Measuring, comparing and improving clinical outcomes in gastrointestinal cancer surgery

Henneman, D.

CHAPTER 2

Ranking and rankability of hospital postoperative mortality rates in colorectal cancer surgery

D. Henneman
A. C. M. van Bommel
A. Snijders
H. S. Snijders
R. A. E. M. Tollenaar
M. W. J. M. Wouters
M. Fiocco

Annals of Surgery 2014 May;259(5):844-9

ABSTRACT

Objectives: To examine to what extent random variation and variation in casemix influence hospital rankings based on mortality rates; to determine the suitability of mortality for ranking hospitals in colorectal surgery.

Background: Comparing and ranking postoperative mortality rates between hospitals becomes increasingly popular. Differences in hospital case-mix, and chance variation related to caseload, may influence rankings. The suitability of mortality for rankings remains unclear.

Methods: Data were derived from the Dutch Surgical Colorectal Audit. Hospital rankings based on fixed (FE) and random effects (RE) logistic regression models, unadjusted and adjusted for casemix were compared with the percentile expected ranks (PCER; the chance that a hospital performs better than a random hospital). Rankability, measuring which part of variation between hospitals is not due to chance, was calculated.

Results: Some 25,591 patients undergoing colorectal resections in 92 hospitals were evaluated. Postoperative mortality rates ranged between 0 and 8.8%. Adjustment for casemix with an FE model caused large changes in rankings. A smaller additional effect on changes in rankings occurred after adjusting with an RE model, with lower volume hospitals moving towards the mean. PCERs ranged between 10% and 85%. Rankability was 38%, meaning that 62% of hospital variation in mortality was due to chance.

Conclusions: Hospital ranks changed after casemix adjustment and random effect models, compared to unadjusted analysis. A large proportion of hospital variation in mortality was due to chance. Caution should be warranted when interpreting hospital rankings

based on postoperative mortality. Percentiles of expected ranks may help to identify hospitals with exceptional performance.

INTRODUCTION

Colorectal cancer surgery is performed commonly, but colorectal resections remain associated with morbidity and mortality, accounting for 24% of all adverse events in general surgery[1]. Hence, complications and mortality are widely used outcomes in colorectal surgery[2-5]. At the same time, society focuses increasingly on effectiveness and efficiency in healthcare. Variations in hospital performances have become subject to research[6] and various quality improvement projects aim at reducing adverse event rates[7].

The Dutch Surgical Colorectal Audit (DSCA) is a nationwide continuous quality improvement program. One of its main focus points is reduction of postoperative mortality rates by providing feedback of results to participating hospitals, with the national average as a benchmark[8]. The Association of Surgeons of the Netherlands agreed on a process in which outcomes of the DSCA will become publicly available in a stepwise fashion throughout the years.

With outcomes made available for the public, explicit ranking of hospitals based on specific outcomes may be attempted to compare quality of care, as is rather popular in the lay press[9-11]. Postoperative mortality may be considered one of the most delicate outcomes, and unjustly stigmatizing a hospital as having a high mortality rate may have great impact on hospital reputation. It is therefore crucial that hospital comparisons, especially rankings, are based on sound methodology and should be reliable. After all, when information becomes public, allocation of reimbursements by insurers and certification by policy makers might be based on such rankings.

Two issues have to be addressed when comparing hospital performances. First, the occurrence of postoperative death may depend on the patient's age, preoperative condition and disease severity. There is an increasing body of evidence that case-mix of different hospitals

varies[12,13]. Second, chance variation may play an important role. For hospitals with a small number of cases, it is difficult to know whether extremely high or low mortality rates are due to chance or caused by actual differences in quality of care. Random Effect (RE) regression models can be fitted to account for the fact that part of the variation in outcomes between hospitals is due to chance[14-16].

Previous studies have examined the influence of random variation and differences in casemix on hospital variation and ranking in performance indicators for various types of treatments[13,17,18], including wound infections and reoperation rates in colorectal surgery. Only one study investigated the effect of adjustment for chance variation on 30-day mortality after colectomy[15], finding a large impact on rankings.

