 0.1990 x (tumour location, adenocarcinomas of the lower third of the oesophagus and Siewert type 1 tumours) – 0.3851 x (tumour location, Siewert type II and type III tumours) - 0.6925 x (tumour location, tumours of the stomach)
Anastomotic leakage (AL)	Log(odds)= - 1.8702 – 0.0041 x (age) – 0.3540 x (female gender) + 0.4164 x (one comorbidity) + 0.5220 x (two comorbidities) + 0.5169 x (three or more comorbidities) - 0.0297 x (ASA grade, mild systemic disease) + 0.3451 x (ASA grade, severe systemic disease) – 0.1911 x (ASA grade, life-threatening disease/moribund patient) – 0.1031 x (ECOG performance status, restricted but walks/does light work) – 0.2274 x (ECOG performance status, walks, full self-care but no work) + 0.0049 x (ECOG performance status, limited self-care/fully disabled) + 0.0941 x (t, tumour invades lamina propria or submucosa) + 0.0571 x (t, tumour invades muscularis propria) - 0.0119 x (t, tumour invades adventitia) + 0.0411 x (t, tumour invades adjacent structures) + 0.0132 x (n, metastasis in 1 to 2 regional lymph nodes) – 0.1418 x (n, metastasis in 3 to 6) - 0.0188 x (n, metastasis in 7 or more) - 0.7059 x (tumour location, adenocarcinomas of the upper and middle oesophagus) – 0.5504 x (tumour location, adenocarcinomas of the lower third of the oesophagus and Siewert type 1 tumours) - 0.4800 x (tumour location, Siewert type II and type III tumours) - 0.9781 x (tumour location, tumours of the stomach) - 0.1101 x (deprivation 2) - 0.2117 x (deprivation 3) – 0.5143 x (deprivation 4) -0.0692 x (deprivation 5 most deprived)

DISCUSSION

This study developed models for case-mix adjustment of postoperative outcomes in oesophago-gastric (O-G) cancer patients undergoing curative resection. Our models are based on the largest contemporary patient cohort and exclusively based on data routinely available from the National Oesophago-Gastric Cancer Audit (NOCGA). Registries in other countries collect similar data items and may adopt the new risk models when pursuing obligatory outcome reporting and comparison between providers, as it is the case in the NHS.

ASA grade and the number of comorbidities were found to be the strongest predictors for both short-term mortality and anastomotic leakage (AL). This is in line with previous literature that identified severely ill patients being more likely to have an increased morbidity risk [22, 23, 29, 38]. Our three case-mix adjustment models, based on routinely available data in the NHS, had similar predictive ability to the ones found in the literature. While model performance might be improved by adding further clinical/laboratory based data items, we recommend against this for national comparisons. First, our review showed that the performance of models including complex clinical/laboratory data (such as the POSSUM score) differed substantially, and second, these clinical data elements are not routinely available in Cancer Registries or through the NOGCA database.

Other predictors identified in the literature include provider related variables such as choice of treatment and volume. But, as we aimed to develop a case-mix adjustment model to monitor outcomes between providers, factors that can be influenced by the provider, are not corrected for. For this reason, only those pre-operative factors were considered, which are found readily available in hospital databases and are not possible to be modified by the provider. The choice of variables might differ in a prognostic model which aims to predict risk in 'new' patients as opposed to a case-mix adjustment models which is usually used in retrospect on the data available. Taking into account patient characteristics that influence the postoperative outcome when comparing performance across providers is necessary to ensure that true differences in performance rather than differences in patient characteristics are being assessed [39]. Nevertheless, outcome differences must be interpreted with caution even after sufficient case-mix adjustment there might be remaining unmeasured confounders which influence the outcome.

Further, the question remains which indicator best reflects quality of surgical care. 30-days mortality rates are decreasing over time. While studies using data from the UK from 1990 and 2002 report an average postoperative 30-day mortality rate of 11.4% [21, 36], a study using data from the period 2005-2009 report a 4% 30-day mortality rate [37]. In our study, using data from April 2011 to March 2013, the 30-day mortality rate was 2.3%. While his is a positive development for clinical practice, 30-day mortality rates become less useful as quality indicators because the estimated mortality rates per hospital are based on smaller numbers of cases and hence more uncertain [39]. Rates of 90-day mortality are higher and research showed that the causes of death at 90-days after surgery are still strongly associated with surgical performance [40-42]. Deciding between measuring 30- or 90-day mortality can be regarded as a trade-off: with shorter follow-up, the included deaths will be mostly related to the surgery, but later deaths will be missed. While with a longer follow up period later deaths are included, potentially at the expense of including deaths unrelated to the surgery. Anastomotic leakages occur more frequently as well, which makes them attractive as quality indicator from a statistical point of view. The models for anastomotic leakage performed relatively poor. This is consistent with prior research which showed that postoperative complications are more difficult to predict on basis of patient characteristics than postoperative mortality [43]. This raises the hypothesis that their occurrence is determined by the quality of surgical care and to a lesser extent by patient characteristics. Thus, for several reasons anastomotic leakage rates seem a valuable quality indicator. However, judging hospital quality based on one indicator is a simplistic approach that should not be advocated. Monitoring several outcome and process indicators together will probably provide the most global picture on hospital performance. This is particularly true where the indicator is based on self-reported data (as in the case of anastomotic leak rates in this study, which should be interpreted in conjunction with return-to-theatre rates and intensive care utilization). Nevertheless, , when comparing outcomes, case-mix adjustment is of crucial importance to make valid comparisons and avoid risk adverse behavior. We therefore aimed to develop the best possible risk adjustment model, although we recognize that some residual confounding will always remain and that also adjusted mortality rates should still be interpreted with caution.

A major strength of this study is its large, national representative, population-based cohort. The use of audit data enabled the analysis of reliable, clinical case mix adjustment information and robust outcome ascertainment by linking to the Office of National Statistics mortality data. Future studies

should address other routinely available information possibly influencing patient outcomes. A potential limitation of this study is that missing data were observed for some key variables and that the coding of complications is subject to coding differences, and potentially under-reporting, between NHS trusts.

In conclusion, we developed well performing case mix adjustment models based on routinely available data for predicting postoperative short-term mortality following O-G cancer surgery. These can be used for the risk adjustment in the assessment of hospital performance in the NHS or other large health systems.

Acknowledgements
We like to acknowledge the health professionals and support personnel in English NHS trusts and Cancer networks who submitted the data to the Audit. We also like to acknowledge the advice of Dr Kate Walker regarding the analytical strategy applied in the development of the model.

Sources of funding
The National Oesophago-Gastric Cancer Audit is commissioned by the Healthcare Quality Improvement Partnership (HQIP) as part of the National Clinical Audit Programme (NCA). HQIP is led by a consortium of the Academy of Medical Royal Colleges, the Royal College of Nursing and National Voices. Its aim is to promote quality improvement, and in particular to increase the impact that clinical audit has on healthcare quality in England and Wales. HQIP holds the contract to manage and develop the NCA Programme, comprising more than 30 clinical audits that cover care provided to people with a wide range of medical, surgical and mental health conditions. The programme is funded by NHS England, the Welsh Government and, with some individual audits, also funded by the Health Department of the Scottish Government, DHSSPS Northern Ireland and the Channel Islands. This publication is based on data collected by or on behalf of the Healthcare Quality Improvement Partnership, who have no responsibility or liability for the accuracy, currency, reliability and/or correctness of this publication.

LITERATURE

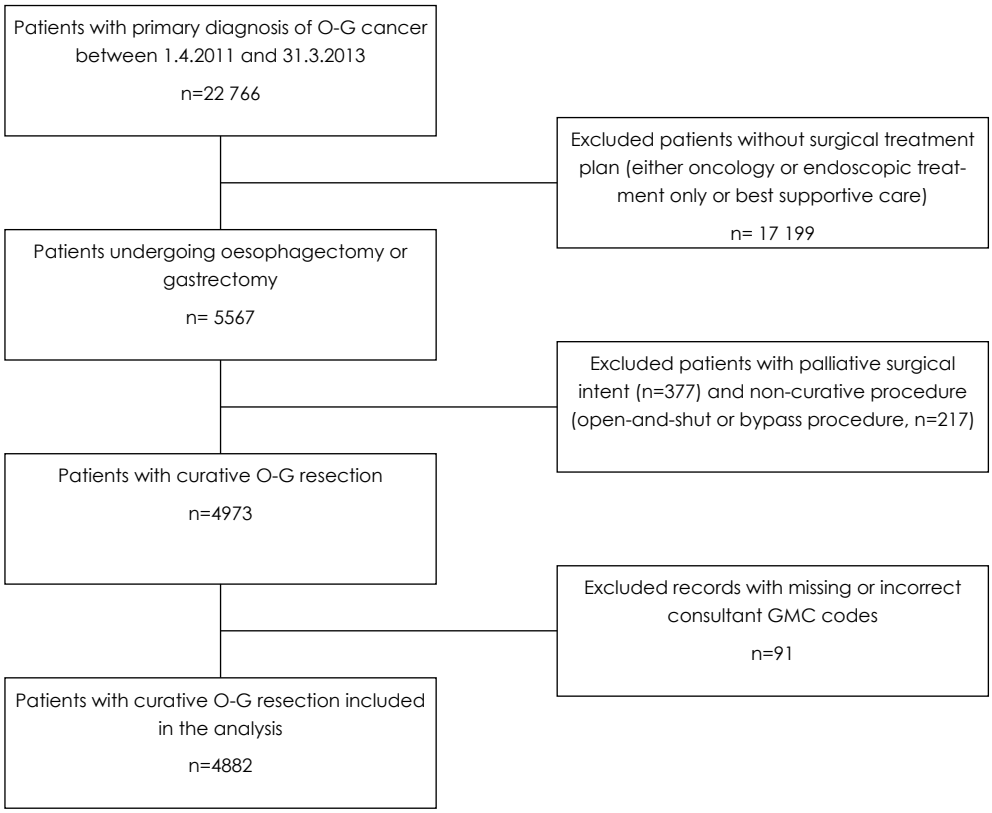
1. National Oesophago-Gastric Cancer Audit 2013. http://www.hscic.gov.uk/catalogue/PUB11093/clin-audi-supp-prog-oeso-gast-2013-rep.pdf [18.03.2014].
2. Findlay JM, Gillies RS, Sgromo B, Marshall RE, Middleton MR, Maynard ND. Individual Risk Modelling for Esophagectomy: A Systematic Review. J Gastrointest Surg 2014.
3. Damhuis RA, Wijnhoven BP, Plaisier PW, Kirkels WJ, Kranse R, van Lanschot JJ. Comparison of 30-day, 90-day and in-hospital postoperative mortality for eight different cancer types. The British journal of surgery 2012;99(8): 1149-1154.
4. Jamieson GG, Mathew G, Ludemann R, Wayman J, Myers JC, Devitt PG. Postoperative mortality following oesophagectomy and problems in reporting its rate. The British journal of surgery 2004;91(8): 943-947.
5. Sobin LH GM, Wittekind C (editors). The TNM Classification of Malignant Tumours 7Th edition. Wiley-Blackwell 2009.
6. Fernandez FG, Meyers BF. Quality of life after esophagectomy. Seminars in thoracic and cardiovascular surgery 2004;16(2): 152-159.
7. The Natioanl Oesophago-Gastric Cancer Audit Data Manual. Version 1.6. http://www/hscic.gov.uk/og 24 November 2014.
8. Steyerberg EW. Clinical prediction models. A practical approach to development, validation, and updating. New York: Springer, 2009.
9. Steyerberg EW, Harrell FE, Jr., Borsboom GJ, Eijkemans MJ, Vergouwe Y, Habbema JD. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. Journal of clinical epidemiology 2001;54(8): 774-781.
10. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. Stat Med 2011;30(4): 377-399.
11. Karl RC, Schreiber R, Boulware D, Baker S, Coppola D. Factors affecting morbidity, mortality, and survival in patients undergoing Ivor Lewis esophagogastrectomy. Annals of surgery 2000;231(5): 635-643.
12. McCulloch P, Ward J, Tekkis PP, surgeons Ago, British Oesophago-Gastric Cancer G. Mortality and morbidity in gastro-oesophageal cancer surgery: initial results of ASCOT multicentre prospective cohort study. Bmj 2003;327(7425): 1192-1197.
13. Mariette C, Taillier G, Van Seuningen I, Triboulet JP. Factors affecting postoperative course and survival after en bloc resection for esophageal carcinoma. The Annals of thoracic surgery 2004;78(4): 1177-1183.
14. Law S, Wong KH, Kwok KF, Chu KM, Wong J. Predictive factors for postoperative pulmonary complications and mortality after esophagectomy for cancer. Annals of surgery 2004;240(5): 791-800.
15. Junemann-Ramirez M, Awan MY, Khan ZM, Rahamim JS. Anastomotic leakage post-esophagogastrectomy for esophageal carcinoma: retrospective analysis of predictive factors, management and influence on longterm survival in a high volume centre. Eur J Cardiothorac Surg 2005;27(1): 3-7.
16. Sunpaweravong S, Ruangsin S, Laohawiriyakamol S, Mahattanobon S, Geater A. Prediction of major postoperative complications and survival for locally advanced esophageal carcinoma patients. Asian journal of surgery / Asian Surgical Association 2012;35(3): 104-109.
17. Noble F, Curtis N, Harris S, Kelly JJ, Bailey IS, Byrne JP, Underwood TJ, South Coast Cancer C-O-G. Risk assessment using a novel score to predict anastomotic leak and major complications after oesophageal resection. J Gastrointest Surg 2012;16(6): 1083-1095.
18. Kassis ES, Kosinski AS, Ross P, Jr., Koppes KE, Donahue JM, Daniel VC. Predictors of Anastomotic Leak After Esophagectomy: An Analysis of The Society of Thoracic Surgeons General Thoracic Database. The Annals of thoracic surgery 2013.
19. Bartels H, Stein HJ, Siewert JR. Preoperative risk analysis and postoperative mortality of oesophagectomy for resectable oesophageal cancer. The British journal of surgery 1998;85(6): 840-844.
20. Liu JF, Watson DI, Devitt PG, Mathew G, Myburgh J, Jamieson GG. Risk factor analysis of post-operative mortality in oesophagectomy. Diseases of the esophagus : official journal of the International Society for Diseases of the Esophagus / ISDE 2000;13(2): 130-135.

21. Zafirellis KD, Fountoulakis A, Dolan K, Dexter SP, Martin IG, Sue-Ling HM. Evaluation of POSSUM in patients with oesophageal cancer undergoing resection. The British journal of surgery 2002;89(9): 1150-1155.
22. Bailey SH, Bull DA, Harpole DH, Rentz JJ, Neumayer LA, Pappas TN, Daley J, Henderson WG, Krasnicka B, Khuri SF. Outcomes after esophagectomy: a ten-year prospective cohort. The Annals of thoracic surgery 2003;75(1): 217-222; discussion 222.
23. Atkins BZ, Shah AS, Hutcheson KA, Mangum JH, Pappas TN, Harpole DH, Jr., D'Amico TA. Reducing hospital morbidity and mortality following esophagectomy. The Annals of thoracic surgery 2004;78(4): 1170-1176; discussion 1170-1176.
24. Steyerberg EW, Neville BA, Koppert LB, Lemmens VE, Tilanus HW, Coebergh JW, Weeks JC, Earle CC. Surgical mortality in patients with esophageal cancer: development and validation of a simple risk score. Journal of clinical oncology : official journal of the American Society of Clinical Oncology 2006;24(26): 4277-4284.
25. Viklund P, Lindblad M, Lu M, Ye W, Johansson J, Lagergren J. Risk factors for complications after esophageal cancer resection: a prospective population-based study in Sweden. Annals of surgery 2006;243(2): 204-211.
26. Lagarde SM, Maris AK, de Castro SM, Busch OR, Obertop H, van Lanschot JJ. Evaluation of O-POSSUM in predicting in-hospital mortality after resection for oesophageal cancer. The British journal of surgery 2007;94(12): 1521-1526.
27. Lai F, Kwan TL, Yuen WC, Wai A, Siu YC, Shung E. Evaluation of various POSSUM models for predicting mortality in patients undergoing elective oesophagectomy for carcinoma. The British journal of surgery 2007;94(9): 1172-1178.
28. Ra J, Paulson EC, Kucharczuk J, Armstrong K, Wirtalla C, Rapaport-Kelz R, Kaiser LR, Spitz FR. Postoperative mortality after esophagectomy for cancer: development of a preoperative risk prediction model. Annals of surgical oncology 2008;15(6): 1577-1584.
29. Wright CD, Kucharczuk JC, O'Brien SM, Grab JD, Allen MS, Society of Thoracic Surgeons General Thoracic Surgery D. Predictors of major morbidity and mortality after esophagectomy for esophageal cancer: a Society of Thoracic Surgeons General Thoracic Surgery Database risk adjustment model. The Journal of thoracic and cardiovascular surgery 2009;137(3): 587-595; discussion 596.
30. Park DP, Welch CA, Harrison DA, Palser TR, Cromwell DA, Gao F, Alderson D, Rowan KM, Perkins GD. Outcomes following oesophagectomy in patients with oesophageal cancer: a secondary analysis of the ICNARC Case Mix Programme Database. Critical care 2009;13 Suppl 2: S1.
31. Bosch DJ, Pultrum BB, de Bock GH, Oosterhuis JK, Rodgers MG, Plukker JT. Comparison of different risk-adjustment models in assessing short-term surgical outcome after transthoracic esophagectomy in patients with esophageal cancer. American journal of surgery 2011;202(3): 303-309.
32. Rutegard M, Lagergren P, Rouvelas I, Lagergren J. Intrathoracic anastomotic leakage and mortality after esophageal cancer resection: a population-based study. Annals of surgical oncology 2012;19(1): 99-103.
33. Morita M, Nakanoko T, Fujinaka Y, Kubo N, Yamashita N, Yoshinaga K, Saeki H, Emi Y, Kakeji Y, Shirabe K, Maehara Y. In-hospital mortality after a surgical resection for esophageal cancer: analyses of the associated factors and historical changes. Annals of surgical oncology 2011;18(6): 1757-1765.
34. Law SY, Fok M, Wong J. Risk analysis in resection of squamous cell carcinoma of the esophagus. World journal of surgery 1994;18(3): 339-346.
35. Tekkis PP, McCulloch P, Poloniecki JD, Prytherch DR, Kessaris N, Steger AC. Risk-adjusted prediction of operative mortality in oesophagogastric surgery with O-POSSUM. The British journal of surgery 2004;91(3): 288-295.
36. Nagabhushan JS, Srinath S, Weir F, Angerson WJ, Sugden BA, Morran CG. Comparison of P-POSSUM and O-POSSUM in predicting mortality after oesophagogastric resections. Postgraduate medical journal 2007;83(979): 355-358.
37. Dutta S, Al-Mrabt NM, Fullarton GM, Horgan PG, McMillan DC. A comparison of POSSUM and GPS models in the prediction of post-operative outcome in patients undergoing oesophago-gastric cancer resection. Annals of surgical oncology 2011;18(10): 2808-2817.
38. Jaroni JL, Wright SM, Lerman C, Epstein LH. Relationship between education and delay discounting in smokers. Addict Behav 2004;29(6): 1171-1175.

39. Lingsma HF, Steyerberg EW, Eijkemans MJ, Dippel DW, Scholte Op Reimer WJ, Van Houwelingen HC, Netherlands Stroke Survey I. Comparing and ranking hospitals based on outcome: results from The Netherlands Stroke Survey. QJM 2010;103(2): 99-108.
40. Talsma AK, Lingsma HF, Steyerberg EW, Wijnhoven BP, Van Lanschot JJ. The 30-day versus in-hospital and 90-day mortality after esophagectomy as indicators for quality of care. Annals of surgery 2014;260(2): 267-273.
41. Rutegard M, Lagergren P, Johar A, Lagergren J. Time Shift in Early Postoperative Mortality After Oesophagectomy for Cancer. Annals of surgical oncology 2015.
42. McMillan RR, Berger A, Sima CS, Lou F, Dycoco J, Rusch V, Rizk NP, Jones DR, Huang J. Thirty-day mortality underestimates the risk of early death after major resections for thoracic malignancies. The Annals of thoracic surgery 2014;98(5): 1769-1774; discussion 1774-1765.
43. Snijders HS, Henneman D, van Leersum NL, ten Berge M, Fiocco M, Karsten TM, Havenga K, Wiggers T, Dekker JW, Tollenaar RA, Wouters MW. Anastomotic leakage as an outcome measure for quality of colorectal cancer surgery. BMJ Qual Saf 2013;22(9): 759-767.
44. Bosch DJ, Pultrum BB, de Bock GH, Oosterhuis JK, Rodgers MGG, Plukker J. Comparison of different risk-adjustment models in assessing short-term surgical outcome after transthoracic esophagectomy in patients with esophageal cancer. The American Journal of Surgery 2011;202(3): 303-309.
45. Koppert LB, Lemmens VE, Coebergh JW, Steyerberg EW, Wijnhoven BP, Tilanus HW, Janssen-Heijnen ML. Impact of age and co-morbidity on surgical resection rate and survival in patients with oesophageal and gastric cancer. The British journal of surgery 2012;99(12): 1693-1700.
46. Risk prediction in surgery. http://www.riskprediction.org.uk/background.php [18.03.2014].