With this background, we aimed to determine to what extent random variation and differences in casemix between hospitals have an impact on hospital comparisons in mortality rates after colorectal cancer resections in the context of the DSCA; and to explore whether postoperative mortality is an appropriate outcome to be used for hospital rankings in colorectal cancer surgery.

METHODS

Data source

Data were derived from the Dutch Surgical Colorectal Audit (DSCA), a continuous national quality improvement project in which many variables concerning patient and disease-specific details, diagnostics, treatment, and outcomes are collected prospectively. Since part of the dataset of the DSCA was designed with the objective of performing casemix adjustment for postoperative mortality, variables were determined as risk factors for postoperative mortality at an early

stage of development of the dataset. These factors were based on existing evidence concerning potential risk factors for mortality and determined by an expert panel using a Delphi method[8,12].

The DSCA contains data registered by all 92 Dutch hospitals performing colorectal cancer surgery[8]. The dataset shows a high level of completeness on most items and a case-ascertainment of approximately 95% when compared to the Netherlands Cancer Registry [8,19]. All information concerning individual patients and hospitals are made anonymous, making it possible to compare hospitals without identifying them.

Primary outcome

The primary outcome was 30-day and/or in-hospital postoperative mortality (death within 30 days after the operation, or during the index admission).

Statistical methods

To assess hospital's performance with respect to mortality, patient- and treatment characteristics (casemix) were included in the logistic regression analysis. Both fixed effect (FE) and random effect (RE) models were investigated. The case-mix factors age, gender, American Society of Anaesthesiologists (ASA) score, Charlson comorbidity index, body mass index, TNM stage, preoperative conditions related to the tumor, tumor location, procedure, preoperative (chemo/radio)therapy, urgent operations, additional resections and multiple synchronous colorectal tumors were included in the models. Details concerning the use of relevant casemix factors have been described elsewhere [12,20].

Fixed effect model

The FE logistic regression model is a classical regression model in which hospitals were included as a categorical variable by considering the situations without and with adjustment for case-mix. From these two models the log odds of mortality, with and without adjustment for differences in casemix between hospitals, and related standard errors for each hospital were estimated. Results from the FE models are referred to as FE estimates in this article.

Random effect model

Random effect (RE) models were used to represent the different source of variation in observed hospital-specific mortality rate. These models were employed to evaluate to what extent hospital variation in postoperative mortality can be attributed to chance. The estimated log odds adjusted for casemix between hospitals were computed along with a model parameter that describes the between-hospital variance (also called heterogeneity). As for the FE model, results from the RE are denoted as RE estimates, also known as Empirical Bayes (EB) estimates. The Bayesian approach, as introduced by Laird and Lewis and Thomas et al., produces shrinkage estimates of individual hospital mortality rates towards the national average and produces a more stable estimator[21,22]. In hospitals with a small number of cases, shrinking is bigger. The confidence intervals produced by the Bayesian methods account for the multiple comparisons problem that arises when identifying hospitals with an exceptional outcome among all hospitals. The variation in hospital-specific mortality rates not due to small sample fluctuations or measureable differences in severity of casemix can be quantified. Previous studies have also looked at the existence of such variation[16,23].

Ranking and rankability

To account for the effect of chance variation on rankings, the expected rank (ER)[21] was used. The ER represents the probability that the performance of a specific centre is better than a randomly chosen hospital. The ER can be transformed in percentiles based on expected ranks (PCER) to scale them between 0% and 100%.

By fitting a RE model, an estimation of the variability between hospitals can be obtained while the FE model provides an estimation of the variance for each hospital as it has been described in the Fixed Effects section. These quantities can be compared to measure which part of the variation between the hospitals is due to true differences. This leads to the measure called rankability, which indicates which part of variation between hospitals is due to true difference, and which part is due to chance[24]. Rankability is computed by relating heterogeneity between hospitals to uncertainty between and within centres. From this definition it follows that rankability can be used to express how reliable the ranking procedure is.

All statistical analyses were performed using R version 2.14. (http://cran.r-project.org/).