APPENDIX

Appendix 1. Flow chart patient inclusion process



Appendix 2. Summary of in-/excluded predictors for postoperative mortality (30-day, 90-day and in-hospital mortality) identified by literature review

	Risk predictors for 30 and 90 day mortality					
Considered for inclusion in model?	**Patient characteristics**	**Comorbidities**	**Tumour characteristics**	**Treatment process**	**Serum levels**	**Other**
Yes	Age; Patient performance score; ASA rating;	Comorbidity count Congestive heart failure/ peripheral vascular disease/ cardiac disease; Pulmonary comorbidity; (Insulin dependent) Diabetes; Renal comorbidity;	TNM stage;			
No: Can be influenced by provider			Urgency of operation;	Neoadjuvant therapy; Amount of blood loss; Incomplete resection; Type of operation; Postoperative pulmonary complications; Pneumonia Need for transfusion		AL[a] Surgeon's assessment on patients fit for surgery; Worse swallowing score;

No: Not routinely available in clinical datasets	Alcohol consumption; (History of previous) Smoking; Race; Steroid use; Mid-arm circumference; Number of stairs climbed;	Charlson score; Peripheral vascular disease; Coronary heart disease~ Coronary artery disease; Hypertension; Hepatic disease; Ascites;			Forced expiratory volume in 1 second <60%; Alkaline phosphatase level more than 125 U/L; FEV1/FVC; Physiological measurements on admission to critical care: Partial pressure of arterial oxygen (PaO2): fraction of inspired oxygen (FiO2) ratio; Lowest arterial pH; Creatinine; Serum albumin; Urea; Mechanical ventilation; Incentive spirometry; Poor cardiac, respiratory, hepatic function;	POSSUM;[b] P-POSSUM; [c] O-POSSUM;[d]
No: Not applicable for this study						Hospital volume; Palliative resection; Year of operation;

[a] Anastomotic leakage

[b] POSSUM (Physiological and Operative Severity Score for the enumeration of mortality and morbidity) includes the following variables: age(y), cardiac history, respiratory history, blood pressure, pulse rate, Glasgow coma score, haemoglobin (g/%), white cell count ($X10^{12}$/L), urea, plasma sodium (mmol/l), plasma potassium (mmol/l), electrocardiogram, operative severity, multiple procedures, total blood loss (ml), peritoneal soiling, presence of malignancy, mode of surgery [45]

[c] P-POSSUM (Portsmouth-modified Physiological and Operative Severity Score for the enumeration of Mortality and morbidity) includes the following variables: age (y), Glasgow Coma Score, cardiac signs, respiratory signs, electrocardiography, systolic pressure (mm Hg), pulse rate(beats/min), haemoglobin level (g/dL), white blood cell count ($X10^{12}$/L), urea level (mmol/L), sodium level(mmol/L), potassium level(mmol/L), surgical severity, multiple procedures, total blood loss, peritoneal soiling, presence of malignancy, mode of surgery [45]

[d] O-POSSUM (Physiological and Operative Severity Score for the enumeration of Mortality and Morbidity Oesophagogastric surgery) includes the following variables: age (y), Glasgow Coma Score, cardiac signs, respiratory signs, electrocardiography, systolic pressure (mm Hg), pulse rate(beats/min), haemoglobin level (g/dL), white blood cell count ($X10^{12}$/L), urea level (mmol/L), sodium level(mmol/L), potassium level(mmol/L), surgical severity, multiple procedures, mode of surgery [45]

Appendix 3. Summary of in-/excluded predictors for postoperative complication/ anastomotic leakage identified by literature review

	Risk predictors for anastomotic leakage					
Considered for inclusion in model?	**Patient characteristics**	**Comorbidities**	**Tumour characteristics**	**Treatment process**	**Serum levels**	**Other**
Yes	Age; ASA rating; Decreased functional status; Gender;	(Congestive) Heart failure; diabetes; Copd; [a] (Insulin dependent) Diabetes; Coronary (artery) disease;	**Tumour stage;**			
No: Can be influenced by provider				Surgical procedure type; Additional organ resection; Procedure duration; Blood transfusion; Operation time;		
No: Not routinely available in clinical datasets	Race; Smoking status; Steroid use; Low BMI; [b] Obesity;	Hypertension; (Peripheral) Vascular disease; Dyspnoea; Coronary disease; Renal insufficiency;		Forced expiratory volume in 1second <60% of predicted; Alkaline phosphatase level of more than 125 U/L; Lower serum albumin concentration; WCC (white cell count); Post-operative CRP (C reactive protein); FEV1/FVC;		POSSUM[c] Increased complexity score?
No: Not applicable of this study						Year of operation;

[a]Chronic obstructive pulmonary disease

[b]Body mass index

[c] POSSUM (Physiological and Operative Severity Score for the enumeration of mortality and morbidity) includes the following variables: age(y), cardiac history, respiratory history, blood pressure, pulse rate, Glasgow coma score, haemoglobin (g/%), white cell count ($X10^{12}$/L), urea, plasma sodium (mmol/l), plasma potassium (mmol/l), electrocardiogram, operative severity, multiple procedures, total blood loss (ml), peritoneal soiling, presence of malignancy, mode of surgery [45]

Appendix 4. Descriptive statistics in the complete case analysis and in the imputed dataset

	Complete dataset		Imputed dataset	
Year of operation				
2012	2417	49.5	2417	49.5
2013	2465	50.5	2465	50.5
Age, years	4873	66.3*	4882	66.3*
Missing values	9	0.2		
Comorbidity count				
No comorbidities	2747	56.3	2747	56.3
One comorbidity	1311	26.8	1311	26.8
Two comorbidities	566	11.6	566	11.6
Three or more comorbidities	258	5.3	258	5.3
Gender				
Male	3 618	74.1	3 618	74.1
ECOG (WHO) performance status				
Carries out all normal activity	2519	51.6	2601	53.3
Restricted but walks/does light work	1557	31.9	1611	33.0
Walks, full self-care but no work	527	10.8	543	11.1
Limited self-care – fully disabled	120	2.5	127	2.6
Missing values	159	3.3		
Size and /or extent of the primary tumour (T)				
No evidence of primary tumour T(0)	202	4.2	205	4.2
Tumour invades lamina propria or submucosa T(1)	929	19.0	957	19.6
Tumour invades muscularis propria T(2)	792	16.2	820	16.8
Tumour invades adventitia T(3)	2323	47.6	2389	48.9
Tumour invades adjacent structures T(4)	490	10.0	511	10.5
Missing values	146	3.0		
Regional lymph nodes (N)				
No regional lymph node metastasis N(0)	2143	43.9	2182	44.7
Metastasis in 1 to 2 regional lymph nodes N(1)	1498	30.7	1544	31.6
Metastasis in 3 to 6 N(2)	615	12.6	634	13.0
Metastasis in 7 or more N(3)	508	10.4	522	10.7
Missing values	118	2.4		

ASA Scale				
Normal healthy patient	816	16.7	866	17.7
Mild systemic disease	2502	51.2	2651	54.3
Severe systemic disease	1248	25.6	1301	26.6
Life-threatening disease/Moribund patient	60	1.2	64	1.3
Missing values	256	5.2		
Histology				
Adenocarcinoma	4336	88.8	4336	88.8
Squamous cell carcinoma	420	8.6	420	8.6
Other carcinoma types	126	2.6	126	2.6
Predominant histology by cancer location				
Squamous cell carcinomas of the oesophagus	492	10.1	492	10.1
Adenocarcinomas of the upper and middle oesophagus	184	3.8	184	3.8
Adenocarcinomas of the lower third of the oesophagus and	1906	39.0	1906	39.0
Siewert type 1 tumours				
Siewert type II and type III tumours	844	17.3	844	17.3
Tumours of the stomach	1456	29.8	1456	29.8
Deprivation				
1 Least deprived	840	17.2	999	20.5
2	860	17.6	1047	21.4
3	846	17.3	999	20.5
4	800	16.4	942	19.3
5 Most deprived	746	15.3	895	18.3
Missing values	790	16.2		
Patient outcomes				
Anastomotic leak	305	6.2	305	6.2
30-day postoperative mortality	112	2.3	112	2.3
90-day postoperative mortality	216	4.4	216	4.4

*Mean

Appendix 5. Univariable analysis in the complete case analysis and in the imputed dataset

Predictor	Original dataset						Imputed dataset					
	30-day Mortality		90-day Mortality		Anastomotic leakage		30-day Mortality		90-day Mortality		Anastomotic leakage	
	OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI
Age per decade, years	**1.3***	**1.1-1.6**	**1.3**	**1.1-1.5**	1.0	0.9-1.1	**1.3****	**1.1-1.6**	**1.3**	**1.1-1.5**	1.0	0.9-1.1
Gender												
Female	0.8	0.5-1.2	0.7	0.5-1.0	**0.7**	**0.5-0.9**	0.8	0.5-1.2	0.7	0.5-1.0	**0.7**	**0.5-0.9**
Comorbidity count												
No comorbidities	1		1		1		**1**		**1**		**1**	
One comorbidity	**1.5**	**1.0-2.4**	**1.5**	**1.1-2.1**	**1.5**	**1.2-2.0**	**1.5**	**1.0-2.4**	**1.5**	**1.1-2.1**	**1.5**	**1.7-2.0**
Two comorbidities	**2.4**	**1.4-4.1**	**2.5**	**1.7-3.7**	**1.7**	**1.2-2.5**	**2.4**	**1.4-4.1**	**2.5**	**1.7-3.7**	**1.7**	**1.2-2.5**
Three or more comorbidities	**2.9**	**1.5-5.6**	**3.0**	**1.8-4.8**	**1.7**	**1.0-2.7**	**2.9**	**1.5-5.6**	**3.0**	**1.8-4.8**	**1.7**	**1.0-2.7**
ECOG (WHO) performance status												
Carries out all normal activity	1		1		1		1		1		1	
Restricted but walks/ does light work	1.2	0.7-1.8	1.3	0.9-1.8	0.9	0.7-1.2	1.2	0.7-1.8	**1.3**	**1.0-1.9**	0.9	0.7-1.2
Walks, full self-care but no work	1.6	0.9-2.9	**2.1**	**1.4-3.1**	0.8	0.5-1.1	**1.7**	**1.0-3.0**	**2.1**	**1.4-3.1**	0.8	0.5-1.2
Limited self-care – fully disabled	**3.7**	**1.7-8.1**	**3.7**	**2.0-6.8**	1.1	0.6-2.3	**3.4**	**1.6-7.4**	**3.8**	**2.1-6.7**	1.1	0.5-2.2
ASA Scale												
Normal healthy patient	1		1		1		1		1		1	
Mild systemic disease	1.9	0.8-4.2	**2.2**	**1.2-4.0**	1.0	0.7-1.4	1.8	0.9-3.9	**2.3**	**1.3-4.0**	1.0	0.7-1.4
Severe systemic disease	**5.0**	**2.3-11.1**	**5.0**	**2.8-9.0**	**1.4**	**1.0-2.0**	**4.7**	**2.2-10.0**	**5.0**	**2.8-8.8**	**1.4**	**1.0-2.0**
Life-threatening disease/ Moribund patient	**8.2**	**2.3-29.0**	**9.5**	**3.8-23.9**	0.8	0.2-2.8	**7.1**	**2.1-24.4**	**8.7**	**3.5-21.6**	0.8	0.2-2.7
Predominant histology by cancer location												
Squamous cell carcinomas of the oesophagus	1		1		1		1		1		1	
Adenocarcinomas of the upper and middle oesophagus	0.9	0.3-2.7	0.7	0.3-1.7	0.5	0.3-1.1	0.9	0.3-2.7	0.7	0.3-1.7	0.5	0.3-1.1
Adenocarcinomas of the lower third of the oesophagus and Siewert type 1 tumours	0.9	0.5-1.6	1.0	0.6-1.6	**0.7**	**0.5-0.9**	0.9	0.5-1.6	1.0	0.6-1.6	**0.7**	**0.5-0.9**
Siewert type II and type III tumours	0.6	0.3-1.3	1.0	0.6-1.6	0.7	0.5-1.1	0.6	0.3-1.3	1.0	0.6-1.6	0.7	0.5-1.1
Tumours of the stomach	0.7	0.4-1.4	0.9	0.5-1.4	**0.4**	**0.3-0.6**	0.7	0.4-1.4	0.9	0.5-1.4	**0.4**	**0.3-0.6**
Histology												
Adenocarcinoma	1		1		1		1		1		1	
Squamous cell	1.3	0.7-2.4	0.9	0.5-1.5	**1.5**	**1.0-2.2**	1.3	0.7-2.4	0.9	0.5-1.5	**1.5**	**1.0-2.2**
Other carcinoma type	1.1	0.3-3.4	1.3	0.6-2.7	1.1	0.5-2.2	1.1	0.3-3.4	1.3	0.6-2.7	1.1	0.5-2.2

Size and/or extent of the primary tumour (T)												
No evidence of primary tumour T(0)	1		1		1		1		1		1	
Tumour invades lamina propria or submucosa T(1)	0.6	0.2-1.4	0.7	0.3-1.7	1.0	0.5-1.8	0.6	0.2-1.4	0.7	0.3-1.5	1.0	0.5-1.8
Tumour invades muscularis propria T(2)	0.8	0.3-1.9	1.2	0.6-1.6	0.9	0.5-1.6	0.8	0.3-1.9	1.2	0.6-2.6	0.9	0.5-1.7
Tumour invades adventitia T(3)	0.6	0.3-1.4	0.9	0.6-1.6	0.9	0.5-1.6	0.6	0.3-1.4	0.9	0.5-1.9	0.9	0.5-1.6
Tumour invades adjacent structures T(4)	0.7	0.3-1.8	1.5	0.5-1.4	0.7	0.4-1.5	0.7	0.3-1.7	1.5	0.7-3.2	0.7	0.4-1.4
Regional lymph nodes (N)												
No regional lymph node metastasis N(0)	1		1		1		1		1		1	
Metastasis in 1 to 2 regional lymph nodes N(1)	1.2	0.8-1.9	**1.3**	**1.0-1.9**	1.0	0.8-1.3	1.2	0.8-1.8	**1.3**	**1.0-1.9**	1.0	0.8-1.3
Metastasis in 3 to 6 N(2)	0.7	0.4-1.4	1.4	0.9-2.1	0.8	0.6-1.2	0.8	0.4-1.5	1.4	0.9-2.2	0.8	0.6-1.2
Metastasis in 7 or more N(3)	1.0	0.5-1.9	**1.7**	**1.1-2.7**	0.8	0.6-1.3	1.0	0.5-2.0	**1.8**	**1.1-2.7**	0.9	0.6-1.3
Deprivation												
1 Least deprived	1		1		1		1		1		1	
2	1.1	0.6-1.9	0.8	0.5-1.2	0.9	0.6-1.4	0.8	0.5-1.4	0.7	0.4-1.1	0.9	0.6-1.3
3	0.6	0.3-1.3	0.8	0.5-1.2	0.8	0.5-1.2	0.6	0.3-1.1	0.8	0.5-1.2	0.8	0.6-1.2
4	0.8	0.4-1.5	0.7	0.4-1.2	0.7	0.4-1.0	0.8	0.4-1.4	0.8	0.5-1.3	**0.6**	**0.4-0.9**
5 Most deprived	1.1	0.6-2.0	1.2	0.8-1.8	0.9	0.6-1.4	0.8	0.5-1.5	1.0	0.7-1.5	0.9	0.7-1.3

* Numbers in bold indicate significance

Chapter 9.

The validity of quality indicators for oesophago-gastric cancer surgery

Claudia Fischer, Hester F Lingsma, Niek S Klazinga, Richard Hardwick, David Cromwell, Ewout W Steyerberg, Oliver Groene

This article is submitted.
Volume-outcome revisited: the effect of hospital and surgeon volumes on multiple outcome measures in oesophago-gastric cancer surgery.

ABSTRACT

Background: Most studies that found a volume outcome effect in resection surgery for oesophago-gastric cancer were conducted before the centralisation of clinical services and reflect a higher baseline risk of surgery that exists today. This study evaluated the relation between hospital- and surgeon volume and different risk-adjusted outcomes after curative oesophago-gastric (OG) cancer surgery in England between 2011 and 2013.

Methods: Multivariable random-effects logistic regression models were used to quantify the effect of surgeon and hospital volume on three outcomes: 30-day and 90-day mortality and anastomotic leakage. The between-cluster heterogeneity was estimated with the median odds ratio (MOR).

Results: The study included patients treated at 42 hospitals and 329 surgeons. Higher hospital volume was associated with lower 30-day mortality (OR: 0.94; 95% CI: 0.91-0.98) and lower anastomotic leakage rates (OR: 0.96; 95% CI: 0.93-0.98) but not 90-day mortality. Higher surgeon volume was only associated with lower anastomotic leakage rates. Hospital volume explained a part of the between-hospital variation in 30-day mortality whereas surgeon volume explained part of the between-hospital variation in anastomotic leakage.

Conclusions: In the setting of centralized O-G cancer surgery in England, we could still observe an effect of volume on short-term outcomes. However, the effect is inconsistent, depending on the type of outcome measure under consideration, and much smaller than in previous studies. Efforts to centralise O-G cancer services further should carefully address the effects of both hospital and surgeon level on outcomes, and their effects on a range of outcome measures that are relevant to patients.

Keywords: oesophago-gastric cancer surgery, hospital volume, surgeon volume, postoperative mortality, complications

INTRODUCTION

For many surgical procedures, patient outcomes have been found to be related to surgical volume (the number of procedures that is performed in a specific unit), with studies typically showing that higher volumes are associated with lower postoperative mortality [1]. As a result, the centralization of high-risk oncological services, including oesophago-gastric cancer (O-G cancer), is occurring in many countries [2-6]. In the United States, the Leapfrog Group, a consortium of leading employers and private health care experts has set a minimum volume threshold of 13 O-G cancer resections per hospital per year [7, 8]. In the Netherlands, a minimum of 20 O-G cancer resections per year has been recently introduced [9]. In the UK, the Department of Health published a recommendation to centralize curative surgical services into specialised cancer centres in 2001 [10] and it is recommended that surgeons perform a minimum of 15 – 20 annual resections [11]. As a consequence, a process of reorganization has taken place in the National Health Services (NHS) during the past decade, which has resulted in a smaller number of acute trusts (hospital organisations) doing this type of surgery.

In O-G cancer surgery, risk-adjusted postoperative mortality and complication rates are widely used as quality indicators [12]. Case volume has also been proposed as a marker for quality in the past because of the substantial evidence of a volume-outcome relationship [13]. However, past studies were based on much higher baseline mortality rates than those observed in current practices. For example, a study from the UK based on patient data from 1990 to 2002 reported an average 30-day postoperative mortality rate of 11.4% [14]. Another study, using more recent data, reported a 30-day mortality rate of 4% [15]. It is unclear whether under these circumstances a volume-outcome relationship is still detectable given that, in a centralized setting, all trusts may exceed recommended thresholds. In addition, the exact mechanism behind the volume-outcome relation is still not fully understood [16, 17]. It is suggested that both the experience of the surgeon and the complete hospital team contribute to surgical outcomes [13]. An exploration of this mechanism is of importance because there is an increasing trend in reporting not only hospital-level, but also surgeon-level volume and outcomes. Finally, there are more outcomes of interest for O-G cancer surgery than the commonly used 30-day mortality. Recent publications suggest that anastomotic leakage rates and 90-day postoperative mortality are also important in assessing the quality of surgical care [18-22].

This study was undertaken to examine the relation between hospital- and surgeon volume and different risk-adjusted outcomes after O-G cancer surgery in a setting of centralized care, and the between-hospital and surgeon differences in outcome.

METHODS

Data

We used data submitted to the National Oesophago-Gastric Cancer Audit (NOGCA), which evaluates the care delivered by all (n=154) English hospitals that provide care to adults diagnosed with invasive, epithelial cancer of the oesophagus or stomach. Data are collected prospectively by hospital staff and have been submitted to the audit since 1 April 2011. Details on the audit method and dataset have previously been published [23]. All patients undergoing curative surgery between 1 April 2011 and 31 March 2013 were included in the study. We excluded patient undergoing curative oncological treatment for squamous cell carcinoma and all palliative patients. Further, we excluded hospitals which operated on less than 10 patients (Appendix 1).