RESULTS

Patients

A total of 25,591 patients that underwent colorectal cancer resections in 92 Dutch hospitals between January 1st, 2009 and December 31st, 2011 were evaluated. The average hospital case volume in the study period was 278 patients (standard deviation 125,2). Patient, tumor and treatment characteristics are displayed in table 1.

Table 1: patient-, tumor- and treatment characteristics.

Characteristic		**N**	**%**
Gender	Male	14072	55%
Age	Mean (standard deviation)	70	11
Body Mass Index	Kg/m^2, mean (standard deviation)	26	4,8
Charlson co-morbidity index	0	14189	55%
	1	5555	22%
	2	3419	13%
	3 or higher	2428	10%
ASA classification	I	5132	20%
	II	13968	55%
	III	5389	21%
	IV	481	1,90%
	V	15	0,10%
	Unknown	606	2%
Pathological TNM stage	X	943	4%
	I	5270	21%
	II	8472	33%
	III	7934	31%
	IV	2972	11%
Preoperative tumor conditions	Perforation	409	2%
	Obstruction	2507	10%
	Anaemia/blood loss	1389	5%
Location of tumor	Right hemicolon	8207	32%
	Left hemicolon	3021	12%
	Sigmoid colon	7104	28%
	Rectum	7259	28%
Preoperative treatment	Short course radiotherapy	3417	13%
	Chemoradiotherapy	2067	8%
	Other	564	2%
Procedure	Ileocaecal resection	267	1%
	Right hemicolectomy	8026	32%
	Transverse colectomy	567	2%

Table 1: patient-, tumor- and treatment characteristics. *Continued*

Characteristic		N	%
	Left hemicolectomy	1854	7%
	Sigmoid colectomy/low anterior resection	11092	44%
	Subtotal colectomy	400	2%
	Abdominoperineal resection	2240	9%
	Panproctocolectomy	245	1%
	Other	622	3%
Urgency of procedure	Urgent/emergency procedure	3840	15%
Additional resections	Locally advanced tumor	2448	10%
	Metastasectomy	821	3%

ASA= American Society of Anesthesiologists

Estimating hospital differences

The average mortality was 4.3% (range 0 - 8,8%). The individual hospital effects are displayed as a sequence of 95% confidence intervals (95% CI) (figures 1a-1b). The 95% CI's represent the estimated range in which the true effect size for each hospital lies with a likelihood of 95%. Similarly, in figure 1c, posterior probability intervals (estimated from the RE model) for the true hospital effects are shown. In these three figures, one specific hospital with no mortality cases is not shown due to the extreme effect sizes and corresponding confidence interval. Although not shown, this hospital was included in the analyses.

Figure 1a shows unadjusted log odds for mortality of all hospitals, ranked from the lowest to the highest mortality rate. In this unadjusted analysis, five hospitals had significantly lower (low outliers) and nine hospitals had significantly higher (high outliers) mortality rates than average (figure 1a). These are hospitals of which the 95% CI's do not cross 0 (horizontal black line). For illustrative purposes, ten arbitrarily chosen hospitals with unadjusted ranks 1, 11, 21, 31

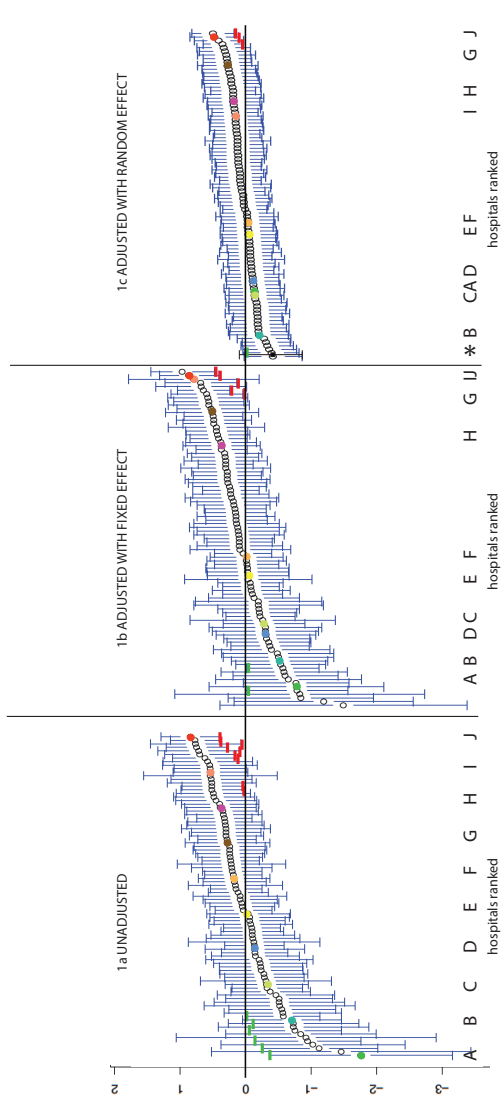

Figure 1: Hospital effect size in mortality and 95% confidence intervals; unadjusted analysis (1a), adjusted for casemix with a fixed effect model (1b), adjusted for casemix with a random effect model (1c). Hospitals A through J, in unadjusted analysis ranked 1st, 11th, 21st, 31st etc., are marked for illustration purposes.

* Hospital with no mortality cases, only displayed in figure c.

etc. up to 91 are marked with a specific colour and letter (A through J).

After adjustment for case-mix by fitting a FE model, hospitals changed ranks, as can be seen by tracking the positions of hospitals A-J in figure 1b compared to figure 1a. Only two hospitals remained low (significantly lower mortality than average) and five hospitals remained high outliers (figure 1b). Hospital A appeared to perform better than the average in the unadjusted analysis (figure 1a), but this proved to be partly due to a favourable casemix as the hospital was not a low outlier after fixed effect adjustment for casemix (figure 1b). After adjustment using a RE model with hospitals as a random effect, confidence intervals shrunk and only three high and one low outlier remained (figure 1c). Hospital J consistently remained a high outlier (significantly higher mortality than average) in all three models.

Ranking

Table 2 shows a quantification of change in ranks between unadjusted analysis and after adjustment for casemix with a FE model (A), and the extent of change in ranks between the FE and RE model analysis (B).

After adjustment for casemix in the FE model, 4 hospitals moved more than 30 places in the ranking. One of these hospitals moved from the lowest 20% (rank 79) to the middle 20% (rank 44), and two hospitals moved from the middle 20% to the bottom 20% (i.e. from the 51st to the 78th place, and from the 56th to the 85th place, respectively). One specific hospital moved from the 57th to the 15th rank. Four other hospitals changed between 21 and 30 ranks, and 23 hospitals changed between 11 and 20 ranks. Overall, three of the 18 'best' hospitals (top 20% in rank) moved out of the top 20%, and six of the 18 'worst' (lowest 20% in rank) moved out of the bottom 20%.

Table 2

	A Difference in rank between unadjusted and FE model	B Difference in rank between FE and RE model
>30 ranks higher	3	0
21-30 ranks higher	1	1
11-20 ranks higher	11	1
6-10 ranks higher	9	12
1-5 ranks higher	24	32
same rank	3	11
1-5 ranks lower	17	20
6-10 ranks lower	8	11
11-20 ranks lower	12	2
21-30 ranks lower	3	1
>30 ranks lower	1	1

Table 2: Change of ranks a) between results from unadjusted analysis and analysis with casemix adjustment in fixed effect (FE) model; b) between results from FE and random effect (RE) model.

By fitting a RE model, three top-20% hospitals moved out of this group, and three of the bottom-20% hospitals moved out of this group. One hospital moved down 45 ranks from the 43rd place to the 88th place; one changed from the 4th to the 29th place; and another one from the 89th to the 68th rank. Eventually three hospitals changed between 11 and 20 ranks, and for 11 hospitals the rank remained as in the FE model analysis.