Predictors and Outcomes

We considered three outcomes: 30-day mortality, 90-day mortality and anastomotic leakage. Mortality was defined as all-cause postoperative mortality within 30 or 90 days after surgery. Date of death was obtained from the Office for National Statistics death certificate register. Anastomotic leakage was defined as severe disruption to the anastomosis, irrespective of whether detected clinically or radiologically, and irrespective of whether it is managed conservatively or by re-operation [24].

Pre-operative patient and tumour characteristics to be used for case-mix adjustment were based on prior research. All regression models included: comorbidity count, age, ASA score, ECOG (WHO) performance status, T stage, N stage, cancer location. Patient gender was also included in the models for the outcomes 90-day mortality and AL, and deprivation was included in the AL model. Distinct domains of deprivation, such as income deprivation, employment deprivation, health deprivation and disability are weighted and combined into a single overall deprivation score. This score is used to rank areas in England according to their inhabitants score on the combined deprivation score [12, 25]. We analysed hospital and surgeon volume. Hospital was defined as NHS trust, which is a division within the English NHS that can consist of up of up to five hospitals. Usually O-G cancer surgery is only performed in one hospital in the trust. Surgeon was defined as the principal operating surgeon.

Hospital volume was defined as the number of O-G cancer surgeries performed at a NHS trust per year. Surgeon volume was defined as the annual number of operations conducted by an individual surgeon.

Statistical analysis

Patient characteristics were described as means or percentages. We described surgeon and hospital volume in the study period with median volume per hospital/surgeon. To describe outcome differences, we divided the hospitals/surgeons in quartiles based on volume and presented the outcome rates for each volume quartile.

The effect of hospital and surgeon volume on the three patient outcomes was tested in multivariable case-mix adjusted logistic regression models, with volume added as a continuous variable. We tested whether the relationship between volume and the outcomes was non-linear by adding squared terms and comparing these with linear terms based on the chi-square statistic.

We assessed differences in mortality and anastomotic leakage rates between hospitals with random effects models, which are considered appropriate for analyses of outcome differences between centres because they can account for correlation between patients within hospitals [26]. First, we analysed between hospitals and surgeons differences without any adjustment. This random effects model includes two random intercepts: one for hospital and one for surgeon and no other covariates. The variance of the random intercepts represents the between hospital/surgeon variation without any adjustment, but taking into account the random variation. In the second model the case-mix adjustment variables were added as covariates, to adjust the between hospital/surgeon variation for differences in patient characteristics. In a next step, we added surgeon and hospital volume one by one as covariates. In a final step, we included both surgeon and hospital volume.

In random effects models the between hospital/surgeon variation is reflected in the variance of the random intercepts ($\tau 2$). We used the median odds ratio (MOR) to quantify this variation. The MOR is a direct function of $\tau 2$ (MOR = exp ($\sqrt{}$ 2 x $\tau 2$ x Φ–1 (0.75)) [27]. The MOR can be equal or greater than 1, an MOR of 1 reflects no variation between the hospitals. The larger the between-hospital variation, the higher the MOR will be [27].

There were no missing values in patient outcomes. Missing data in predictors were imputed using a 'multiple imputation by chained equations' model including the outcome measures and the independent variables deprivation, age at diagnose, ECOG performance status, ASA score, gender, tumour location, number of comorbidities, size and/or extent of the primary tumour (T), regional lymph nodes (N). Imputation and statistical analysis were performed Stata software, version 11(StataCorp.2009.Stata Statistical Software: Release: 11) and in R statistical software 2.7.2 using the Hmisc, the lm4 and rms packages (R Foundation for Statistical Computation, Vienna) Syntax code for R is provided in Appendix 2.

RESULTS

Descriptives

The study included 4868 patients treated at 42 hospitals and 329 surgeons. Patients had a mean age of 66 years and the majority were male (n=3610; 74%). Fifty six percent of patients had no recorded co-morbidities (n=2735). Most patients scored an ECOG performance score of 0 (carries out all normal activity) (n=2515; 52%) or 1 (restricted but walks/does light work) (n=1550; 32%). The majority of the patients had a T stage of 3 (Tumour invades adventitia) (n=2318; 48%) and an ASA score of 2 (mild systemic disease) (n=2498; 51%). The majority of patients had an adenocarcinoma of the lower third of the oesophagus and a Siewert type 1 tumour (n=1904; 39%). Overall, 30-day mortality was 2.3% (n=111), 90-day mortality was 4.4% (n=215), and 6.3% (n=305) of the patients developed an anastomotic leakage (AL).

Variation on surgeon and hospital level

Volume varied between both surgeons and hospitals. The median hospital volume was 55 patients per year, with an interquartile range of 41 and 68 (Figure 1a). Many surgeons did not perform the recommended minimum of 15-20 annual resections. The median surgeon operated 6 patients per year, and the interquartile range was 4 and 9 (Figure 1b).

Table 1. Characteristics of the patients included in the study

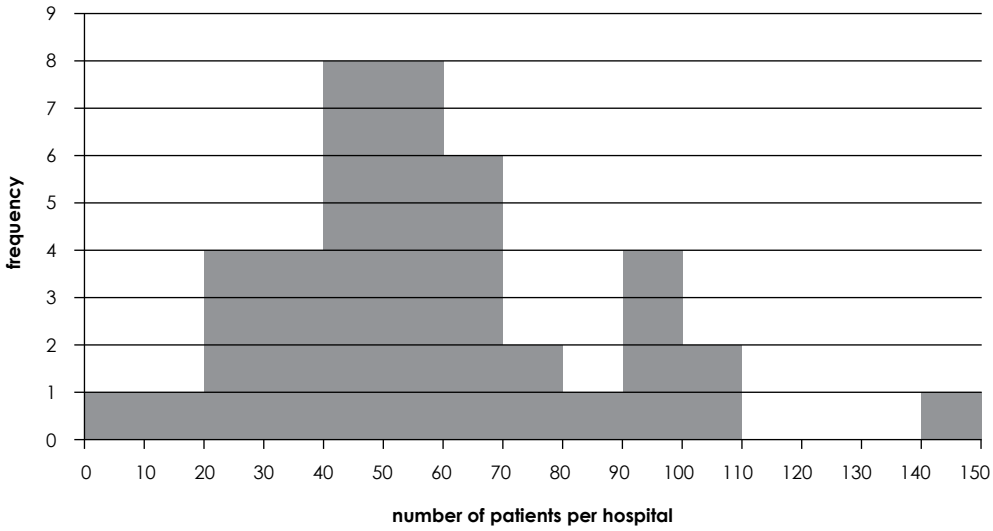
Patient and prognostic information	No. of patients	%
Age, years	4859	66*
Missing values	9	0.2
Comorbidity count		
No comorbidities	2735	56.2
One comorbidity	1309	26.9
Two comorbidities	566	11.6
Three or more comorbidities	258	5.3
Gender		
Male	3610	74.2

ECOG (WHO) performance status		
Carries out all normal activity	2515	51.7
Restricted but walks/does light work	1550	31.8
Walks, full self-care but no work	526	10.8
Limited self-care - fully disabled	120	2.5
Missing values	157	3.2
Size and/or extent of the primary tumour (T)		
No evidence of primary tumour T(0)	202	4.1
Tumour invades lamina propria or submucosa T(1)	927	19.0
Tumour invades muscularis propria T(2)	790	16.2
Tumour invades adventitia T(3)	2318	47.6
Tumour invades adjacent structures T(4)	489	10.0
Missing values	142	2.9
Regional lymph nodes (N)		
No regional lymph node metastasis N(0)	2137	43.9
Metastasis in 1 to 2 regional lymph nodes N(1)	1496	30.7
Metastasis in 3 to 6 N(2)	614	12.6
Metastasis in 7 or more N(3)	507	10.4
Missing values	114	2.3
ASA Scale		
Normal healthy patient	811	16.7
Mild systemic disease	2498	51.3
Severe systemic disease	1246	25.6
Life-threatening disease/ Moribund patient	60	1.2
Missing values	253	5.2
Cancer location		
Squamous cell carcinomas of the oesophagus	491	10.1
Adenocarcinomas of the upper and middle oesophagus	183	3.8
Adenocarcinomas of the lower third of the oesophagus and Siewert type 1 tumours	1904	39.1
Siewert type II and type III tumours	840	17.3
Tumours of the stomach	1450	29.8
Deprivation		
1 Least deprived	838	17.2
2	859	17.6
3	844	17.3
4	799	16.4
5 Most deprived	745	15.3
Missing values	783	16.1

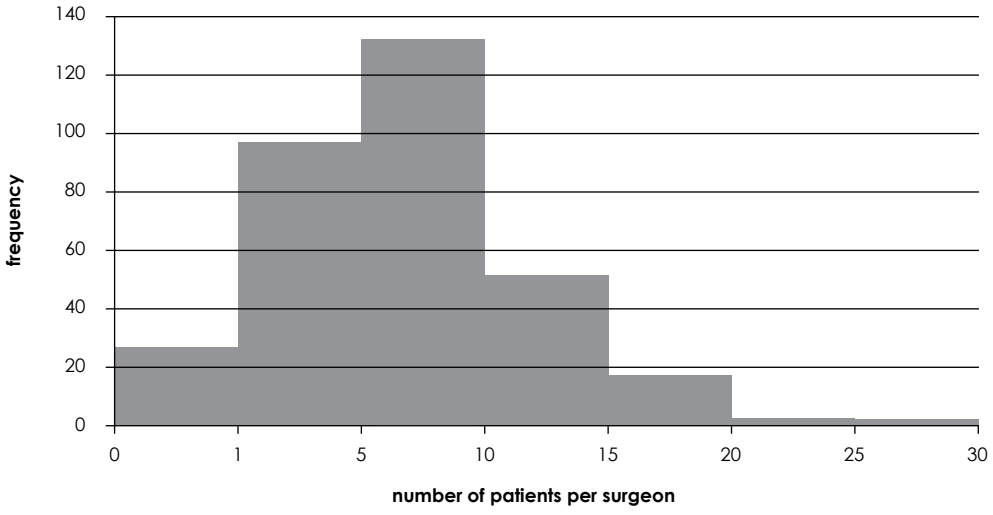
*Mean

Figure 1ab. Hospital volume (1a), surgeon volume (1b)

Median hospital volume = 55
Interquartile range = 41 - 68



Median surgeon volume = 6
Interquartile range = 4 - 9



The descriptive analysis of outcomes by quartiles of hospital volume and surgeon volume showed an overall lower risk for all outcomes in the highest volume quartiles compared to the lowest volume quartiles (Table 2). For example, 30-day mortality was 3.0% in the lowest hospital volume quartile compared 1.3% in the highest quartile. The anastomotic leakage risk was 7.9% in the lowest average annual surgeon volume quartile compared to the highest quartile, in which the risk for anastomotic leakage was 1.3%.

Table 2. 30-day, 90-day mortality and anastomotic leakage risk according to quartiles of hospital volume and surgeon volume

		Average hospital volume				**Average surgeon volume**			
	Total	Q 1 (0-49)	Q 2 (50-65)	Q 3 (66-91)	Q 4 (92-148)	Q 1 (0-5)	Q 2 (6-9)	Q 3 (10-13)	Q 4 (14-28)
N		1,253	1,148	1,360	1,107	1,144	1,156	1,292	1,169
30-day mortality, %	2.3	3.0	3.1	1.7	1.3	2.4	2.3	2.6	0.7
90-day mortality, %	4.4	5.0	5.0	3.8	3.9	4.5	5.4	4.0	1.4
Anastomotic leakage,%	6.3	7.1	8.9	6.3	2.5	7.9	7.1	4.6	1.3

Table 3. Association between adjusted volume predictors and the outcomes 30-day mortality, 90-day mortality and anastomotic leakage

	30-day mortality		**90-day mortality**		**Anastomotic leakage**	
	OR	**95%CI**	**OR**	**95%CI**	**OR**	**95%CI**
Surgeon volume*						
Adjusted for case-mix	1.03	0.84-1.25	1.01	0.87- 1.16	**0.88**	**0.78-1.00**
Adjusted for case-mix and hospital volume	0.92	0.76-1.12	0.97	0.85-1.11	**0.81**	**0.72-0.92**
Hospital volume*						
Adjusted for case-mix	**0.94**	**0.91-0.98**	0.98	0.96-1.00	**0.95**	**0.93-0.97**
Adjusted for case-mix and surgeon volume	**0.94**	**0.91-0.98**	0.98	0.96-1.01	**0.96**	**0.93-0.98**

* ORs represent the effect of 5 extra patients per year

Higher hospital volume was a significant predictor for lower 30-day mortality, after adjustment for case-mix and surgeon volume (OR: 0.94 (95%CI: 0.91-0.98) per increase of 5 patients per year) (table 3). Surgeon volume had no significant effect on 30-day mortality. Higher surgeon and higher hospital volume were independent predictors of lower risk of an anastomotic leakage (hospital volume OR: 0.96 (95%CI: 0.93-0.98), surgeon volume: OR: 0.88 (95% CI: 0.78-1.00). Neither hospital volume nor surgeon volume were significant predictors for 90-day mortality when controlling for each other (Table 3).

Hospital volume explained part of the variation in 30-day mortality between the hospitals. The median odds ratio (MOR) decreased from 1.38 without controlling for hospital volume in the model to 1.30 when hospital and surgeon volume were added to the models as covariates (Figure 2). The MOR represents the odds of dead within 30 days of a patient from a randomly selected hospital compared to the odds if he/she would go to another randomly selected hospital. In this model the MOR represents outcome differences between centres, after taking into account the random variation, differences in patient characteristics and that patients were treated by surgeons and hospitals with varying procedure volumes. Differences between hospitals in this model thus may be due to quality of care related factors, other than volume. Surgeon volume did not explain between-hospital or between-surgeon variation in 30-day mortality (Figure 2, Figure 3).

Hospital volume explained more of the between-surgeon variation in anastomotic leakages (change in MOR from 1.67 to 1.56) (Figure 2). The other way around, surgeon volume explained a very small part of the variation between hospitals (change in MOR from 1.29 to MOR 1.27) (Figure 3) (Appendix 3).

While hospital volume explained a minimal amount of between-hospitals variation in 90-day mortality, surgeon volume did not explain any variation in 90-day mortality at all (Figure 2, Figure 3).

Figure 2. Between-hospital variation in outcomes

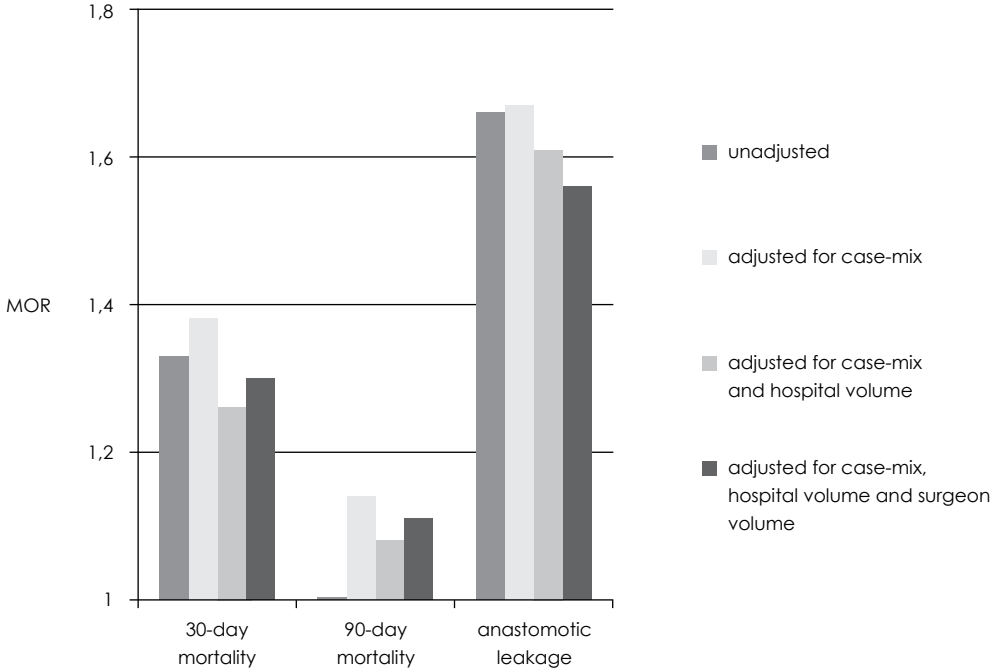


Figure legend: Change in median odds ratio (MOR) on hospital level considering different case mix and prediction factors for 30-day, 90-day mortality and anastomotic leakage

Figure 3. Between-surgeon variation in outcomes

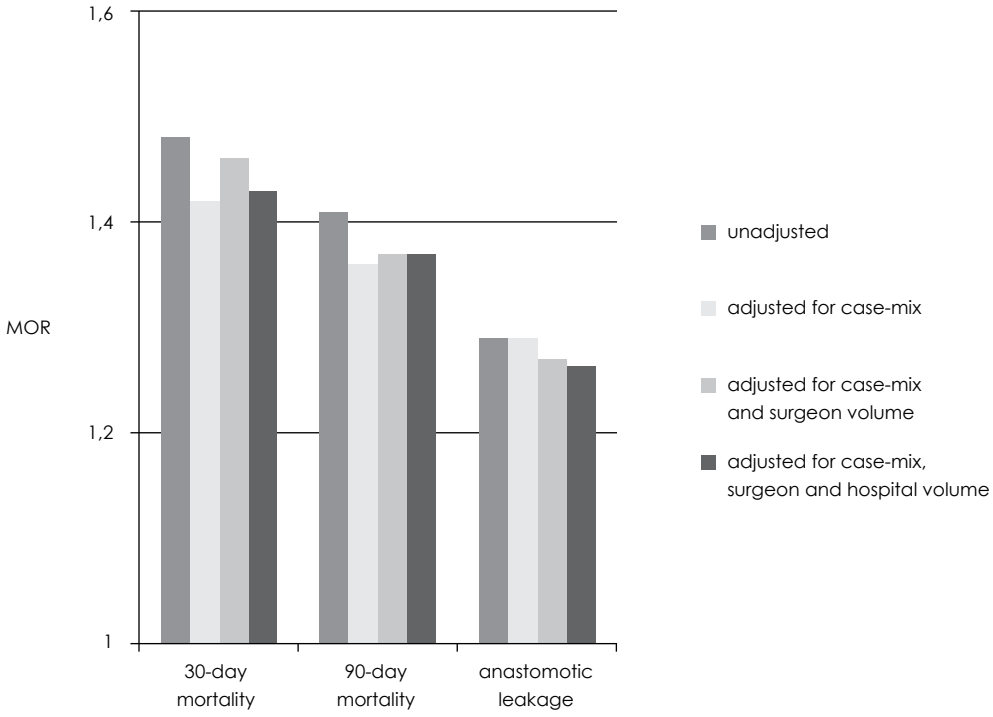


Figure legend: Change in median odds ratio (MOR) on surgeon level considering different case mix and prediction factors for 30-day, 90-day mortality and anastomotic leakage

DISCUSSION

Main findings

This study has shown that despite centralization, differences between hospitals and surgeons in patient outcomes after O-G surgery still exist in England. These between-hospital and between-surgeon differences are partly explained by surgeon and hospital volume but the volume outcome relation is different for the different outcomes (30-day and 90-day mortality and anastomotic leakage). Higher hospital volume was associated with lower 30-day mortality, but high surgeon volume was not. Neither hospital volume nor surgeon volume affected 90-day mortality. Higher surgeon and hospital volume were both associated with fewer anastomotic leakages, but surgeon volume was the stronger predictor.

Volume-outcome relation

Higher volume hospitals have lower 30-day mortality rates; this in line with earlier research in both O-G cancer and other surgical oncological procedures [3, 28]. However, no clear definition of 'high volume' exists yet [29]. A recent, study on patients undergoing oesophagectomy in the Netherlands showed that increasing annual hospital volume was associated with a nonlinear decrease in mortality up to 40-60 oesophagectomies, after that a plateau was reached [30]. This finding was not replicated in our data: we found a linear relationship while the majority of hospitals treated more than 50 patients. Despite the centralization that took place in the UK, we still found an effect of volume.

In contrast, we did not find surgeon volume to be a significant predictor of 30-day mortality. Surgeon volume and hospital volume possibly reflect different aspects of quality of care. It has been suggested that next to the surgeon skills there are other hospital factors in high volume hospitals that reduce mortality risk, such as post-operative care, care pathways and multidisciplinary team work [31-34].

Surgical volume and hospital volume had only a minor effect on 90-day mortality. Although still a substantial amount of patients dies between 30- and 90-days after surgery, their death seems not to be influenced by the surgeon or hospital volume. Previous research showed that although part of the deaths occurring after 30 days is still related to surgery, and increasing proportion related to cancer recurrence [19]. Both, higher surgeon and hospital volume showed to be related to fewer anastomotic leakages, but surgeon volume had a stronger effect. It can be imagined that anastomotic

leakages are closely related to technical surgical skills that high volume surgeons have better developed.

When calculating surgeon volume for highly complicated but rare procedures, it has been suggested to count also related procedures [35]. In our study we had no data to take into account other operations as O-G resections. We could however capture surgeries that were performed by one surgeon in different centres. This is especially important when some surgeons operate in multiple hospitals and others do not.

Recommendations

In different countries different volume norms are used, which is the reflection of mixed scientific evidence. We found that even in the centralized UK setting there is a relation between volume and 30-day mortality and anastomotic leakages. This finding suggests that further increasing hospital volumes, and to a lesser extent surgeon volumes, might improve short-term outcomes. However, on-going centralization might also have negative effects on for example access and equity, which we did not study.

We studied different outcomes that are used as quality indicators. Anastomotic leakages (AL) rates were affected by hospital volume. This is in line with previous findings that surgical complications seem to be a good indicator for surgical quality as they are closely related to the surgical process and are not so much influenced by patient characteristics [12, 36, 37]. This makes AL rates 'actionable' hospital quality indicators. Apart from the clinical relevance, the relatively frequent occurrence makes them attractive as a quality indicator from a statistical point of view, as rates per hospital can be estimated relatively certain. A large disadvantage of complications as an outcome indicator is their possibly unclear definition, especially compared to mortality. Different definitions of complications, or a different interpretation of the same definition, may lead to biased hospital comparisons [38].

Adjusted 30-day mortality differences between hospitals were partly explained by hospital volume. However, a large disadvantage of this measure is the low event rate. Even in our data pooled from two years, the absolute numbers of deaths within 30-days per hospital were small (median absolute number per hospital = 2), which makes the estimates per hospital uncertain and challenges the comparison between hospitals [39, 40].

In that sense 90-day mortality rate is more attractive as more deaths occur. However, 90-day mortality differences between hospitals were not ex-

plained by hospital volume. Possibly the 90-day mortality rates might reflect aspects of quality of care not related to volume. But we consider it is more likely that a longer time period introduces effects of confounding factors that dilute the relation between quality of care and outcome. Previous research showed the 'ideal' timeframe to measure quality of O-G cancer surgery lies somewhere between 30 and 90 days [21]. In all, always multiple indicators should be monitored as this gives a more comprehensive picture and hospitals often do not score good or bad on all indicators [41].

Next to variation between hospitals, we did observe variation in outcome between surgeons. But these were not explained by any of the volume indicators. In addition, the absolute numbers of deaths of ALs per surgeon are extremely low, which makes outcome rates per surgeon unsuitable as quality indicators: the difference between 0 or 1 death per year cannot tell us anything about quality, only about bad luck.

In summary, in the setting of centralized O-G cancer surgery in England, we could still observe an effect of hospitals volume on 30-day mortality and AL rates, suggesting that further centralization might be considered based on our data. However, the effects are much weaker and less consistent than those observed in previous research. Surgeon volume only affected AL rates, so represents only part of the surgical process. AL rates and 30-day mortality rates per hospital could be useful as quality indicators, but both have their own large disadvantage: definition problems for AL and low frequency for mortality. 90-day mortality likely reflects other things than the quality of surgery and thus is less suitable as a quality indicator. Surgeon level outcomes are very infrequent and volume did not explain the differences.

Acknowledgments
We like to acknowledge the health professionals and support personnel in English NHS trusts and Cancer networks who submitted the data to the Audit.

Competing Interests
None declared.

Sources of funding
The National Oesophago-Gastric Cancer Audit is commissioned by the Healthcare Quality Improvement Partnership (HQIP) as part of the National Clinical Audit Programme (NCA). HQIP is led by a consortium of the Academy of Medical Royal Colleges, the Royal College of Nursing and National

Voices. Its aim is to promote quality improvement, and in particular to increase the impact that clinical audit has on healthcare quality in England and Wales. HQIP holds the contract to manage and develop the NCA Programme, comprising more than 30 clinical audits that cover care provided to people with a wide range of medical, surgical and mental health conditions. The programme is funded by NHS England, the Welsh Government and, with some individual audits, also funded by the Health Department of the Scottish Government, DHSSPS Northern Ireland and the Channel Islands.

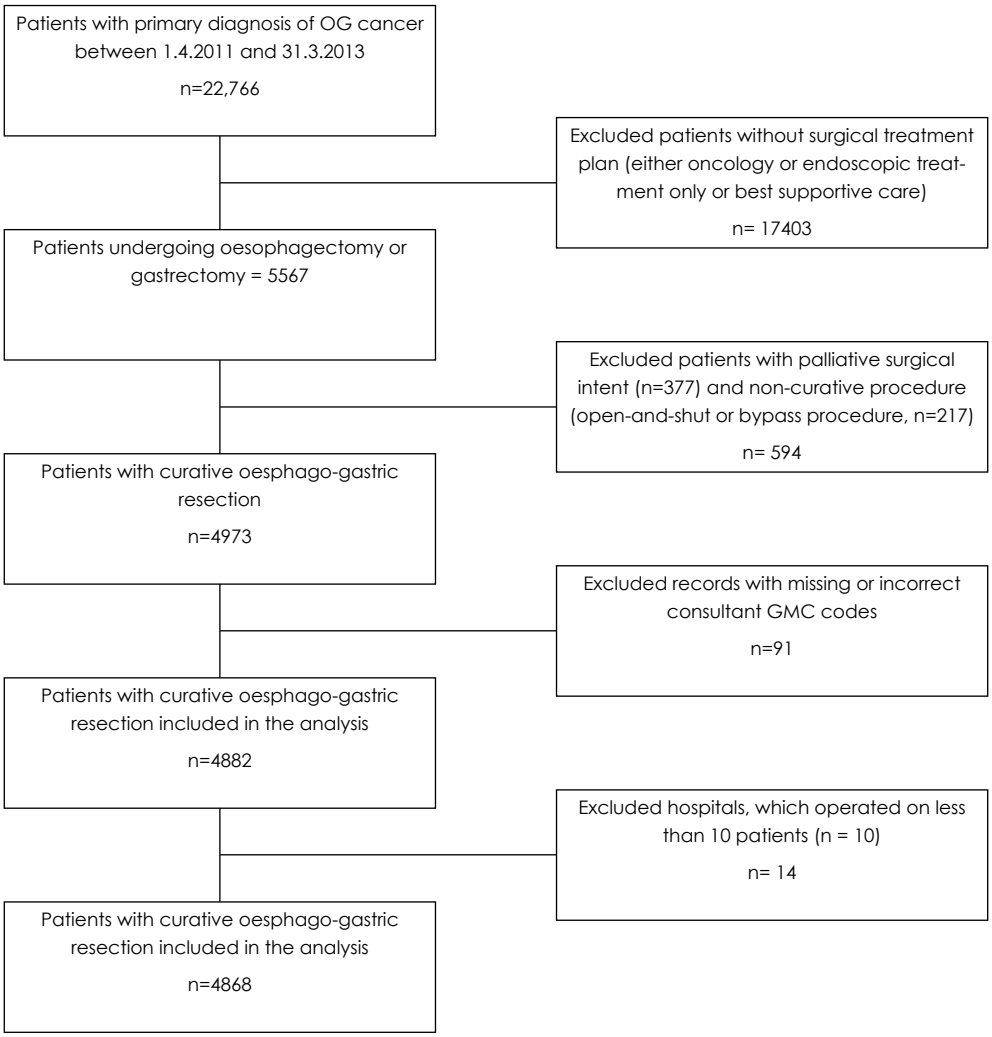
LITERATURE

1. Birkmeyer JD, Siewers AE, Finlayson EV, Stukel TA, Lucas FL, Batista I, et al. Hospital volume and surgical mortality in the United States. N Engl J Med. 2002 Apr 11;346(15):1128-37.
2. Gooiker GA, van der Geest LG, Wouters MW, Vonk M, Karsten TM, Tollenaar RA, et al. Quality improvement of pancreatic surgery by centralization in the western part of the Netherlands. Annals of surgical oncology. 2011 Jul;18(7):1821-9.
3. Begg CB, Cramer LD, Hoskins WJ, Brennan MF. Impact of hospital volume on operative mortality for major cancer surgery. JAMA. 1998 Nov 25;280(20):1747-51.
4. Finlayson EV, Goodney PP, Birkmeyer JD. Hospital volume and operative mortality in cancer surgery: a national study. Arch Surg. 2003 Jul;138(7):721-5; discussion 6.
5. Finks JF, Osborne NH, Birkmeyer JD. Trends in hospital volume and operative mortality for high-risk surgery. N Engl J Med. 2011 Jun 2;364(22):2128-37.
6. Epstein AM. Volume and outcome--it is time to move ahead. N Engl J Med. 2002 Apr 11;346(15):1161-4.
7. Varghese TK, Jr., Wood DE, Farjah F, Oelschlager BK, Symons RG, MacLeod KE, et al. Variation in esophagectomy outcomes in hospitals meeting Leapfrog volume outcome standards. The Annals of thoracic surgery. 2011 Apr;91(4):1003-9; discussion 9-10.
8. Milstein A, Galvin RS, Delbanco SF, Salber P, Buck CR, Jr. Improving the safety of health care: the leapfrog initiative. Eff Clin Pract. 2000 Nov-Dec;3(6):313-6.
9. Association of Surgeons of the Netherlands (NVvH). Normering Chrirurgische behandelingen 3.0 (2012). Available from: http://www.heelkunde.nl/uploads/4w/qz/4wqzdizoxd5GDUvTc1Ixgg/NVvH-Normen-30pdf. Accessed 30 March 2015.
10. Natioanl Oesophago-Gastric Cancer Audit. Natioanal Oesophago-Gastric Cancer Audit. Third Annual Report-Patient Summary (2012). Available from: http://www.opa.org.uk/news/national-oesophago-gastric-cancer-audit--third-annual-report-patient-summary.html 30 March 2015.
11. Association of Upper Gastrointestinal Surgeons of Great Britain and Ireland (AUGIS) (2010). Guidance on minimum surgeon volumes. Available from: http://www.augis.org/wp-content/uploads/2014/05/AUGIS_recommendations_on_Minimum_Volumes.pdf. Accessed 30 March 2015.
12. Fischer C LH, Hardwick R, et al. The development of a case-mix correction model for comparing outcomes of oesophago-gastric cancer surgery. 2015.
13. Brusselaers N, Mattsson F, Lagergren J. Hospital and surgeon volume in relation to long-term survival after oesophagectomy: systematic review and meta-analysis. Gut. 2014 Sep;63(9):1393-400.
14. Zafirellis KD, Fountoulakis A, Dolan K, Dexter SP, Martin IG, Sue-Ling HM. Evaluation of POSSUM in patients with oesophageal cancer undergoing resection. The British journal of surgery. 2002 Sep;89(9):1150-5.
15. Dutta S, Al-Mrabt NM, Fullarton GM, Horgan PG, McMillan DC. A comparison of POSSUM and GPS models in the prediction of post-operative outcome in patients undergoing oesophago-gastric cancer resection. Annals of surgical oncology. 2011 Oct;18(10):2808-17.
16. Dudley RA, Johansen KL. Invited commentary: Physician responses to purchaser quality initiatives for surgical procedures. Surgery. 2001 Sep;130(3):425-8.
17. Khuri SF. Invited commentary: Surgeons, not General Motors, should set standards for surgical care. Surgery. 2001 Sep;130(3):429-31.
18. Lingsma HF, Steyerberg EW, Eijkemans MJ, Dippel DW, Scholte Op Reimer WJ, Van Houwelingen HC, et al. Comparing and ranking hospitals based on outcome: results from The Netherlands Stroke Survey. QJM. 2010 Feb;103(2):99-108.
19. Talsma AK, Lingsma HF, Steyerberg EW, Wijnhoven BP, Van Lanschot JJ. The 30-day versus in-hospital and 90-day mortality after esophagectomy as indicators for quality of care. Annals of surgery. 2014 Aug;260(2):267-73.
20. McMillan RR, Berger A, Sima CS, Lou F, Dycoco J, Rusch V, et al. Thirty-day mortality underestimates the risk of early death after major resections for thoracic malignancies. The Annals of thoracic surgery. 2014 Nov;98(5):1769-74; discussion 74-5.
21. Rutegard M, Lagergren P, Johar A, Lagergren J. Time Shift in Early Postoperative Mortality After Oesophagectomy for Cancer. Annals of surgical oncology. 2015 Feb 4.

22. Snijders HS, Henneman D, van Leersum NL, ten Berge M, Fiocco M, Karsten TM, et al. Anastomotic leakage as an outcome measure for quality of colorectal cancer surgery. BMJ Qual Saf. 2013 Sep;22(9):759-67.
23. Hardwick G GO, Cromwell D. National Oesophago-Gastric Cancer Audit 2013. The Royal College of Surgeons of England, 2013.
24. Lee HS, Ha AW, Kim WK. Effect of resveratrol on the metastasis of 4T1 mouse breast cancer cells in vitro and in vivo. Nutrition research and practice. 2012 Aug;6(4):294-300.
25. McLennan D BH, Noble M, et al,. The english indices of deprivation 2010 2011. Available from: http://data.gov.uk/dataset/index-of-multiple-deprivation. Accessed 21.04.2015.
26. Lingsma H. Measuring quality of care. Methods and applications to acute neuroloigcal diseases [thesis]. Alphen a/d Rijn: ErasmusMC; 2010.
27. Larsen K, Merlo J. Appropriate assessment of neighborhood effects on individual health: integrating random and fixed effects in multilevel logistic regression. Am J Epidemiol. 2005 Jan 1;161(1):81-8.
28. Hannan EL, Radzyner M, Rubin D, Dougherty J, Brennan MF. The influence of hospital and surgeon volume on in-hospital mortality for colectomy, gastrectomy, and lung lobectomy in patients with cancer. Surgery. 2002 Jan;131(1):6-15.
29. Christian CK, Gustafson ML, Betensky RA, Daley J, Zinner MJ. The volume-outcome relationship: don't believe everything you see. World journal of surgery. 2005 Oct;29(10):1241-4.
30. Henneman D, Dikken JL, Putter H, Lemmens VE, Van der Geest LG, van Hillegersberg R, et al. Centralization of esophagectomy: how far should we go? Annals of surgical oncology. 2014 Dec;21(13):4068-74.
31. Aiken LH, Clarke SP, Cheung RB, Sloane DM, Silber JH. Educational levels of hospital nurses and surgical patient mortality. JAMA. 2003 Sep 24;290(12):1617-23.
32. Nguyen NT, Paya M, Stevens CM, Mavandadi S, Zainabadi K, Wilson SE. The relationship between hospital volume and outcome in bariatric surgery at academic medical centers. Annals of surgery. 2004 Oct;240(4):586-93; discussion 93-4.
33. Daley J, Forbes MG, Young GJ, Charns MP, Gibbs JO, Hur K, et al. Validating risk-adjusted surgical outcomes: site visit assessment of process and structure. National VA Surgical Risk Study. J Am Coll Surg. 1997 Oct;185(4):341-51.
34. Gort M, Broekhuis M, Otter R, Klazinga NS. Improvement of best practice in early breast cancer: actionable surgeon and hospital factors. Breast Cancer Res Treat. 2007 Apr;102(2):219-26.
35. Modrall JG, Rosero EB, Chung J, Arko FR, 3rd, Valentine RJ, Clagett GP, et al. Defining the type of surgeon volume that influences the outcomes for open abdominal aortic aneurysm repair. J Vasc Surg. 2011 Dec;54(6):1599-604.
36. Junemann-Ramirez M, Awan MY, Khan ZM, Rahamim JS. Anastomotic leakage post-esophagogastrectomy for esophageal carcinoma: retrospective analysis of predictive factors, management and influence on longterm survival in a high volume centre. Eur J Cardiothorac Surg. 2005 Jan;27(1):3-7.
37. Fischer C, Lingsma HF, van Leersum N, Tollenaar RA, Wouters MW, Steyerberg EW. Comparing colon cancer outcomes: The impact of low hospital case volume and case-mix adjustment. Eur J Surg Oncol. 2015 Apr 30.
38. Low DE, Alderson D, Cecconello I, Chang AC, Darling GE, D'Journo XB, Griffin SM, Hölscher AH, Hofstetter WL, Jobe BA, Kitagawa Y, Kucharczuk JC, Law SY, Lerut TE, Maynard N, Pera M, Peters JH, Pramesh CS, Reynolds JV, Smithers BM, van Lanschot JJ. International Consensus on Standardization of Data Collection for Complications Associated With Esophagectomy: Esophagectomy Complications Consensus Group (ECCG). Ann Surg 2015 Aug; 262(2):286-94 2015.
39. Daley J, Henderson WG, Khuri SF. Risk-adjusted surgical outcomes. Annu Rev Med. 2001;52:275-87.
40. 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.
41. Almoudaris AM, Burns EM, Bottle A, Aylin P, Darzi A, Vincent C, et al. Single measures of performance do not reflect overall institutional quality in colorectal cancer surgery. Gut. 2013 Mar;62(3):423-9.

APPENDIX

Appendix 1. Flow chart patient inclusion process inclusion process



Appendix 2.

R code

#Load required packages

library(lm4) #fits random effect logistic regression models

library(foreign) #can import foreign data files

#Import STATA datafile

MyData <- read.dta("S:/MyDocuments/……..dta", convert.dates=TRUE, convert.factors=TRUE, missing.type=TRUE, convert.underscore=TRUE)

#Random effects model with 30-day mortality as outcome, patient characteristics, surgeon and #hospital volume as covariates, and two random effects: one for surgeon and one for hospital #(Figure 1, Figure 2, outcome 30-day mortality bar group, last bar)

MyModel <- glmer(30day.mortality ~ volume.surgeon + volume.hospital + age + as.factor(comorbidity) + as.factor(performancestatus) + as.factor(t)+ as.factor(n) + as.factor(asa)+ as.factor(tumour.location) + (1| surgeon) + (1|hospital), data=MyData, family=binomial)

#MOR between surgeons

exp(0.674*(sqrt(as.numeric(VarCorr(MyModel)[[1]]))))

#MOR between hospitals

exp(0.674*(sqrt(as.numeric(VarCorr(MyModel)[[2]]))))

Appendix 3. Table 1. Between surgeon and between-hospital variation in 30-day, 90-day mortality and anastomotic leakage

	Unadjusted	Adjusted for patient characteristics	Adjusted for patient characteristics and hospital volume	Adjusted for patient characteristics and surgeon volume	Adjusted for patient characteristics, surgeon and hospital volume
	MOR*	**MOR**	**MOR**	**MOR**	**MOR**
30-day mortality					
Combined between surgeons	1.48 (1-2.93)	1.42 (1-2.87)		1.46 (1-2.91)	1.43 (1-2.88)
Combined between hospitals	1.33 (1-1.81)	1.38 (1 – 1.86)	1.26 (1 – 1.74)		1.30 (1 – 1.78)
90-day mortality					
Combined between surgeons	1.41 (1-1.81)	1.36 (1 – 1.76)		1.37 (1 – 1.77)	1.37 (1 – 1.77)
Combined between hospitals	1 (1-1.55)	1.14 (1 – 1.69)	1.08 (1 – 1.63)		1.11 (1 – 1.66)
Anastomotic leakage					
Combined between surgeons	1.29 (1-1.71)	1.29 (1 – 1.71)		1.27 (1 – 1.69)	1.27 (1 – 1.69)
Combined between hospitals	1.66 (1.63-2.01)	1.67 (1.64 – 2.02)	1.61 (1.58 – 2.02)		

*Median odds ratio (95% Confidence Interval)

Part IV

Discussion and summary

Chapter 10.

General discussion

GENERAL DISCUSSION

Hospital quality indicators are widely implemented for purposes such as increasing accountability and transparency as well as achieving the overarching goal of quality improvement. However, it is not clear whether currently used hospital quality indicators actually reflect quality of care.[1] The aim of this thesis was to expand our knowledge on how to measure quality of hospital care, particularly with regard to its use for external comparison. We specifically investigated reliability and validity, as these are key aspects of quality indicators.[2] For reliability we focused on data quality, indicator definitions and statistical uncertainty. For validity we focused on case-mix correction.