For illustrative purposes, table 3 shows the respective ranks for postoperative mortality for hospitals A – J with the three different models as used in figure 1a-1c. In addition, the PCER for these specific hospitals are displayed. The PCER can be interpreted as the

Table 3

hospital	unadjusted rank	FE adjusted rank	RE adjusted rank	ER	Percentile ER
A	1	6	19	31.3	65,6%
B	11	13	7	24.1	73,8%
C	21	23	18	30.2	67,6%
D	31	20	22	33.1	64,2%
E	41	36	35	40.6	56,2%
F	51	41	38	41.5	55,4%
G	61	81	82	67.3	27,3%
H	71	72	72	60.2	35,2%
I	81	89	68	57.3	38,1%
J	91	90	90	83.1	9,9%

Table 3: ranks on postoperative mortality of hospitals A-J (ranked 1st, 11th, 21st, 31st and so on in unadjusted analysis) based on results from different models. FE= fixed effect; RE=random effect; ER= expected rank.

probability that a specific hospital has a better performance than a randomly selected hospital *i*.

Hospital A was highest in rank in unadjusted analysis, but moved to the 6th and 19th place in adjusted analysis and in the RE model analysis. Hospital A had PCER equal to 66%; this means that there is 66% probability that hospital A would have better mortality rates than a randomly selected hospital. Hospitals A, C, and D, in unadjusted analysis ranked 1, 21, and 31 had quite similar PCERs (64%-68%).

The distribution of the PCERs for all hospitals is shown in figure 2. As Figure 2 shows none of the hospitals could be classified as the worst or the best hospital (0% or 100%). Hospitals' percentiles of all hospitals ranged from 10% to 85%. PCER for hospital B was 74%, implying that there is 74% probability that hospital B performs better than a randomly selected hospital. On the other hand, hospital J with a PCER of 10% still had a 10% chance of not being the worst perform-

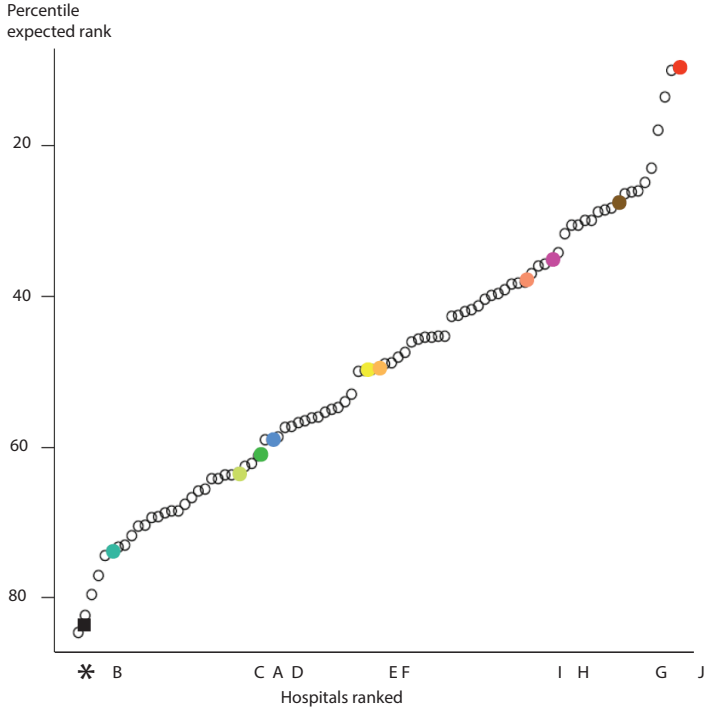

Figure 2: percentiles expected rank (PCER) on postoperative mortality for all hospitals. Hospitals A through J, in unadjusted analysis ranked 1st, 11th, 21st, 31st etc., are marked for illustration purposes.

* Hospital with no mortality cases.

ing hospital. The hospital with the highest PCER of all 92 hospitals was ranked 12th in unadjusted analysis (not shown in table 2).

Rankability

Rankability without and with adjustment for case-mix was equal to 44% and 38% respectively. Rankability can be interpreted as part of

the observed differences between centres not due to chance; while the rest is due to natural variations or chance. A value of rankability equal to 38% implies that 62% of the variation in postoperative mortality between hospitals was due to chance while 38% may be considered due to true differences in hospital performance.