The specific research questions were:

1. *Which aspects of data quality affect the reliability of quality indicators for hospital care?*
2. *How clear are the definitions of currently used quality indicators for hospital care?*
3. *To what extent does statistical uncertainty affect the reliability of outcome indicators for surgical colon and oesophago-gastric cancer care?*
4. *To what extent does case-mix correction affect the validity of outcome indicators for surgical colon and oesophago-gastric cancer care?*

For research question 1 and 2 we studied the data underlying Dutch national indicators (DHTP (Dutch Health Transparency Program), in the time period 2008/09) and performed a literature study on readmission rates. We found that administrative data is usually suboptimal for calculating quality indicators for external purposes, such as pay for performance schemes. Data quality problems include heterogeneous data information systems, inaccurately registered data (e.g. estimated instead of calculated indicator scores), and incomplete data (e.g. due to the inability to follow patients if they went to other hospitals, low number of secondary diagnoses).

Our literature studies showed that indicator definitions are often ambiguous. Clear indicator definition is a crucial aspect of reliability as unspecific indicator definitions lead to different interpretations of the indicator, which may result in a biased estimation of the indicator score.

To answer research question 3 and 4 we analysed Dutch colon cancer audit data (Dica, years 2011/12) and English oesophago-gastric (O-G) cancer audit data (years 2011/12).

We found that statistical uncertainty leads to uncertain estimates of the outcome rate per hospital. Taking into account statistical uncertainty with random effect models showed that differences between hospitals in outcome after colon cancer surgery were to a large extent attributable to chance. Especially when event rates are low, such as in the case of postoperative mortality, and when comparing outcomes per surgeon, it is crucial to take into account uncertainty.

We found that case-mix explains a certain extent of outcome variation between hospitals in 30-day mortality and to a lesser extent in anastomotic leakages in both colon and oesophago-gastric cancer and therefore case-mix is important to take into account.
We developed a case-mix correction model using the English O-G cancer audit data (years 2011/2012). This model can be used for correcting for case-mix when comparing short-term outcomes following O-G surgery.

This final chapter is divided into two parts. The first part discusses our main findings per research question. In the second part we interpret our findings and present implications for science and policy.

Data quality
Our first research question addressed the impact of data quality on the reliability of hospital quality indicators. High-quality data, namely data that are as complete and as error-free as possible, form an essential requirement for reliable hospital quality indicators.

In Chapter 3 we presented an overview of factors threatening the reliability of the outcome indicator readmission rate. Most quality indicators are calculated using administrative data. Although this type of data has several advantages, such as being easy to obtain and including large patient populations, we found that the low quality of administrative data hampers the calculation of reliable quality indicators due to the data's inaccuracy and incompleteness (Chapter 3, Chapter 5, Chapter 6). While basic case-mix information, such as age, is mostly available in administrative databases, clinical information, such as disease severity, is often lacking.

Our literature study showed that in currently used administrative databases patients often cannot be followed. This is problematic because outcomes of interest for quality of care measurement, such as death or readmission, occur after hospital discharge. Therefore, the ability to follow patients beyond hospital discharge is essential.[3] Especially with increasingly centralized

care, patients may be readmitted in centres other than the one where the primary admission took place. Missing these readmissions creates an underestimation of the outcome in the index hospitals. Unique patient identifiers, which allow following patients across hospitals, are essential (Chapter 3).

We found that coding plays an important role in determining the level of data quality. Variations in coding practices can have a significant effect on the validity of the registered data elements. This variation is usually caused by the fact that different hospitals put different types of staff in charge of the coding process (e.g. medical vs. administrative staff) and by the fact that the indicators are broadly defined, allowing subjective judgements to be made.

Another factor threatening the reliability of indicators is the burden of the data collection. For the DHTP indicators it was shown that data collection and computation strategies vary widely between Dutch hospitals. For example, when data elements are relatively inaccessible and require a large amount of labour to extract the necessary information, some hospitals delivered an estimation of the indicator score. As a result, many self-reported quality indicators show implausible indicator scores. Indicators that are based on data elements collected in a standardized manner by the national cancer registry were much more plausible (Chapter 5).

Indicator definitions

Our research question 2 addressed the impact of indicator definitions on reliability. The definition of an indicator determines which patient group is included (denominator) and which events are counted (numerator). Both, the denominator and the numerator can be affected by vague definitions. In our literature studies we showed that the numerator of the readmission rate is seldom precisely defined (Chapter 2, Chapter 4). In the majority of the studies we looked at, the definitions in the studies only included the timeframe in which the outcome was investigated. Usually events between 28 and 31 days were captured.

Summary of data issues

In our reviews we found that poor data quality and unclear indicator definitions are common problems encountered when using administrative data for the purpose of measuring quality of care. Data collection and computation strategies vary widely in practice. Our empirical study showed that the data elements that are difficult to abstract show low reliability because sometimes they are just estimated instead of calculated. We found that

imprecise indicator definitions are generally seen in the literature, the basic distinction between planned and unplanned events is seldom made. Current administrative databases often do not allow this distinction to be made. The usual time frame for investigating events is around 30-days, irrespective of the underlying disease. We showed that low data quality and imprecise indicator definitions lead to an over- or underestimation of an indicator score. Ignoring this aspect of low reliability leads to biased comparisons between hospitals.

Statistical uncertainty

Our third research question addressed the impact of statistical uncertainty on the reliability of hospital quality indicators.

Our empirical studies have shown that the number of postoperative deaths and anastomotic leakages are low in colon cancer and oesophago-gastric cancer patients. For example, in the Dutch colon cancer cohort the median postoperative mortality rate per hospital was 3% (absolute number: 5 patients) (Chapter 7). As was shown in our analysis, when numbers are that low, the impact of statistical uncertainty is significant, which results in uncertain estimates of the outcome rate per hospital. The effect of statistical uncertainty was smaller in our second outcome, anastomotic leakage following colon cancer surgery, which occurred more frequently than postoperative deaths (anastomotic leakage in median hospital: 7%, 9 patients). This is explained by the fact that not only the total number of treated patients per hospital, but the frequency of the outcome, determines the size of the impact of statistical uncertainty.[4]

Advanced statistical models, specifically random effect models, can be effectively used to take statistical uncertainty into account. These models assume that the hospitals are all part of a larger population. As a result, extreme but uncertain estimates in individual hospitals are shrunken towards the overall mean. When event rates are high the models based on the fixed and random effect will reveal similar results.[5]

In Chapter 7 it was shown that it is crucial which method is chosen when comparing outcomes between hospitals where outcome rates are usually low. For postoperative mortality following colon cancer surgery, we found that taking the random variation into account diluted all observed variation in postoperative mortality between hospitals. There was no variation in postoperative mortality beyond random variation.

The effect of statistical uncertainty will increase even further when the unit of comparison becomes smaller, for instance, when comparing outcomes between surgeons (Chapter 9). In those cases relevant outcome differences are unlikely to be detected and therefore they are not recommended for external reporting.

Besides the indicator scores for individual hospitals, there may also be interest in the variation in outcome between hospitals. Quantifying such overall outcome differences can be done in several ways, for example, with the interquartile range or standard deviation of the outcome rates. Measures to quantify outcome differences from random effect models (taking into account statistical uncertainty) are not easily interpretable. Several measures have been proposed based on the tau^2, which is the variance of the random effects.[6] In Chapter 9 we used the median odds ratio (MOR) to quantify the outcome variation between hospitals. The MOR is a direct function of tau^2 and quantifies the variation between hospitals (or surgeons) by comparing two persons with the same covariates, chosen randomly from different hospitals. The MOR can be equal to or greater than 1; an MOR of 1 reflects no variation between the hospitals. The larger the between-hospital variation, the higher the MOR will be.[7] For example, an MOR of 2 means that if two 'similar' patients are treated in two randomly selected hospitals, one patient will have twice the odds of a poor outcome compared to the other. A further advantage of the MOR is that it can be directly compared with the odds ratios of other covariates in the model, for example the effect of a certain patient characteristic or treatment.[7] An alternative measure is the intraclass correlation coefficient (ICC). However, the ICC is a less straightforward measure for interpreting dichotomous outcomes compared to continuous outcomes, because in linear regression the individual and the hospital-level variance are measured on the same scale, while in logistic regression individual and hospital level variance are not directly comparable.[8] Overall, the MOR seems a more attractive measure for quantifying variation in outcome. The MOR is directly comparable with fixed-effects odds ratios,[7] is not statistically dependent on the number of events, and it quantifies the level variance on the commonly known odds ratio scale.[8]

Summary on statistical uncertainty

Disregarding statistical uncertainty can have a significant negative impact on the reliability of outcome indicators for quality of hospital care.

We found that estimated outcome rates per hospital can be very uncertain. Ignoring this statistical uncertainty, by applying fixed-effect models, leads to an overestimation of outcome differences between hospitals. Therefore,

random effect models are preferable because they prevent over-interpretation. The lower the patient numbers and event rates, the bigger the problem of statistical uncertainty. Consequently outcomes that occur more frequently are more attractive as outcome indicators.

Case-mix correction

Not every patient carries the same risk of a specific outcome; for example, more severely ill patients carry a higher risk of having complications or dying compared to less severely ill patients. This risk difference would not be of significant importance if patients were randomly assigned to hospitals (or surgeons). As this is not the case,[9-11] the difference in patient characteristics may explain a part of the observed outcome variation between hospitals.

Research question 4 addressed the influence of case-mix correction on the validity of hospital quality indicators for colon and O-G cancer.

In Chapter 8 we demonstrated that postoperative mortality in O-G cancer was affected to some extent by patient characteristics. The effect of patient characteristics on postoperative mortality in colon cancer patients was more substantial (Chapter 7). Patients' characteristics, which are generally present in administrative data, showed a relatively high predictability (c-statistic: 0.76 (a measure of the discriminative ability of the model)). Cancer care is becoming increasingly centralized. This makes risk adjustment, especially for tumour- and disease-related factors, even more important since patient profiles will increasingly vary between specialized and non-specialized hospitals. Adding these factors increased the c-statistic to 0.82.

We also tested the effect of adding these specific characteristics to the case-mix model on the variation in postoperative mortality following colon cancer surgery between hospitals. It was found that, compared to the unadjusted mortality rates, the effect of general case-mix correction on the postoperative mortality estimates per hospital was limited. This may be explained by the fact that a case-mix variable is important when it is predictive of the outcome, and varies between hospitals.[12] The c-statistic only reflects the between-patient outcome variation, but does not take into account the differences in the distribution of the patient characteristics across the hospitals and therefore could be misleading. In our study population the general patient characteristics, such as age, were equally distributed across the centres. Disease-specific variables, such as pre-operative tumour complications, varied more across the centres. Therefore, including disease-specific correction variables in addition to the general case-mix variables

did affect the mortality estimates per hospital more. We therefore conclude that disease-specific case-mix correction is important and that the c-statistic alone is not very useful for quantifying the credibility and accuracy of a case-mix correction model.[13]

Our findings for the outcome anastomotic leakage were different. Compared to mortality, anastomotic leakages for both cancer conditions were to a much smaller extent determined by patient characteristics (e.g. colon cancer: c-statistic: 0.58). Adding tumour-specific information did not increase the predictive performance of the model (c-statistic: 0.60). A comparable AUC value was found in the O-G cancer patients (c-statistic: 0.59). The low predictability may be caused by anastomotic leakages being less attributable to patient characteristics. Another explanation may be that we missed important predictors for anastomotic leakage. However, this is rather unlikely, since we had two patient cohorts with rich patient and tumour information available, and similar findings for both cohorts.

We also tested the effect of adding tumour-specific characteristics to the case-mix model for comparison of anastomotic leakage rates between hospitals. For the overall hospital comparison, extended case-mix correction was shown to have a limited effect on the outcome anastomotic leakage.

In Chapter 3 we discussed why a case-mix correction model should contain variables that are related to the outcome and show variation between hospitals. Our literature review showed that for the readmission rate these variables have not been clearly identified yet. Although numerous case-mix correction models with varying methodology, geographical characteristics, underlying patient groups and included variables have been developed, little consensus exists on which variables actually are better at predicting readmissions.[14,15] Information with regard to the patient's disease severity and socioeconomic status, factors which are thought to majorly contribute to the likelihood of this outcome, are frequently missing in prediction models. Administrative data, which are usually used to calculate the indicator, do not contain this kind of information.[16,17]

Summary on case-mix correction

When outcomes are (partly) determined by patient characteristics and these patient characteristics vary across hospitals, case-mix correction is crucial for the validity of indicators used to measure quality of hospital care. There is no 'perfect' case-mix model, but our studies showed that disease-specific variables changed the results of the hospital comparisons. There-

fore, disease-specific variables should be included in case-mix correction models for comparing outcomes between hospitals.
Our empirical studies showed that patient characteristics do not have the same effect on all outcome measures. Some quality indicators are more robust and have less need for case-mix correction than others (e.g. anastomotic leakages vs. postoperative mortality).

GENERALIZABILITY OF FINDINGS

Our case studies on data issues were based on national quality indicators based on hospital administrative data of the years 2011 and 2012. We investigated two indicator sets based on administrative data. Since every indicator set can show different levels of reliability, depending on the amount of labour needed to calculate the indicators, we conclude that our findings on reliability aspects of hospital quality indicators based on administrative self-reported data can only be generalized with caution for the Netherlands and other health care systems.

However, we may assume that our findings have implications for other administrative databases which use a self-reporting system and commonly defined indicators. This assumption is confirmed by international studies that have found significant variation in the data accuracy of administrative data.[18-21] However, some of the causes of low administrative data quality may have already been addressed in other countries. The Nordic countries, for example, have a long tradition of registry-based epidemiological research, which uses unique patient identifiers to enable the linkage between registries. This results in rich disease and treatment information (e.g. drug description in ambulatory care) and allows researchers to follow patients beyond hospital discharge.[22] Our finding that many indicators are inaccurately defined supports previous findings from other countries[23] (Chapter 4). Even in the US, where readmission rates are used for pay-for-performance schemes, research highlights problems in data quality and indicator definition.[24-26]

In this thesis we studied the influence of statistical uncertainty and case-mix correction using Dutch and English clinical audit data. Since the patient profiles of Dutch and English patients are not expected to significantly differ from those of oncological patients in other Western countries, our findings on the influence of case-mix correction when measuring quality of hospital care can be generalized to other health care settings where patient characteristics are expected to vary across hospitals. Statistical uncertainty is always influential when patient numbers and especially event rates are low, irrespective of the health care system in which they are measured.

IMPLICATIONS FOR RESEARCH

In this thesis we first addressed the reliability of indicators measuring quality of hospital care. We found that data quality is a very important determinant of the indicators' reliability. Future research should investigate ways to optimize data quality determining factors. Data quality is determined in part by the feasibility of abstracting and registering the necessary data elements in the data registration system.

Our research showed that indicators that were hard to abstract from the data registration systems showed low reliability, because they were difficult to calculate (it is believed that) these indicators were estimated (by hospital staff) instead of calculated. Data elements that require extensive efforts to be delivered need to be identified, their reliability tested and, in case of reliability issues, their generation needs to be facilitated. Further research is needed to try to minimise the burden of data registration. This may be done by investigating possible ways to link databases and to automatically calculate necessary data elements. In the Netherlands a promising initiative has been taken by the 'Netherlands Federation of University Medical Centres'.[27] The federation's aim is to create a uniform and one-off registration of medical information for multiple purposes, based on international standards for medical data.

In our studies ambiguous indicator definitions were shown to affect the validity of indicators measuring quality of hospital care. The identification of events related to quality of care represents a substantial challenge. Future research needs to deepen our understanding of what we are actually measuring. Therefore we need to focus on the different elements of the definition of a quality indicator (e.g. timeframe, type of outcome). Specifically for the readmission rate indicator, this implies investigating the optimal time frame of defining an admission as readmission for the different diseases. Further, the consequences of shortening or extending this timeframe need to be researched. Our research showed that even the basic distinction between planned and unplanned events is rarely made. Even if data systems do allow the distinction between planned and unplanned events to be made, some uncertainty remains. The decision whether a readmission was planned or unplanned is not always clear and objective. Research is needed to investigate alternative approaches, such as defining diagnoses in a way that will indicate a planned readmission before the patient is discharged. When a patient gets admitted within a certain timeframe the system can define the admission as an admission or readmission based on the a priori defined diagnoses.

In contrast to choosing a common timeframe, we propose that the timeframe should be investigated and chosen based on the index disease and the purpose of the quality indicator. A timeframe that is too short may miss related events, while a timeframe that is too large increases the likelihood of including unrelated events. On the other hand, a short timeframe may give more accurate insight into technical quality of hospital care, while a longer timeframe may be able to reflect overall system performance.

Patients do not always get readmitted to the index hospital nor do they always die within the in-hospital period. More research is needed to identify the patterns of patient journeys and the consequences of missing these events. Further, algorithms need to be developed to abstract quality indicators from administrative data.

Nevertheless, even when data registration is perfect, the person in charge of the data registration and coding may also influence the data quality. It will always be difficult to prevent the intentional manipulation of data.

Statistical uncertainty is largely caused by low patient and event rates. For statistical reasons, patient and event rates must not be too small. To increase the number of patients and events, the time period of investigation can be extended. Our research suggests that for O-G and colon cancer a time period of about five calendar years would provide the necessary number of patients to have enough statistical power to detect significant outliers between surgeons for typical short-term outcomes, such as mortality or anastomotic leakage. Using data over such a long time period, however, decreases the actionability and relevance of the data. Another way to increase the number of events and patients is to increase the unit of observation. For example, researchers could aggregate events over hospitals instead of at the surgeon level. But even when data is aggregated on the hospital level and for a relevant time period, event rates may be still too low. An alternative approach would be to expand the number of events by combining different outcomes such as surgical complications, including bleedings, infections, or anastomotic leakages, or readmissions and length of stay. The events could also be weighted to create composite scores. However, such aggregated results may be (more) difficult to interpret. Given these problems, research needs to focus on improving statistical methods in order to better deal with statistical uncertainty. Specifically, it should be determined whether there is a minimum number of patients and events that needs to be met in order to have an acceptable level of certainty, as was proposed by the Dutch Health Transparency Program (DHTP).

Further, researchers should determine what is the best way to present the results of between-hospital comparisons.

In our studies we showed that patient characteristics have a substantial impact on the results of patient-outcome comparisons across providers. Numerous questions regarding case-mix correction still need to be answered. Case-mix correction models for complications and readmissions were shown to have poor to average performance. It must be determined whether we can identify measurable, relevant variables that improve the case-mix correction models. In our research it was shown that disease severity and socio-economic status are crucial considerations for case-mix correction. However, these two factors are also the hardest to capture using current data systems. The ability to study the effect of these variables needs to be enhanced, either by linking administrative data to detailed clinical databases or by collecting socioeconomic data of patients. The required extent of correction could be debated. For example, correcting for socio-economic status does not provide hospitals an incentive to improve their quality of the care specifically for lower socio-economic groups.

Further, there are case-mix factors that have not yet been well researched which might be especially important in a specialized/centralized care setting. For example, if a patient is admitted for palliative care, death is not an undesirable outcome. However, current data systems do not allow researchers to distinguish curative from palliative deaths. Further, severely ill patients are transferred to specialized centres. Despite the fact that these patients carry a higher risk of undesired outcomes, whether a patient is transferred or not is usually not captured in the data.

In the context of centralization, where severely ill patients are treated in specialized centres and their risk profile differs significantly from that of patients in small, not specialized centres, it is of interest to investigate whether the effect of case-mix variables varies between specialized and non-specialized centres. This can be done by testing the interaction between centres and the case-mix variable of interest or by including random slopes in the random effect models.

In our studies it was also found that the discriminative ability was non-optimal in judging the quality of a case-mix correction model. Research needs to focus on how to better quantify the performance of a model and to determine whether a model can be regarded as 'good enough' to be accepted as a valid model.

A number of indicators are currently used to measure quality of hospital care. We investigated the suitability of some of them in this thesis. Future research needs to investigate complications as outcome measures for surgical care since the occurrence of complications may be a more reliable and valid indicator of quality of care than mortality. We mainly studied outcome indicators. Process indicators, however, are usually more actionable, since they are directly attributable to the care process. Therefore process indicators should be further researched for their ability to quantify and compare quality of hospital care.

Indicators that capture the patient's journey through primary and secondary care need to be further developed, as such indicators provide a more complete picture of the care delivered for chronic patients.

In summary, future research should focus on the following topics:

- Increase feasibility of abstracting and registering relevant data elements in data registration systems.
- Minimize the burden of data registration (e.g. through data linkage).
- Investigate the elements of the definition of a quality indicator (e.g. type of outcome).
- Identify patterns of patient journeys.
- Investigate the consequences of missing events that occur in hospitals other than the index hospital.
- Further develop ways to abstract quality indicators from administrative data.

- Improve statistical methods to deal with statistical uncertainty.
- Further improve case-mix correction models and define criteria to judge case-mix correction models.

IMPLICATIONS FOR POLICY

The availability of information on quality of care has become a cornerstone of healthcare, perhaps even more so in health care systems with a market mechanism. However, for the Netherlands it was recently stated that the absence of reliable and valid quality information is a major weakness of the health care system.[32] In different countries, thousands of quality indicators are collected yearly; the data are made publicly available and are used for benchmarking and reimbursement strategies both on the hospital level as well as on the surgeon level. As this thesis shows, quality indicators have severe shortcomings that affect their reliability and validity. These shortcomings are not sufficiently taken into account currently in the Netherlands. The indicators we evaluated have flaws in all studied aspects: data quality, indicator definition, case-mix correction and statistical uncertainty. Ignoring these problems leads to a biased estimation of the indicator scores and leads to wrong conclusions on quality of care. Consequently, hospitals may be unjustifiably punished and reputations damaged.

Ideally, measuring quality of care leads to quality improvement.[33-35] It is argued, for example, that patients use hospital rankings to choose the 'best' provider. However, this is questionable. Research has shown that most patients actually do not use the information for decision making.[32] With regard to care providers, the evidence on whether reporting health care performance leads to quality improvement is mixed, especially when financial incentives are linked to performance.[33-36]

In order to be able to use quality information to improve quality of care, the quality measures must first be able to identify hospitals with real quality problems and second must provide insight into how to improve quality.. When there is a large amount of interest in quality information—for example, because of public reporting or when funding is linked to performance—it needs to be realized that invalid and unreliable quality indicators may cause perverse effects such as avoiding admission of high-risk patients. Consequently, to create the right incentives for quality improvement, quality indicators for hospital care need to be reliable and valid.

In this thesis the main outcome indicators investigated were readmissions, mortality and surgical complications. These indicators have good face validity and can potentially provide very useful insight into quality of hospital care. However, we identified many methodological issues with these outcome indicators.

Mortality has strong face validity as a measure of hospital care, as death is the most undesirable outcome for patients. However, our research showed that mortality is more influenced by patient characteristics than quality of care. Other research has also shown that actually only a small percentage of deaths that occurred in hospitals are preventable. Of the 1000 cases reviewed, 5.2% were judged to have been preventable (i.e. having a greater than 50% probability that better care would have prevented death).[37,38] Further, it was shown that patients dying in hospital are often already near the end of their lives. It was estimated that only 0.5% (95%CI; 0.3% - 0.7%) of the patients whose deaths were possibly related to quality of care were expected in the absence of quality of care problems to live at least 3 more months in good cognitive health.[37,39]

Readmissions are an important indicator of quality of care as they are associated with high costs. However, the indicator is often misunderstood. Readmissions do not seem to solely measure quality of hospital care, but instead reflect the overall quality of the health system and social care. Depending on the underlying disease, they are also determined by what happens to the patients post discharge, their social network, and the patient's compliance. Only unplanned readmissions are expected to relate to quality. Readmissions occurring soon after discharge (within seven days) are most likely related to hospital care and therefore might be the most preventable.[40]

The studied surgical quality indicator, anastomotic leakage, proved to be a possibly useful indicator for measuring surgical care. Anastomotic leakage occurs relatively frequently, is relatively robust for patient characteristics, and is closely related to the surgical intervention. Yet, research has shown that there is no universally accepted definition of anastomotic leakage; in a systematic review, 56 different definitions of anastomotic leakage were identified.[41] Having no uniform definition of the indicator complicates the assessment and comparison across hospitals.

Although there is a tendency to prefer outcome indicators over process indicators, the discussion on which type of indicator better reflects quality of hospital care is ongoing. Both indicator types have advantages and disadvantages. While outcome indicators are highly relevant for patients and payers and provide information on all aspects of delivered care, the events they quantify occur infrequently, they are confounded by factors unrelated to quality of care, and opportunities for improvement are hard to identify based on outcomes. Process indicators are generally evidence-based, are directly attributable to delivered care (actionable), usually do not require

case-mix correction, and have no time lags. However, process indicators also just represent what can be measured; they are less meaningful to patients.

One indicator can never provide a full picture of all aspects of the delivered care. Therefore we recommend investigating sets of indicators that include structure, process and outcome indicators. This approach is also needed because some indicators may outweigh others. For example, a hospital with a long length of stay may score low on the readmission rate indicator, while a hospital with a high score on the readmission rate indicator may score low on the mortality rate indicator. The difficulty is that it is also known that patients prefer quality of care information to be presented in a simple form, such as a single number.[35] Therefore different stakeholders may require different presentation of quality information.

RECOMMENDATIONS

Based on this thesis some specific recommendations can be made for the development and use of quality indicators that are used to assess the quality of hospital care in the Netherlands for external purposes.

Indicator selection

- Create indicator sets that contain structure, process and outcome indicators for reasons of actionability.
- For outcome indicators, choose indicators and data sources that allow case-mix correction, including disease-specific case-mix factors.
- For outcome indicators, choose indicators for which a reasonable number of patients in the numerator can be expected.

Data registration and indicator definition

- Strengthen data quality control.
- Maximize the opportunity to use nationally unique patient identifiers (UPIs) to enable linkage of administrative databases and clinical data.
- Make indicator definitions as precise as possible to eliminate differences in interpretation.
- Reduce the total registration burden by reducing the number of indicators, making more efficient use of existing data, and automating data collection.
- Request hospitals to submit patient-level data in order to uniformly abstract the numerator and denominator for the indicator.

Analysis

- Use random effect models to take statistical uncertainty into account.
- Adjust for case-mix differences between hospitals.

Use

- Interpret currently publicly presented outcomes with caution as their validity and reliability is often poor.
- Focus on outcome indicators that are relatively robust to case-mix differences and occur relatively frequently, such as complications after surgery.
- Do not use outcome indicators that are not sufficiently corrected for case-mix factors or occur infrequently for public reporting or financial purposes.

REFERENCES

1. Copnell B, Hagger V, Wilson SG, Evans SM, Sprivulis PC, Cameron PA. Measuring the quality of hospital care: an inventory of indicators. Intern Med J. 2009 Jun;39(6):352-60.
2. Kimberlin CL, Winterstein AG. Validity and reliability of measurement instruments used in research. Am J Health Syst Pharm. 2008 Dec 1;65(23):2276-84.
3. Pouw ME, Peelen LM, Moons KG, Kalkman CJ, Lingsma HF. Including post-discharge mortality in calculation of hospital standardised mortality ratios: retrospective analysis of hospital episode statistics. Bmj. 2013;347:f5913.
4. Talsma AK, Lingsma HF, Steyerberg EW, Wijnhoven BP, Van Lanschot JJ. The 30-day versus in-hospital and 90-day mortality after esophagectomy as indicators for quality of care. Annals of surgery. 2014 Aug;260(2):267-73.
5. Lingsma H. Measuring quality of care. Methods and applications to acute neuroloigcal diseases [thesis]. Alphen a/d Rijn: ErasmusMC; 2010.
6. Lingsma HF, Roozenbeek B, Perel P, Roberts I, Maas AI, Steyerberg EW. Between-centre differences and treatment effects in randomized controlled trials: a case study in traumatic brain injury. Trials. 2011;12:201.
7. Larsen K, Merlo J. Appropriate assessment of neighborhood effects on individual health: integrating random and fixed effects in multilevel logistic regression. Am J Epidemiol. 2005 Jan 1;161(1):81-8.
8. Merlo J, Chaix B, Ohlsson H, Beckman A, Johnell K, Hjerpe P, et al. A brief conceptual tutorial of multilevel analysis in social epidemiology: using measures of clustering in multilevel logistic regression to investigate contextual phenomena. J Epidemiol Community Health. 2006 Apr;60(4):290-7.
9. Kolfschoten NE, Marang van de Mheen PJ, Gooiker GA, Eddes EH, Kievit J, Tollenaar R, et al. Variation in case-mix between hospitals treating colorectal cancer patients in the Netherlands. European Journal of Surgical Oncology (EJSO). 2011;37(11):956-63.
10. Khuri SF, Najjar SF, Daley J, Krasnicka B, Hossain M, Henderson WG, et al. Comparison of surgical outcomes between teaching and nonteaching hospitals in the Department of Veterans Affairs. Annals of surgery. 2001;234(3):370.
11. Friese CR, Earle CC, Silber JH, Aiken LH. Hospital characteristics, clinical severity, and outcomes for surgical oncology patients. Surgery. 2010;147(5):602-9.
12. McNamee R. Regression modelling and other methods to control confounding. Occupational and environmental medicine. 2005;62(7):500-6.
13. Austin PC, Reeves MJ. The relationship between the C-statistic of a risk-adjustment model and the accuracy of hospital report cards: a Monte Carlo Study. Med Care. 2013 Mar;51(3):275-84.
14. Brooke BS, De Martino RR, Girotti M, Dimick JB, Goodney PP. Developing strategies for predicting and preventing readmissions in vascular surgery. J Vasc Surg. 2012 Aug;56(2):556-62.
15. van Walraven C, Wong J, Hawken S, Forster AJ. Comparing methods to calculate hospital-specific rates of early death or urgent readmission. Canadian Medical Association Journal. 2012;184(15):E810-E7.
16. Fischer C, Lingsma HF, Marang-van de Mheen PJ, Kringos DS, Klazinga NS, Steyerberg EW. Is the readmission rate a valid quality indicator? A review of the evidence. PLoS One. 2014;9(11):e112282.
17. Johnson T, Bardhan J, Odwazny R, Harting B, Skarupski K, McNutt R. Hospital care may not affect the risk of readmission. Qual Manag Health Care. 2012 Apr-Jun;21(2):68-73.
18. Burns EM, Rigby E, Mamidanna R, Bottle A, Aylin P, Ziprin P, et al. Systematic review of discharge coding accuracy. J Public Health (Oxf). 2012 Mar;34(1):138-48.
19. Aylin P, Lees T, Baker S, Prytherch D, Ashley S. Descriptive study comparing routine hospital administrative data with the Vascular Society of Great Britain and Ireland's National Vascular Database. Eur J Vasc Endovasc Surg. 2007 Apr;33(4):461-5; discussion 6.
20. Quach S, Blais C, Quan H. Administrative data have high variation in validity for recording heart failure. Can J Cardiol. 2010 Oct;26(8):306-12.
21. Powell H, Lim LL, Heller RF. Accuracy of administrative data to assess comorbidity in patients with heart disease. an Australian perspective. Journal of clinical epidemiology. 2001 Jul;54(7):687-93.
22. Furu K, Wettermark B, Andersen M, Martikainen JE, Almarsdottir AB, Sorensen HT. The Nordic countries as a cohort for pharmacoepidemiological research. Basic Clin Pharmacol Toxicol. 2010 Feb;106(2):86-94.

23. Fischer C, Anema HA, Klazinga NS. The validity of indicators for assessing quality of care: a review of the European literature on hospital readmission rate. Eur J Public Health. 2012 Aug;22(4):484-91.
24. Sacks GD, Dawes AJ, Russell MM, Lin AY, Maggard-Gibbons M, Winograd D, et al. Evaluation of hospital readmissions in surgical patients: do administrative data tell the real story? JAMA Surg. 2014 Aug;149(8):759-64.
25. Fisher ES, Whaley FS, Krushat WM, Malenka DJ, Fleming C, Baron JA, et al. The accuracy of Medicare's hospital claims data: progress has been made, but problems remain. Am J Public Health. 1992 Feb;82(2):243-8.
26. Kressin NR, Chang BH, Hendricks A, Kazis LE. Agreement between administrative data and patients' self-reports of race/ethnicity. Am J Public Health. 2003 Oct;93(10):1734-9.
27. Hazelzet J, Georgieva P. Registratie aan de bron. Visie op documentatie en gebruik van zorggegevens 2013-2020. NFU- Nederlandse Federatie van Universitair Medische Centra. 2013; Available from: http://www.nfu.nl/img/pdf/13.3694_Brochure_Registratie_aan_de_bron_versie_4-7-13.pdf [Accessed 11.07.2015]
28. Kleef van R SE, Ven van de W, . Evaluatie zorgstelsel en risicoverevening. Acht jaar na invoering Zorgverzekeringswet: succes verzekerd? : Erasmus Universiteit Rotterdam instituut Beleid & Management Gezondheidszorg2014.
29. Berwick DM, James B, Coye MJ. Connections between quality measurement and improvement. Med Care 2003; 41(1 Suppl): I30-8.
30. Fung CH, Lim YW, Mattke S, Damberg C, Shekelle PG. Systematic review: the evidence that publishing patient care performance data improves quality of care. Ann Intern Med 2008; 148(2): 111-23.
31. Hibbard JH. What can we say about the impact of public reporting? Inconsistent execution yields variable results. Ann Intern Med 2008; 148(2): 160-1.
32. lake C DH. Why don't the Dutch use quality information in their hospital choice? Results from a survey among 479 patients from a Dutch hospital. Health 2014;6(1):1-5.
33. Asch DA, Werner RM. Paying for performance in population health: lessons from health care settings. Prev Chronic Dis. 2010 Sep;7(5):A98.
34. Petersen LA, Woodard LD, Urech T, Daw C, Sookanan S. Does pay-for-performance improve the quality of health care? Ann Intern Med. 2006 Aug 15;145(4):265-72.
35. Fung CH, Lim YW, Mattke S, Damberg C, Shekelle PG. Systematic review: the evidence that publishing patient care performance data improves quality of care. Ann Intern Med. 2008 Jan 15;148(2):111-23.
36. Shekelle PG LY, Mattke S, et al. Does public release of performance results improve quality of care? A systematic review. The Health Foundation; 2008; Available from: http://www.health.org.uk/public/cms/75/76/313/554/Public%20release%20of%20performance%20result.pdf?realName=UWXIXp.pdf [Accessed 23.05.2015].
37. Shojania KG. Deaths due to medical error: jumbo jets or just small propeller planes? BMJ Qual Saf. 2012 Sep;21(9):709-12.
38. Hogan H, Healey F, Neale G, Thomson R, Vincent C, Black N. Preventable deaths due to problems in care in English acute hospitals: a retrospective case record review study. BMJ Qual Saf. 2012 Sep;21(9):737-45.
39. Hayward RA, Hofer TP. Estimating hospital deaths due to medical errors: preventability is in the eye of the reviewer. JAMA. 2001 Jul 25;286(4):415-20.
40. van Walraven C, Jennings A, Taljaard M, Dhalla I, English S, Mulpuru S, et al. Incidence of potentially avoidable urgent readmissions and their relation to all-cause urgent readmissions. CMAJ. 2011 Oct 4;183(14):E1067-72.
41. Bruce J, Krukowski ZH, Al-Khairy G, Russell EM, Park KG. Systematic review of the definition and measurement of anastomotic leak after gastrointestinal surgery. The British journal of surgery. 2001 Sep;88(9):1157-68.

SUMMARY

Introduction
Information on quality of care plays a central role in healthcare nowadays. As a result of the growing demand for health care, increasing costs and evidence of variation in quality of care, the interest in the quantitative assessment of health care quality has increased. Furthermore, today's society demands ever more and more transparency, which requires the health care sector—as well as other public sectors—to provide insight into their performance.
Different stakeholders may use quality of care information for different purposes. Health care professionals use it to evaluate their performance in order to try to improve quality. Government agencies use quality of care information to monitor and regulate quality of care, while insurance companies use it to select hospitals they want to contract, and patients may use it to compare hospitals in order to make informed treatment decisions.
As quality of care information is used for such a variety of purposes, it is crucial that quality measures are valid and reliable and actually do represent quality of care. However, despite a rapidly growing body of scientific literature on this topic, there is currently no consensus on how to measure quality of care.

Quality of care is commonly measured with quality indicators. Quality indicators are measurement tools, screens, or flags that are used as guides to monitor, evaluate, and improve quality of care. They are often classified into three types: structure (e.g. nurse-to-bed ratio), process (e.g. proportion of patients who received discharge instructions) and outcome (e.g. proportion of patients who had to undergo reoperation) indicators. While in the scientific literature no consensus exists on which type of quality indicator is best able to assess quality of care, in the last few years there has been a clear preference for outcome indicators.

Using outcome indicators as a measure of quality of hospital care requires that differences in outcomes between hospitals represent differences in underlying quality of care. Therefore outcome indicators need to be evaluated in terms of their psychometric criteria: how 'good' is the indicator?

In this thesis we concentrate on the psychometric criteria reliability and validity. We study the following elements that determine the reliability and validity of an outcome indicator: data quality, definitions, statistical uncertainty and case-mix (Table 1).

Table 1. Psychometric criteria to evaluate an outcome indicator studied in this thesis

Psychometric characteristic		
Reliability:	Data quality:	Uniformity in data (collection)
	Definitions:	Accurate definition of numerator and denominator and other data elements that are used to calculate the indicator
	Statistical uncertainty:	Random variation caused by low numbers of patients/outcomes
Validity:	Case-mix:	Differences in patient populations between hospitals

We investigate different outcome indicators for quality of hospital care in different disease fields with the main aim of expanding our knowledge on how to measure quality of hospital care for between-hospital comparisons.

The specific research questions are the following:

1. *Which aspects of data quality affect the reliability of quality indicators for hospital care?*
2. *How clear are the definitions of currently used quality indicators for hospital care?*
3. *To what extent does statistical uncertainty affect the reliability of outcome indicators for surgical colon and oesophago-gastric cancer care?*
4. *To what extent does case-mix correction affect the validity of outcome indicators for surgical colon and oesophago-gastric cancer care?*

These questions are answered in eight chapters. Table 2 presents an overview of these chapters. The table shows the indicators studied, the clinical fields, and the data sources, including year and country.

Reliability and validity

In part II (literature studies) and part III (empirical studies) the impact of data issues, statistical uncertainty and case-mix correction on the reliability and validity of indicators measuring quality of hospital care is studied.

In the systematic literature review in Chapter 2, we demonstrate that the quality indicator rate of readmission is often used in the European context. However, it is seldom accurately defined. The 486 included studies showed limited consensus on the time frame that should be used to calculate the indicator and the types of readmission to consider.

Table 2. Outcome indicators for the quality of hospital care studied in this thesis

Outcome indicator	Clinical field	Datasource	Country	Year	Chapter
Readmission	General	Literature	International	1999-2010	2
	General	Literature	International	2001-2013	3
	Heart failure	Literature	International	/ -2014	4
Relapse	Breast cancer	Administrative	NL	2009-2011	5
Complications	Hip and knee replacement	Administrative	NL	2009-2011	5 and 6
	Colon cancer	Clinical	NL	2011-2012	7
	Oesophago-gastric cancer	Clinical	UK	2011-2013	8 and 9
Mortality*	Colon cancer	Clinical	NL	2011-2012	7
	Oesophago-gastric cancer	Clinical	UK	2011-2013	8 and 9

*postoperative mortality, 30-day mortality, 90-day mortality.

The literature study in Chapter 3 summarizes methodological aspects of the definition and measurement of the readmission rates that need to be considered when interpreting readmission rates as a reflection of quality of hospital care. We found that administrative data is usually suboptimal for calculating quality indicators for external purposes, such as pay-for–performance schemes. Data quality problems include heterogeneous data information systems, inaccurate data registration, and incomplete data. Further, we found that defining the context in which readmissions are used is crucial. This context includes the patient group and which elements of quality of care the readmissions rate is expected to reflect. Finally, the multi-faceted nature of quality of care and the correlation between readmissions and other outcomes limit the indicator's validity.

In Chapter 4 the association between heart failure process indicators and the outcome indicator readmission rate was investigated. We found that readmission rates after heart failure hospitalization are mostly unrelated to the evidence-based ACC/AHA in-hospital process indicators for heart failure. The strongest evidence was found for the process indicator 'ACE inhibitor/ARB blocker use'; four out of ten studies showed a significant correlation between this indicator and readmission rates. It is unclear whether readmissions are largely determined by processes of hospital care; they may be more strongly influenced by post-discharge care or patient characteristics.