DISCUSSION

Our study is the first to look at the rankability concept of postoperative mortality in colorectal surgery. In our study we estimated that 62% of the variation in postoperative mortality between hospitals was due to chance. We found that the differences in hospitals' ranks between unadjusted and FE adjusted analysis were considerable. To a lesser extent, rankings changed again after fitting RE models. Finally, we illustrated the differences between rankings on mortality rates based on FE and RE models and compared these rankings to result from the PCER measure, which estimated the probability that a hospital has a better mortality rate than a randomly chosen hospital.

Dimick et al., assessing mortality rates in 18,454 colectomy patients from 181 hospitals participating in the American National Surgical Quality Improvement Program (NSQIP), found large differences between hospital rankings based on FE and RE models[15]. In our study, we also detected effect of the RE model on rankings, although in our study, drastic relevant changes in rank (e.g. moving from the top or bottom 20% in comparison with ranking based on FE modelling) only occurred in 3 of the bottom-20% and 3 of the top-20% hospitals when comparing ranks based on the FE model with the RE model results. Dimick et al. found that the FE model ranking potentially misclassified 25% of the top-20% hospitals and 25% of the bottom-20%

hospitals when comparing with the RE model ranking. Our dataset might be more homogenous in terms of variation in hospital volumes and mortality rates between hospitals, which could be a possible explanation for the different findings. Both studies suggest the use of EB methodology when ranking hospital outcomes and this method should still be preferred over the more traditional FE models.

The novelty of our work concerns the use of rankability, as introduced by van Houwelingen et al.[24], for postoperative mortality in colorectal surgery. Rankability gives an idea of how a specific outcome accurately reflects hospitals' performance. Previous studies have looked at rankability in different outcomes. Van Dishoeck et al. studied seven performance indicators[18]. For the performance indicator 'unplanned reoperations after colorectal surgery', rankability was relatively high (71%), but for all the other remaining indicators rankability was lower (e.g., 58% for in-hospital mortality following myocardial infarction, 38% for pressure ulcer incidence). Unfortunately, the authors were unable to perform adjustments for casemix in this study. In a similar study, looking at surgical site infections for various procedures in 34 Dutch hospitals, the same authors found a rankability of only 8% for this outcome when all procedures were combined; however for colectomies, the rankability for surgical site infections was 80%. An exact rule that can be used to assess the reliability of a specific ranking does not exist. Lingsma et al. suggest that any ranking is meaningless when rankability is smaller than 50%, and that ranking can be used if rankability is bigger than 75% [25]. In our study, the rankability for postoperative mortality after colorectal surgery was rather low, 38%, suggesting that this measure is not appropriate for ranking hospitals. Should it be attempted, the PCER can be used. The PCER can be interpreted as an estimate of the probability that the mortality rate of a specific hospital is smaller than

the mortality rate of a randomly selected hospital. In this study the PCERs ranged from 10% to 85%.

The necessity of adjusting for casemix factors when comparing outcomes between hospitals is well established[13,26]. Kolfschoten at al. showed differences in patient characteristics between hospitals in the DSCA dataset, leading to different expected mortality rates[12]. Siregar et al. showed that hospital rankings on mortality following cardiac surgery are greatly influenced by adjustment for casemix with an RE model.

We came to the same conclusion for colorectal cancer surgery. We furthermore showed that outlier status (ie. hospitals having a significantly lower or higher mortality rate than average) changes after adjusting for case-mix factors.

Our study has some limitations. Firstly, although many casemix variables were available, there may have been unknown confounding variables not available in the dataset that may have influenced variation in outcomes between hospitals. However, colorectal cancer surgery experts constructed the DSCA dataset, and special attention was given to case-mix variables necessary for fair hospital comparisons[8,27].

One specific hospital had no observed cases of mortality. In unadjusted and adjusted FE analysis this hospital had an enormous (negative) effect size with extraordinarily large confidence intervals. For this reason, the estimates for this hospital were not displayed in the figures. This example illustrates the shortcomings of a FE model when there are no observed events in a hospital, which is not unlikely to happen in a sample with many hospitals. Arguably this hospital may perform well since there were no postoperative deaths in three years. Nonetheless, it is difficult to draw statistically valid conclusions from this observation in FE analysis. The EB methodol-

ogy overcomes this problem and an estimated hospital effect can be found in hospitals with no events. The PCER for this specific hospital was equal to 83% with an ER of 16.7.