In Chapters 5 and 6 we present our studies on the data quality of the Dutch administrative database DHTP (Dutch Healthcare Transparency Program) in the years 2008 and 2009. In Chapter 5 we investigated the extent to which

local hospital data collection and indicator computation strategies differ in the Netherlands and how these differences affect the plausibility of self-reported indicator scores on hip and knee replacement and breast cancer indicators. We used survey results from 42 Dutch hospitals. We found that the data collection and indicator computation strategies varied widely between the hospitals. The hip and knee replacement quality indicator scores were often implausible, which was mostly due to poor data registry. In contrast, breast cancer quality indicator scores were more plausible.
In Chapter 6 we used the national data of the DHTP scores to assess the construct validity of structure, process and outcome indicators for hip replacement. Among the 28 hypothesized correlations between the hospitals' scores on different indicators (e.g. high percentage of antibiotics provided is correlated with a low percentage of wound infections), just three correlations were in the direction hypothesized and statistically significant. Hospitals with a low percentage of wound infections scored high on scheduling postoperative appointments (p-value=0.001) and high on not transfusing homologous blood (p-value=0.05). Hospitals with higher scores on scheduling complication meetings also scored higher on providing thrombosis prophylaxis (p-value= 0.04). Our results showed that although the individual indicators might be valid and actionable, drawing overall conclusions based on the whole indicator set should be done carefully, as construct validity could not be established. Factors that may explain the lack of construct validity are poor data quality, no adjustment for case-mix and statistical uncertainty.

The literature study in Chapter 3 showed that hospital quality indicators are usually based on administrative data. However, administrative data contain a limited set of (clinical) patient characteristics which makes sufficient case-mix correction difficult. Administrative data often do not contain disease-specific information or information on the patient's socio-economic status, while both have been shown to be important case-mix correction variables.
In Chapter 7 we investigated the impact of statistical uncertainty and case-mix correction on postoperative mortality and anastomotic leakage rates amongcolon cancer patients. We studied 13,120 patients who underwent colon cancer resection in 85 Dutch hospitals. We addressed differences between hospitals in rates of postoperative mortality and anastomotic leakage using fixed (ignoring statistical uncertainty) and random effects (incorporating statistical uncertainty) logistic regression models with general and disease-specific case-mix correction. It was found that overall statistical uncertainty had a larger effect on hospital performance than extended case-mix correction, although some individual hospital outcome rates were affected

by more detailed case-mix correction. This finding highlights the importance of considering statistical uncertainty caused by low hospital case volume when comparing outcomes between hospitals.
In Chapter 8 we developed a case-mix correction model to adjust for patient characteristics when comparing short-term outcomes in oesophago-gastric cancer patients. We used data from the English national audit on oesophago-gastric cancer, which included 4882 patients. The models for 30-day mortality, 90-day mortality and anastomotic leakage were internally validated with bootstrapping techniques, showing moderate discriminative ability (AUC 0.65 for 30-day mortality, 0.66 for 90-day mortality and 0.59 for anastomotic leakage). The models can be used for risk adjustment when assessing hospital performance in the NHS or other large health systems.
In Chapter 9 we assessed differences in short-term outcomes following oesophago-gastric cancer surgery between hospitals and surgeons and the influence of hospital and surgeon volume on these outcomes. We used 4868 cases of the National Oesophago-Gastric Cancer Audit (NOGCA) from the year 2011 to 2013 to study the outcomes 30-day mortality, 90-day mortality and anastomotic leakage. Multivariate logistic regression models were used to quantify the effects of surgeon and hospital volume on outcome. Outcome differences between surgeons and hospitals were quantified by multivariable random effects logistic regression models, including a surgeon and hospital level, to take into account case-mix and statistical uncertainty. Our analysis showed that higher hospital volume was associated with lower 30-day mortality. Higher surgeon volume was associated with lower anastomotic leakage rates. Hospital volume explained a part of the between-hospital variation in 30-day mortality, whereas surgeon volume explained part of the between-hospital variation in anastomotic leakage. The effect of centralization was inconsistent and depended on the type of outcome measure under consideration. Therefore, efforts to centralise O-G cancer services further should carefully address the effects this will have on outcomes at both the hospital and surgeon level.

Discussion

The aim of this thesis was to expand our knowledge on how to measure quality of hospital care for between-hospital comparisons. We studied reliability and validity of quality indicators and specifically focused on data quality, indicator definitions, statistical uncertainty and case-mix correction.

The specific research questions were:

1. *Which aspects of data quality affect the reliability of quality indicators for hospital care?*
2. *How clear are the definitions of currently used quality indicators for hospital care?*
3. *To what extent does statistical uncertainty affect the reliability of outcome indicators for surgical colon and oesophago-gastric cancer care?*
4. *To what extent does case-mix correction affect the validity of outcome indicators for surgical colon and oesophago-gastric cancer care?*

We found that poor data quality and unclear indicator definitions are common problems when using administrative data for the purpose of measuring quality of care. Important issues that affect data quality are heterogeneous data information systems, inaccurate data registration, and incomplete data. We further found that many indicators are not clearly defined. For example, the distinction between planned and unplanned readmission is seldom made. Ignoring these reliability factors may lead to biased comparisons between hospitals.

Our analyses showed that disregarding statistical uncertainty and case-mix correction can have a significant negative impact on the reliability and validity of outcome indicators for quality of hospital care. We found that estimated outcome rates per hospital can be very uncertain and ignoring this statistical uncertainty can lead to an overestimation of the differences between hospitals. We found that patient characteristics do not have the same effect on all outcome measures. Some quality indicators were shown to be more robust towards case-mix correction than others (e.g. anastomotic leakage vs. postoperative mortality). When outcomes are (partly) determined by patient characteristics and these patient characteristics vary across hospitals, case-mix correction, including disease-specific variables, is crucial for the reliability and validity of outcome indicators.

Our findings have several implications regarding future research. Future research is needed to optimize data quality. As data quality is determined by the feasibility of abstracting and registering the necessary data, research should focus on minimizing the burden of data registration. This may be done by investigating possibilities to link databases and to automatically calculate necessary data elements. Further, more research is needed to identify the patterns of patient journeys and to gain insight into the consequences of missing events that occur outside the hospital.

Further, research needs to focus on statistical methods on how to better deal with statistical uncertainty and how to present the results of between-hospital comparisons. Specifically, it should be determined whether there is a specific minimum number of patients and events that needs to be met in order to have an acceptable level of uncertainty. With regard to case-mix, we found that case-mix correction models for complications and readmissions had just poor to average performance. It needs to be determined whether the case-mix correction models perform poorly because the events are not predicted by patient factors or because relevant variables have been missed. Further, research needs to focus on how to quantify the performance of a case-mix correction model and how to decide if a model can be regarded as 'good'.

From a policy perspective, in order to be able to use quality information to improve quality of care, the quality measures must first be able to identify hospitals with real quality problems and second must provide insight into how to improve quality. When there is a large amount interest in quality information—for example, because of public reporting or when funding is linked to performance—it needs to be realized that invalid and unreliable quality indicators may cause perverse effects such as avoiding high-risk patients. Consequently, to create the right incentives for quality improvement, quality indicators for hospital care need to be reliable and valid.

To improve the reliability and validity of quality indicators that are used to assess the quality of hospital care in the Netherlands for external purposes, some specific recommendations can be made based on this thesis:

Indicator selection

- Create indicator sets that contain structure, process and outcome indicators for reasons of actionability.
- For outcome indicators, choose indicators and data sources that allow case-mix adjustment, including disease-specific case-mix factors.
- For outcome indicators, choose indicators for which a reasonable number of patients in the numerator can be expected.

Data registration and indicator definition

- Strengthen standard based data quality control.
- Maximize the opportunity to use nationally unique patient identifiers (UPIs) to enable linkage of administrative databases and clinical data.
- Make indicator definitions as precise as possible to eliminate differences in interpretation.
- Reduce the total registration burden by reducing the number of indicators, making more efficient use of existing data, and automating data collection.
- Request that hospitals submit patient-level data in order to uniformly abstract the numerator and denominator for the indicator.

Analysis

- Use random effect models to take statistical uncertainty into account.
- Adjust for case-mix differences between hospitals.

Use

- Interpret currently publicly presented outcomes with caution as their validity and reliability is often poor.
- Focus on outcome indicators that are relatively robust to case-mix differences and occur relatively frequently such as complications after surgery.
- Do not use outcome indicators that are not sufficiently corrected for case-mix factors or occur infrequently for public reporting or financial purposes.

SAMENVATTING

Introductie
Informatie over kwaliteit van zorg speelt tegenwoordig een centrale rol binnen ons zorgstelsel. Als gevolg van een toenemende vraag naar gezondheidszorg, stijgende kosten en aanwijzingen voor variatie in kwaliteit van zorg, is er een groeiende interesse in kwantitatieve beoordeling van de kwaliteit van de gezondheidszorg. Daarnaast is er vanuit de maatschappij steeds meer behoefte aan transparantie omtrent de prestaties van de zorgsector en andere publieke sectoren.

Informatie over kwaliteit van zorg wordt door verschillende belanghebbenden voor uiteenlopende doelen gebruikt. Zorgverleners kunnen hun eigen kwaliteit evalueren en verbeteren. Overheidsinstanties monitoren en reguleren kwaliteit van zorg, verzekeringsmaatschappijen selecteren ziekenhuizen om zorg in te kopen en patiënten kunnen ziekenhuizen vergelijken om daarna een ziekenhuis te kiezen.
Omdat de informatie over kwaliteit van zorg voor zoveel doeleinden wordt gebruikt, is het van groot belang dat de kwaliteitsmaten valide en betrouwbaar zijn. Ondanks de toename in wetenschappelijke literatuur over dit onderwerp, is er tot noch toe geen consensus bereikt over hoe kwaliteit van zorg moet worden meten.

Kwaliteit van zorg wordt over het algemeen gemeten aan de hand van kwaliteitsindicatoren. Kwaliteitsindicatoren zijn meetbare aspecten van kwaliteit van zorg die worden gebruikt om de kwaliteit van zorg te monitoren, evalueren en verbeteren. Deze indicatoren worden vaak in drie groepen verdeeld: structuur- (bijv. aantal verpleegkundigen t.o.v. het aantal bedden), proces- (bijv. proportie van patiënten die instructies krijgen bij ontslag uit het ziekenhuis), en uitkomstindicatoren (bijv. het aantal patiënten dat een heroperatie moet ondergaan). Terwijl in de wetenschappelijke literatuur geen consensus bestaat over welk type indicatoren zou moeten worden gebruikt om kwaliteit van zorg te beoordelen, is op dit moment een duidelijke trend te zien in de richting van uitkomstindicatoren. Zo worden bijvoorbeeld sterftecijfers gebruikt als maat voor kwaliteit van ziekenhuiszorg.

Om uitkomstindicatoren te kunnen interpreteren als maat voor kwaliteit van ziekenhuiszorg is het van belang dat verschillen in uitkomsten tussen ziekenhuizen daadwerkelijke de kwaliteit van de geleverde zorg representeren. Daarom dienen de indicatoren nauwkeurig te worden geëvalueerd aan de hand van psychometrische criteria; hoe 'goed' zijn eigenlijk de indicatoren zelf?

Dit proefschrift richt zich op twee psychometrische criteria: betrouwbaarheid en validiteit. We onderzoeken verschillende elementen die de betrouwbaarheid en validiteit van een uitkomstindicator kunnen beïnvloeden, namelijk de kwaliteit van de data waaruit de indicator berekend wordt, de gebruikte definities, statistische onzekerheid en case-mix (tabel 1).

Tabel 1. Psychometrische criteria ter evaluatie van uitkomstindicatoren die worden behandeld in dit proefschrift

Psychometrisch kenmerk		
Betrouwbaarheid:	Kwaliteit van data:	Uniformiteit van de data (dataverzameling)
	Definities:	Nauwkeurige definitie van teller en noemer en andere gegevens die worden gebruikt om de indicator te berekenen
	Statistische onzekerheid:	Willekeurige variatie veroorzaakt door lage aantallen patiënten/ uitkomsten
Validiteit:	Case-mix:	Verschillen in patiëntenpopulatie tussen ziekenhuizen

In dit proefschrift wordt de betrouwbaarheid en validiteit van verschillende uitkomstindicatoren voor verschillende ziektebeelden onderzocht. Doel hiervan is het vergroten van wetenschappelijke kennis over het meten van kwaliteit van ziekenhuiszorg om vergelijkingen tussen ziekenhuizen mogelijk te maken.

De specifieke onderzoeksvragen zijn:

1. *Welke aspecten van de kwaliteit van de data hebben invloed op de betrouwbaarheid van kwaliteitsindicatoren voor ziekenhuiszorg?*
2. *Hoe duidelijk zijn de definities van de kwaliteitsindicatoren die momenteel worden gebruikt voor de ziekenhuiszorg?*
3. *Hoe belangrijk is statistische onzekerheid voor de betrouwbaarheid van uitkomstindicatoren voor de kwaliteit van operatieve zorg bij slokdarm- en darmkanker?*
4. *Hoe belangrijk is case-mixcorrectie voor de validiteit van de uitkomstindicatoren voor de kwaliteit van operatieve zorg bij slokdarm- en darmkanker?*

Deze vragen zullen worden beantwoord in tien hoofdstukken. Tabel 2 geeft een overzicht van deze hoofdstukken. In de tabel zijn de onderzochte indicatoren, het ziektebeeld en de databron (met jaar en land van herkomst) weergegeven.

Tabel 2. Uitkomstindicatoren voor de kwaliteit van zorg die in dit proefschrift worden onderzocht

Uitkomstindicatoren in deze studies					Ziektebeeld	Databron	Land	Jaar
Heropname	**Sterfte***	**Naadlekkage**	**Wondinfecties**	**Terugval**				
Hoofdstuk 2 + 3					Algemeen	Literatuur		
Hoofdstuk 4					Hartfalen	Literatuur		
				Hoofdstuk 5	Borstkanker	Administratief	NL	2009-2011
			Hoofdstuk 5 +6		Heup/knieprothese	Administratief	NL	2009-2011
	Hoofdstuk 7	Hoofdstuk 7			Dikkedarmkanker	Klinisch	NL	2011-2012
	Hoofdstuk 8 +9	Hoofdstuk 8 +9			Slokdarmkanker	Klinisch	UK	2011-2013

*postoperatieve sterfte, 30-dagen sterfte, 90-dagen sterfte;

Betrouwbaarheid en validiteit

In deel I (literatuurstudies) en deel II (empirische studies) wordt de impact van dataregistratie en indicatordefinities op de betrouwbaarheid van indicatoren die de kwaliteit van zorg meten onderzocht. Uit de literatuurstudie in hoofdstuk 2 blijkt dat de kwaliteitsindicator "heropname in het ziekenhuis" vaak wordt gebruikt binnen Europa. Heropnames worden echter zelden nauwkeurig gedefinieerd. In de 486 geïncludeerde studies was er weinig consensus te vinden wat betreft de tijdsperiode waarbinnen de heropname zou moeten vallen en welke types van heropname zouden moeten worden meegeteld.

De literatuurstudie in hoofdstuk 3 vat de methodologische aspecten die zouden moeten worden overwogen bij het definiëren en meten van heropnames samen. We vonden dat administratieve data in veel gevallen

suboptimaal zijn voor het berekenen van kwaliteitsindicatoren voor externe doeleinden als het vergelijken van ziekenhuizen. Problemen met de datakwaliteit zijn onder andere: heterogene informatiesystemen, onnauwkeurige of foutieve dataregistratie en incomplete data. Daarnaast bleek de context waarin de heropname plaatsvindt van groot belang, bijvoorbeeld welke aspecten van kwaliteit van zorg de heropname naar verwachting weerspiegelen. Tenslotte wordt de betrouwbaarheid van de indicatoren beperkt door de vele verschillende aspecten die deel uit maken van kwaliteit van zorg.

In de literatuurstudie in hoofdstuk 4 hebben we het verband tussen procesindicatoren en de uitkomstindicator heropname onderzocht bij hartfalen. We vonden dat het aantal heropnames na een opname door hartfalen per ziekenhuis niet gerelateerd was aan de scores op de procesindicatoren voor hartfalen. Onduidelijk is daarom of heropnames worden bepaald door processen binnen het ziekenhuis. Heropnamen kunnen ook worden beïnvloed door de zorg die wordt geboden na ontslag uit het ziekenhuis.

In hoofdstuk 5 en 6 worden studies gepresenteerd over de betrouwbaarheid van de Nederlandse 'Zichtbare Zorg' indicatoren over het jaar 2008/09.
In hoofdstuk 5 onderzochten we de mate waarin de dataverzameling en de berekening van indicatoren verschilt tussen ziekenhuizen en welk effect dit heeft op de betrouwbaarheid van de indicatorscores op indicatoren voor heup- en knievervanging en borstkanker. Voor het onderzoek werden vragenlijsten naar de Nederlandse ziekenhuizen gestuurd. We vonden dat er grote verschillen zijn in dataverzameling en het berekenen van de indicatoren tussen de ziekenhuizen. De kwaliteitsindicatorscores voor de heup- en knieprotheses waren vaak onwaarschijnlijk, meestal als gevolg van slechte dataregistratie. De scores op de indicatoren voor borstkanker waren aannemelijker.

In hoofdstuk 6 werd gebruik gemaakt van de nationale database van 'Zichtbare Zorg' indicatoren voor het vaststellen van de constructvaliditeit van structuur-, proces- en uitkomstindicatoren voor heupvervangingen. Van de 28 correlaties tussen ziekenhuisscores op verschillende indicatoren die te verwachten waren (bv. een hoog percentage antibiotica voorgeschreven is gecorreleerd met een laag percentage aan wondinfecties), waren er slechts drie statistisch significant en in de veronderstelde richting gecorreleerd. Deze resultaten laten zien dat, al hoewel de indicatoren op zichzelf valide en bruikbaar zijn, men voorzichtig moet zijn in het trekken van algemene conclusies over meerdere indicatoren, aangezien samenhang tussen

indicatoren beperkt was. Factoren die invloed zouden kunnen hebben gehad op het gebrek aan correlatie zijn: slechte kwaliteit van de data, geen correctie voor case-mix en statistische onzekerheid.

In hoofdstuk 7 wordt de invloed van statistische onzekerheid en case-mix-correctie op postoperatieve sterfte en naadlekkage bij dikke darmkanker patiënten onderzocht. We gebruikten gegevens van 13.120 patiënten die een darmkanker resectie ondergingen in één van de 85 Nederlandse ziekenhuizen. We vergeleken de ziekenhuizen op basis van postoperatieve sterfte en de anastomotische lekkage met gebruik van zogenaamde *fixed effects modellen* (waarbij geen rekening wordt gehouden met statistische onzekerheid) en *random effects modellen* (waarbij wel rekening wordt gehouden met statistische onzekerheid). We corrigeerden zowel voor generieke als ziekte specifieke case-mix factoren. We vonden dat de statistische onzekerheid een groter effect had op de geschatte verschillen tussen ziekenhuizen dan de uitgebreide case-mixcorrectie. Alhoewel de uitkomsten van enkele ziekenhuizen wel werden beïnvloed door een gedetailleerde case-mixcorrectie. Deze bevinding onderstreept hoe belangrijk het is om rekening te houden met statistische onzekerheid bij het vergelijken van uitkomsten tussen ziekenhuizen, zeker als de uitkomst (per ziekenhuis) niet zo vaak voorkomt.

De literatuurstudie in hoofdstuk 3 laat zien dat kwaliteitsindicatoren voor ziekenhuizen vaak gebaseerd zijn op administratieve data. Echter, administratieve databases bevatten maar beperkte informatie over de (klinische) patiëntkarakteristieken. Administratieve data bevatten bijvoorbeeld vaak geen ziektespecifieke-informatie of informatie over de sociaal-economische status van de patiënt, terwijl beiden belangrijke case-mixcorrectie factoren zijn gebleken. De mogelijkheden voor case-mixcorrectie zijn dus beperkt op basis van administratieve data.

In hoofdstuk 8 is er een case-mixcorrectie model ontwikkeld om te corrigeren voor patiëntkarakteristieken, bij het vergelijken van korte termijn uitkomsten van slokdarmkankerpatiënten. Er werd gebruik gemaakt van Engelse nationale audit data met informatie over 4882 patiënten. De verschillende modellen voor 30-dagen mortaliteit, 90-dagen mortaliteit en naadlekkage hadden een beperkt onderscheidend vermogen. Desondanks kunnen de modellen worden gebruikt voor correctie voor patiëntkarakteristieken bij het vergelijken van uitkomsten van slokdarmkanker tussen Engelse ziekenhuizen.