In this study, we used pooled data from three registration years of registration. The longitudinal aspect of the data is beyond the scope of this manuscript, but in a future work it will be investigated whether results based on correlation across years might be used to make predictions about center effects.

There is an on-going debate whether outcome measures such as postoperative mortality adequately reflect quality of care. Some advocate using process measures (e.g. guideline adherence) in measuring quality of care, because these factors can be improved more concretely by hospitals with poor performance. However, what counts for patients are outcomes. Most probably, quality of care is best expressed as a combination of process and outcome measures, or even composite measures comprising both[28]. For this study, postoperatively mortality was chosen because it is well defined and may be considered quite delicate: unjustly stigmatizing a hospital as having high postoperative mortality may have a dramatic impact on hospital reputation and reimbursements. We found that 62% of variation in mortality between hospitals is due to chance, which implies that great caution should be used when interpreting hospital comparisons and rankings on this outcome. However, since this is an important outcome for patients, it seems worthwhile to continue measuring postoperative mortality rates. Another important reason to continue collecting and reporting postoperative mortality information is that evidence shows that feedback of surgical outcomes to physicians can lead to improvement[29,30]. Recently, in the UK, postoperative mortality data per surgeon has become publicly available on the Internet. Presumably, the influence of chance variation is

even greater in that situation, since the number of patients for each surgeon is rather small.

Measuring quality of care may have internal and external purposes. The DSCA is used as a system for benchmarking: surgical teams from the participating hospitals can compare results and improve in relation to the national average. This is an internal purpose. With an increasing demand for transparency of quality information, however, more information becomes public. Eventually, (risk-adjusted) outcome information will become public too. In this situation, third parties may compare outcomes between hospitals. Payers have limited resources and want to allocate them to the best performers. Ranking can be used in this context and therefore people should be aware of the reliability of such lists. The magnitude of differences between two hospitals is lost: one hospital is simply higher in rank than the other. Moreover, a hospital can move down on a ranking list as a result of another one moving up, even when performance remains the same. The advantage of the PCER measure is that it can be interpreted on its own, and it can be very useful in helping payers and patients to make decisions. The uncertainty concerning the outcome is included in the percentage ascribed to each hospital: the chance that the selected hospital has a better outcome than a randomly selected hospital. We suggest that when outcome information such as postoperative mortality becomes public, PCERs should be published with them. Reliable information on specific performance indicators may be extremely useful, if properly analysed and interpreted.

REFERENCES

1. Schilling PL, Dimick JB, Birkmeyer JD. Prioritizing quality improvement in general surgery. J Am Coll Surg 2008;207:698-704.
2. Morris EJ, Taylor EF, Thomas JD, et al. Thirty-day postoperative mortality after colorectal cancer surgery in England. Gut 2011;60:806-13.
3. Ghaferi AA, Birkmeyer JD, Dimick JB. Variation in hospital mortality associated with inpatient surgery. N Engl J Med 2009;361:1368-75.
4. Faiz O, Haji A, Bottle A, Clark SK, Darzi AW, Aylin P. Elective colonic surgery for cancer in the elderly: an investigation into postoperative mortality in English NHS hospitals between 1996 and 2007. Colorectal Dis 2011;13:779-85.
5. Cone MM, Herzig DO, Diggs BS, Rea JD, Hardiman KM, Lu KC. Effect of Surgical Approach on 30-Day Mortality and Morbidity After Elective Colectomy: a NSQIP Study. J Gastrointest Surg 2012;16:1212-7.
6. Elferink MA, Wouters MW, Krijnen P, et al. Disparities in quality of care for colon cancer between hospitals in the Netherlands. Eur J Surg Oncol 2010;36 Suppl 1:S64-73.
7. van Gijn W, Wouters MW, Peeters KC, van de Velde CJ. Nationwide outcome registrations to improve quality of care in rectal surgery. An initiative of the European Society of Surgical Oncology. J Surg Oncol 2009;99:491-6.
8. Van Leersum NJ, Snijders HS, Henneman D, et al. The Dutch Surgical Colorectal Audit. Eur J Surg Oncol 2013; 39(10):1063-70.
9. Wang W, Dillon B, Bouamra O. An analysis of hospital trauma care performance evaluation. J Trauma 2007;62:1215-22.
10. Green J, Wintfeld N. Report cards on cardiac surgeons. Assessing New York State's approach. N Engl J Med 1995;332:1229-32.
11. Chevreul K, Prigent A, Durand-Zaleski I, Steg PG. Does lay media ranking of hospitals reflect lower mortality in treating acute myocardial infarction? Archives of cardiovascular diseases 2012;105:489-98.
12. Kolfschoten NE, Marang van de Mheen PJ, Gooiker GA, et al. Variation in case-mix between hospitals treating colorectal cancer patients in the Netherlands. Eur J Surg Oncol 2011;37: 956-63.
13. Lingsma HF, Dippel DW, Hoeks SE, et al. Variation between hospitals in patient outcome after stroke is only partly explained by differences in quality of care: results from the Netherlands Stroke Survey. J Neurol Neurosurg Psychiatry 2008;79:888-94.
14. Morris C. Parametric empirical Bayes inference: theory and applications. Journal of the American Statistical Association 1983;78:47-65