In hoofdstuk 9 worden verschillen tussen ziekenhuizen en tussen chirurgen in kortetermijnuitkomsten (30-dagen mortaliteit, 90-dagen mortaliteit en naadlekkage) na een slokdarmkanker resectie bekeken. Daarnaast werd de invloed van ziekenhuis- en chirurgvolume op deze uitkomsten onderzocht. Voor de studie werd gebruik gemaakt van 4868 patiënten uit de Engelse nationale audit (2011 tot 2013). We gebruikten multivariabele logistisch regressie modellen om het effect van chirurg- en ziekenhuisvolume op de uitkomst te schatten. De verschillen in uitkomsten tussen ziekenhuizen en tussen chirurgen werden geschat met een multivariabel *random-effects* logistisch regressiemodel, met een level voor chirurg en voor ziekenhuis, zodat er rekening kon worden gehouden met zowel case-mix verschillen als statistische onzekerheid. We vonden dat een hoger ziekenhuisvolume geassocieerd is met een lagere 30-dagen mortaliteit. Een hoger chirurg-volume is geassocieerd met minder naadlekkage. Het bleek dat ziekenhuisvolume een deel van de variatie tussen ziekenhuizen in 30-dagen mortaliteit verklaarde. Chirurgvolume verklaarde een deel van de variatie in naadlekkage tussen ziekenhuizen. Het effect van volume verschilde dus tussen uitkomstmaten. Daarom zou bij eventuele verdere centralisatie van de medische zorg voor slokdarmkanker, moeten worden gekeken naar zowel de effecten van ziekenhuisvolume als van chirurgvolume.

Discussie

Het doel van dit proefschrift was om de kennis over het meten van kwaliteit van ziekenhuiszorg om ziekenhuizen te kunnen vergelijken te vergroten, specifiek over de betrouwbaarheid en validiteit van de kwaliteitsindicatoren. De onderzoeksvragen waren:

1. *Welke aspecten van de kwaliteit van de data hebben invloed op de betrouwbaarheid van kwaliteitsindicatoren voor ziekenhuiszorg?*
2. *Hoe duidelijk zijn de definities van de kwaliteitsindicatoren die momenteel worden gebruikt voor de ziekenhuiszorg?*
3. *Hoe belangrijk is statistische onzekerheid voor de betrouwbaarheid van uitkomstindicatoren voor de kwaliteit van operatieve zorg bij slokdarm- en darmkanker?*
4. *Hoe belangrijk is case-mixcorrectie voor de validiteit van de uitkomstindicatoren voor de kwaliteit van operatieve zorg bij slokdarm- en darmkanker?*

We vonden dat onvoldoende datakwaliteit en onduidelijke definities van indicatoren een negatieve invloed hebben op de betrouwbaarheid van kwaliteitsindicatoren, met name bij het gebruik van administratieve data

voor het meten van kwaliteit van zorg. Belangrijke aspecten die invloed hebben op de datakwaliteit zijn heterogeniteit in data informatiesystemen, inaccurate dataregistratie en incomplete data. Verder vonden we dat veel indicatoren niet duidelijk genoeg zijn gedefinieerd, bijv. het verschil tussen geplande en ongeplande heropname wordt maar zelden gemaakt. Het negeren van deze betrouwbaarheids-aspecten kan leiden tot vertekening in de vergelijking tussen ziekenhuizen.

We vonden dat als geen rekening wordt gehouden met statistische onzekerheid en case-mix, dit een grote negatieve invloed heeft op de validiteit van de uitkomstindicatoren voor kwaliteit van ziekenhuiszorg. De geschatte uitkomsten per ziekenhuis kunnen erg onzeker zijn en het negeren van deze statistische onzekerheid leidt tot het overschatten van verschillen tussen ziekhuizen. Case-mixcorrectie had een wisselend effect op de verschillende uitkomstmaten. Sommige kwaliteitsindicatoren bleken minder beïnvloed te worden door case-mixcorrectie(bijv. naadlekkage) dan anderen (bijv. mortaliteit). Als uitkomsten (deels) worden bepaald door patiëntkarakteristieken en als er een variatie is in deze karakteristieken tussen ziekhuizen, dan is een case-mixcorrectie essentieel voor de validiteit van de uitkomstindicatoren.

Op basis van deze conclusies kunnen aanbevelingen geformuleerd worden. Er is meer onderzoek nodig om de datakwaliteit voor de berekening van kwaliteitsindicatoren te optimaliseren. Omdat de datakwaliteit afhankelijk is van de mogelijkheid om de benodigde data te verkrijgen en te registreren, zou onderzoek zich moeten richten op het minimaliseren van de registratielast. Dit zou bijvoorbeeld gedaan kunnen worden door naar mogelijkheden te kijken om databases aan elkaar te koppelen en door het automatisch berekenen van bepaalde data-elementen. Verder is er ook meer onderzoek nodig om inzicht te krijgen in de consequenties van het niet meenemen van uitkomsten die buiten het ziekenhuis plaatsvinden.

Met betrekking tot de statistische onzekerheid, zou onderzoek zich moeten richten op statistische methoden die beter kunnen omgaan met statistische onzekerheid en op hoe ziekenhuisverschillen zouden moeten worden gepresenteerd aan verschillende belanghebbenden. Een specifieke vraag is of er een minimum aantal patiënten en uitkomsten is dat een acceptabel niveau van statistische onzekerheid garandeert. Voor de case-mixcorrectie modellen werd gevonden dat de modellen voor complicaties en voor heropnames matig presteerden. Er moet worden bekeken of deze modellen matig presteerden omdat de uitkomsten moeilijk kunnen worden voorspeld door patiënten factoren of dat er nog relevante variabelen missen in het

model. Verder moet onderzoek zich richten op hoe de prestatie van case-mixcorrectie modellen te kwantificeren en hoe te definiëren wanneer een model als 'goed' kan worden beschouwd.

Vanuit beleidsperspectief is een vereiste voor het verbeteren van kwaliteit op basis van kwaliteitsinformatie dat de indicatoren de slecht presterende ziekenhuizen kunnen identificeren, maar vervolgens ook inzicht geven in hoe deze zorg verbeterd zou kunnen worden. Naarmate de consequenties die aan informatie over kwaliteit van zorg worden verbonden groter zijn, bijvoorbeeld bij openbaarmaking of bij financiële consequenties, neemt de kans op ongewilde resultaten toe. Bijvoorbeeld het vermeiden van patiënten met hoge risico's door ziekenhuizen. Hieruit volgt dat om het gewilde resultaat te bereiken – het verbeteren van de kwaliteit van zorg – de kwaliteitsindicatoren voor ziekenhuiszorg betrouwbaar en valide moeten zijn.
Om de betrouwbaarheid en validiteit van de kwaliteitsindicatoren die in het Nederlandse zorgstelsel worden gebruikt voor vergelijkingen tussen ziekenhuizen te vergroten, doen wij enkele aanbevelingen op basis van dit proefschrift:

Selectie van indicatoren:
- Maak indicatorsets die zowel structuur-, proces- en uitkomstindicatoren bevatten omdat ziekenhuizen dan aanwijzingen hebben wat ze kunnen verbeteren.
- Voor uitkomstindicatoren: kies indicatoren en databronnen die case-mixcorrectie mogelijk maken, ook voor de ziektespecifieke case-mix-factoren
- Voor uitkomstindicatoren: kies indicatoren met een verwacht aantal patiënten in de teller van ten minste 15 per eenheid. Bijv. minstens 15 verwachte sterfgevallen per ziekenhuis per jaar.

Dataregistratie en indicator definities
- Strengere kwaliteitscontrole op de data
- Maak zoveel mogelijk gebruik van unieke nationale 'patient identifiers' om deze manier klinische en administratieve databases aan elkaar te kunnen koppelen
- Maak de indicatordefinities zo nauwkeurig mogelijk om verschillen in interpretatie te voorkomen
- Verlaag de registratielast door vermindering van het aantal indicatoren, een efficiënter gebruik van bestaande gegevens en geautomatiseerde dataverzameling

- Vraag ziekenhuizen om data op patiëntenniveau in te voeren, zodat teller en noemers van de indicatoren uniform uit de database kunnen worden gehaald

Analyse
- Gebruik random-effect modellen om rekening te houden met statistische onzekerheid
- Corrigeer voor case-mix verschillen tussen ziekenhuizen

Gebruik
- Wees voorzichtig bij het interpreteren van de uitkomstindicatoren die op dit moment worden gepubliceerd, de validiteit en betrouwbaarheid zijn vaak beperkt
- Gebruik uitkomstindicatoren die relatief weinig worden beïnvloed door case-mix verschillen en die relatief vaak voorkomen, zoals bijvoorbeeld complicaties na een chirurgische ingreep
- Maak uitkomstindicatoren waarbij niet voldoende is gecorrigeerd voor case-mix factoren, of die niet vaak genoeg voorkomen (minder dan 15 uitkomsten per jaar/per eenheid) niet openbaar en verbind er geen financiële consequenties aan

ACKNOWLEDGEMENTS

I definitely could have chosen a warmer country, but do I regret coming to the Netherlands? Definitely not! These past years here in the Netherlands have been such a great journey and doing my PhD was certainly a highlight of it.

Completing this PhD thesis would not have been possible without the guidance, contributions and encouragement of certain people who I would like to thank.

My gratitude goes first of all to Prof. Dr. **Ewout Steyerberg** my promotor. Ewout, I want to thank you for your trust from the early stages of my PhD project. You gave me a great deal of freedom and responsibility in my work, which made the work very exciting and helped me to learn so much over the last few years. You are an extraordinary researcher. I always left meetings that I had with you feeling positive, newly motivated and with fresh ideas (and a lot of work to do). Thank you for all your input, time and support.

I would like to thank Prof. **Niek Klazinga**, my promotor. Niek, I first met you when I sneaked into your course on Methods of Health Services Research at the Nihes Program in 2008. I am so grateful I did that because it was with you my work in the field of health services research started. Back then I had no idea that the following years we would work together but this happened and started with me working as an intern at the AMC. After this I worked with you in the European Health Services Research project and you were also involved during my PhD period in diverse projects. Throughout these years you have provided me with intellectual guidance, mentorship and support (and you accomplished all of this without usually being in the same country). I always enjoyed working with you and want to thank you for all that you have done for me.

Dr. **Hester Lingsma**, I couldn't have wished for a better copromotor. Hester, I am grateful for all your input; for reading the many drafts and for teaching me so much about science. I would also like to thank you for being patient with me and always supporting my ideas. Your ability to make complicated things easy and your sense of humor made it again and again a pleasure to work with you.

I would like to acknowledge the time and effort of the members of my **doctoral commission** for reviewing my thesis and for their presence at the defense ceremony.

Over the past few years I had the opportunity to work with many people on different projects, such as the Netherlands Federation of University Medical Centres (NFU) and the Dutch Healthcare Authority (NZA) project, the Integrated Cancer Centre South (IKZ), the Dutch Institute for Clinical Auditing (DICA), and the National Oesopahgo-gastric Cancer Audit (NOGCA) groups. I want to thank all of my **co-authors** and **project partners** for their collaboration, time and input.

A special thanks to Dr. **Oliver Groene** who provided me the opportunity to work with the NOGCA data at the Royal College of Surgeons in London for several months. Oliver, you and your lovely colleagues at the department made me feel very welcome. I had a very nice time in London (although you would sometimes understand my English better than my Austrian German). I am happy you are a part of my doctoral committee at my defense.

My kind **colleagues at the ErasmusMC** also did their part. Bouwine and Sandra, I am glad to have shared my time at the ErasmusMC with you. Thanks for being such "gezellige" roommates and for all your support. Sandra, our conference trip to Rio is such a nice memory. Nikki, Iris, and Hester thanks a lot for saving me a lot of aggravation and time by translating the summary of my thesis into Dutch. I also want to thank Sanne and the rest of the secretaries and ICT team for their great support. Caspar, thanks for your help with my statistical questions.

At the beginning of my PhD project I spent time working at the **Academic Medical Center (AMC)**. Many thanks to my former roommates and colleagues, I miss my time at the AMC!

The research master program of the Vrije Universiteit Amsterdam brought me to the Netherlands. **Maurits van Tulder** and **Marcel Adriaanse**, this master program enabled me to obtain a PhD position - thanks for all you have taught me. My **fellow** eight **LCD'ers**, thanks for the many nice memories I have of our study time. Tsjitske, Mariska and Mine, I'll never forget our trip to Turkey!

My **paranimfem Tsjitske**, I am so glad you are next to me at my defense. You have been a dear friend since the first day we met at the VU for our LCD master program. My **paranimfem Sandra**, we started our PhD projects together and we finished it together side by side at each other's defenses. So nice!

But doing a PhD requires more than just the direct, work-related help. Many dear people, from close by and from far away, indirectly helped me to finish this book by offering their love and support. I want to thank all of those **friends**, but especially my dear **family**.
Ein Doktorat zu machen, im Ausland zu leben (und dabei noch eine gute Tochter zu sein), ist nicht immer einfach. Wie sehr freu ich mich doch immer aufs nach Hause Kommen und euch alle wieder zu sehen. Xandi, ich habe so ein Glück dich als Schwester zu haben. Papa, Mama und Andi, Xandi: Danke für eure unermüdliche und liebevolle Unterstützung.
Dragi Ivane, srce moje, zahvalnost za toliko divnih stvari sto si ucinio za mene, kao i moje osjecaje prema tebi ne mogu opisati u par recenica. Zivot je pun iznenadenja i nikad ne znamo tocno sta nas ceka. Vjerujem da ono najbolje je jos pred nama.

ABOUT THE AUTHOR

Claudia Fischer was born in Sankt Andrae im Lavanttal in Austria on the 17th of June 1985. After obtaining her A-levels in 2004 she started studying health care management at the University of Applied Sciences in Feldkirchen, Austria. She finished her 'magistra' degree in 2008. Once finishing her magistra-thesis, which focused on the implementation of trainee programs from the for-profit sector in the hospital setting, she started the research master program 'Lifestyle and Chronic Disorders' at the Vrije Universiteit in Amsterdam in the Netherlands, from which she obtained her MSc degree in 2010.

After finishing her master degree she worked as a junior researcher at the department of Public Health at the Academic Medical Centre, Amsterdam. In 2010 she started a PhD project on methodological aspects of measuring quality of hospital care at the department of Public Health of the ErasmusMC under the supervision of Prof. Dr. Ewout Steyerberg, Prof. Dr. Niek Klazinga, and Dr. Hester Lingsma, which resulted in this thesis. After finishing her PhD thesis Claudia Fischer worked as 'team leader methodology' at the Dutch Institute for Clinical Auditing in the Netherlands. In December 2015 she will start to work as a research scientist at Philips Research in Cambridge/UK.

LIST OF PUBLICATIONS

Peer reviewed journals

Fischer C, Lingsma HF, Hardwick R, Cromwell D, Steyerberg EW, Groene O. Risk adjustment models for short-term outcomes after surgical resection for oesophago-gastric cancer patients. British Journal of Surgery. 2015: In press.

Fischer C, Steyerberg EW, Fonarow GC, Ganiats TG, Lingsma HF. A systematic review and meta-analysis on the association between quality of hospital care and readmission rates in patients with heart failure. American Heart Journal. 2015: In press.

Fischer C, Lingsma H, van Leersum N, Tollenaar RA, Wouters MW, Steyerberg EW. Comparing colon cancer outcomes: The impact of low hospital case volume and case-mix adjustment. European Journal of Surgical Oncology. 2015, 41(8):1045-53.

Fischer C, Lingsma H, Marang van de Mheen P, Kringos DS, Klazinga NS, Steyerberg EW. Is the readmission rate a valid quality indicator? A review of the evidence. PloS One. 2014: 9 (11).

Aghaei HA, Ravaghi H, Kringos DS, Ogbu UC, Fischer C, Azami SR, Klazinga NS. Using quality measures for quality improvement: the perspective of hospital staff. PLoS One. 2014: 23;9(1).

Anema HA, van der Veer SN, Kievit J, Krol-Warmerdam E, Fischer C, Steyerberg E, Dongelmans DA, Reidinga AC, Klazinga NS, de Keizer NF. Influences of definition ambiguity on hospital indicator scores: examples from The Netherlands. European Journal of Public Health. 2014: 24(1):73-8.

Anema HA, Kievit J, Fischer C, Steyerberg EW, Klazinga NS. Influences of hospital information systems, indicator data collection and computation on reported Dutch hospital performance indicator scores. BMC Health Services Research. 2013: 13(1):212.

Fischer C, Anema H, Klazinga N. The validity of indicators for assessing quality of care: A review of the European literature on the readmission rate. European Journal of Public Health. 2011: 1-8. (Received the editor's choice)

Klazinga N, Fischer C, ten Asbroek A. Health services research activities related to performance indicators and benchmarking in Europe. Journal of Health Services Research and Policy. 2011: 16 Suppl 2:38-47.

Fischer C, Brug J, Tak N, Yngve A, te Velde S. Differences in fruit and vegetable intake and their determinants among 11-year-old schoolchildren between 2003 and 2009. International Journal of Behavioral Nutrition and Physical Activity. 2011: 8:141.

Fischer C, Yıldırım M, Salmon J, Chinapaw M. Comparing different accelerometer cut points for sedentary time in children. Pediatrics exercise science. 2012: 24(2):220-8.

Reports

Kringos DS, Fischer C, Lingsma HF, Marang-van de Mheen PK, Steyerberg EW, Klazinga NS. Onderzoek naar de bruikbaarheid van informatie over heropnames en heroperaties uit het DBC Informatiesysteem (DIS) voor het genereren van kwaliteitsindicatoren.

Kringos DS, Anema HA, ten Asbroek AHA, Fischer C, Botje D, Kievit J, Steyerberg EW, Klazinga NS. Beperkt Zicht; Onderzoek naar de betrouwbaarheid, validiteit en bruikbaarheid van prestatie-indicatoren over de kwaliteit van de Nederlandse ziekenhuiszorg. Amsterdam: Nederlandse Federatie van Universitair Medische Centra (NFU). December 2012. ISBN: 978 90 9027307 3. 99 p.

Policy Briefs

HSR Europe (2011). *Policybrief-health-services-research.* Retrieved January 08, 2015, from http://www.nivel.nl/sites/default/files/bestanden/Policybrief-health-services-research.pdf.

PHD PORTFOLIO

Summary of PhD training and teaching activities

Name of PhD student: Claudia Fischer
Erasmus MC department: Public Health
PhD period: 2010 - 2015
Promotors: Prof.dr. E.W.Steyerberg
Prof.dr. N.S.Klazinga
Copromotor: Dr. H.F.Lingsma

	Year	Workload (ECTS)
1. PhD training		
General academic skills		
Workshop grant proposal	2014	0.3
Project and time management course	2012	1.4
Research skills		
Master of Public Health, Netherlands Institute for health Sciences (Nihes) Rotterdam, the Netherlands	2011-2013	70
Master Policy and Politics, institute of Health Policy & Management	2011	5
Seminars and workshops		
SMDM core course: introduction to medical decision analysis (Decision-Analytic Modeling)	06/2014	0.3
Seminars department of Public Health, Erasmus MC	2010-2015	3.6
Meetings clinical decision making, department of Public Health, Erasmus MC	2010-2015	1.8
Presentations at national and international conferences		
Presentations within Erasmus MC	2010-2015	5
National conferences		
Dutch Institute of Clinical Audit (DICA) congress, Amsterdam	06/2013	1
NCVG, Amsterdam	11/2010	1
International meetings and conferences		
Health is Wealth (Liverpool Health Partners), Liverpool	01/2015	0.9
ISQUA, Rio de Janeiro	10/2014	0.9
Young Expert Panel on Health, Microsoft Innovation Centre, institutional workshop at the WHO, Brussels	06/2014	0.3
SMDM, Antwerp	06/2014	0.9
Seminar 'Social Europe-Solidarity in health', European Parliament, Brussels	01/2014	0.3
International QUIRN meeting, Lisbon	05/2013	0.3
European Health Forum, Gastein	10/2012	0.9
Digital Futures workshop (DG INFSO) Brussels	04/2012	0.6
Young Researchers Meeting, FAHRE, Porto	03/2012	0.3
European Health Forum, Gastein	10/2011	0.9
SMDM, Oslo	10/2011	0.9
ISQUA, Hong Kong	06/2011	0.9
Academy Health Annual Research Meeting, Seattle	04/2011	0.9
EUPHA, Amsterdam	05/2010	0.6
EGEA, Brussels	05/2010	0.6
Health Services Research Working Conference, the Hague	04/2010	0.3

2. Teaching activies	2010-2015	
Lecturing		
Master thesis supervision of medical and health policy students, ErasmusMC	2013	1.4
Leading workgroups on quality of care, medical master students, Amsterdam Medical Centre, University of Amsterdam;	2012-2013	0.2
Supervising medical bachelor students on public health research project; ErasmusMC	2012-2013	1.2
Guest lecturer, study course "between health care process and health care policy", Quality indicators for medical/ medical informatics students at Amsterdam Medical Centre, University of Amsterdam;	2011-2012	0.1

ISQUA = International society for quality in health care
WHO = World Health Origination
SMDM = society for medical decision making
NCVG = nederlands congress for volksgezondheid
EUPHA = European public health association
FAHRE= Strengthening Euopean Food and Health Research