15. Dimick JB, Ghaferi AA, Osborne NH, Ko CY, Hall BL. Reliability adjustment for reporting hospital outcomes with surgery. Ann Surg 2012;255:703-7.

16. Dimick JB, Staiger DO, Birkmeyer JD. Ranking hospitals on surgical mortality: the importance of reliability adjustment. Health Serv Res 2010;45:1614-29.

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

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

19. Dutch Institute for Clinical Auditing, Annual Reports 2012. http://www.clinicalaudit.nl. 2013. Accessed 01-08-2013

20. Kolfschoten NE, Wouters MW, Gooiker GA, et al. Nonelective Colon Cancer Resections in Elderly Patients: Results from the Dutch Surgical Colorectal Audit. Dig Surg 2012;29:412-9.

21. Laird NM, Louis TA. Empirical Bayes Ranking Methods. Journal of Educational Statistics 1989;14:29-46.

22. Thomas N, Longford NT, Rolph JE. Empirical Bayes methods for estimating hospital-specific mortality rates. Stat Med 1994;13:889-903.

23. Park RE, Brook RH, Kosecoff J, et al. Explaining variations in hospital death rates. Randomness, severity of illness, quality of care. JAMA 1990;264:484-90.

24. van Houwelingen HC, Brand R, Louis TA. Empirical Bayes methods for monitoring health care quality. https://www.msbi.nl/dnn/%20People/Houwelingen/Publications/tabid/158/Default.aspx. Accessed 20-12-2012.

25. Lingsma HF, Steyerberg EW, Eijkemans MJ, Dippel DW, Scholte Op Reimer WJ, Van Houwelingen HC. Comparing and ranking hospitals based on outcome: results from The Netherlands Stroke Survey. QJM 2010;103:99-108.

26. McArdle CS, Hole DJ. Outcome following surgery for colorectal cancer: analysis by hospital after adjustment for case-mix and deprivation. Br J Cancer 2002;86:331-5.

27. Snijders HS, Henneman D, van Leersum NL, et al. Anastomotic leakage as an outcome measure for quality of colorectal cancer surgery. BMJ Qual Saf 2013; 22(9):759-67.

28. Kolfschoten NE, Kievit J, Gooiker GA, et al. Focusing on desired outcomes of care after colon cancer resections; hospital variations in 'textbook outcome'. Eur J Surg Oncol 2013;39:156-63.

29. Khuri SF, Daley J, Henderson WG. The comparative assessment and improvement of quality of surgical care in the Department of Veterans Affairs. Arch Surg 2002;137:20-7.

30. Bridgewater B, Grayson AD, Brooks N, et al. Has the publication of cardiac surgery outcome data been associated with changes in practice in northwest England: an analysis of 25,730

patients undergoing CABG surgery under 30 surgeons over eight years. Heart 2007;93: 744-8.