Observational Study
Medicine®
OPEN

The prevalence of potentially preventable deaths in an acute care hospital

A retrospective cohort

Daniel M. Kobewka, MD, MSc[a,b,*], Carl van Walraven, MD, MSc[a,b,c,d], Monica Taljaard, PhD[b,c], Paul Ronksley, PhD[e], Alan J. Forster, MD, MSc[a,b,c,d,f]

Abstract

Studies estimate that 6% to 27% of deaths in hospitals might be prevented with higher quality care. These estimates may be inaccurate because they fail to account for the uncertainty associated with classifying preventability. The purpose of this study was to measure the prevalence of preventable deaths, accounting for the uncertainty in preventability ratings.

We created standardized structured case abstracts for all deaths at a multisite academic teaching hospital over a 3-month period. Each case abstract was evaluated independently by 4 reviewers who rated death preventability on a 100-point scale ranging from 0 ("Definitely not preventable") to 100 ("Definitely preventable"). Ratings were categorized into a 4-level ordinal scale and latent class analysis was used to measure the prevalence of each preventability class and estimate the probability that deaths in each class were preventable.

There were 480 deaths (3.4% of all admissions) during the study period. The latent class model (LCM) found that 91.6% (95% CI: 88.4–94.8%) of deaths were "nonpreventable" and 8.4% (5.2–11.6%) were "possibly preventable." "Possibly preventable" deaths could be identified with 90% certainty, but due to error in reviewer ratings, a "possibly preventable" death had a 50% probability of being receiving a rating of less than 25/100 by any single reviewer. Only 5 of 31 deaths classified as a "possibly preventable" (1.0% of all deaths) were judged to likely be alive in 3 months with perfect care.

After accounting for uncertainty associated with rating the preventability of hospital deaths, we found that 8.4% of deaths were deemed possibly preventable. There was only moderate probability that these deaths were truly preventable.

Abbreviations: AIC = Akaike information criterion, BIC = Bayesian information criterion, HSMR = Hospital Standardized Mortality Ratio, ICC = intraclass correlation, IRR = interrater reliability, LCA = latent class analysis.

Keywords: adverse events, patient safety, preventable deaths

1. Introduction

Nearly 25 years ago, a seminal study by Brennan et al[1] found that 4% of patients admitted to acute care hospitals experienced unintended injury or complications caused by medical care. Many subsequent studies have confirmed that healthcare causes harm and, in some cases, death.[2–14] Estimates of the prevalence of preventable deaths in hospital vary widely, ranging from 0.1% to 0.7% of all admissions and from 6% to 27% of all deaths.[2,3,7,15] While these statistics have motivated large-scale patient safety efforts, some have voiced concerns over the validity of these risk estimates because measuring death "preventability" can be unreliable and inaccurate.[16–19]

Almost all studies in this area gauge death preventability using medical record peer review, wherein a death is classified as preventable if the majority of reviewers judge the death to be caused by an error in healthcare delivery. This method does not account for the subjective nature of measuring death preventability and its associated error. Accounting for this error in reviewer preventability ratings is critical because poor interrater reliability (IRR) will artificially inflate the avoidable death prevalence, a bias described by Oppenheimer and Hayward that occurs when an imperfect test is used to measure an infrequent condition (Appendix 1, http://links.lww.com/MD/B575).[20,21] Despite the importance of this issue, only 1 study of preventable deaths accounted for the uncertainty in preventability ratings and recognized the effect of poor IRR.[7]

Since preventability cannot be directly measured, latent class analysis (LCA) could be used to classify death preventability based on classifications by multiple reviewers.[22,23] In general, LCA can be used when a variable cannot be measured directly, but only indirectly using 2 or more variables that are themselves measured with error. In this study, we use LCA to estimate preventable death prevalence based on multiple physician reviews.

Editor: Jesper Kers.

The authors have no conflicts of interest to disclose.

Supplemental Digital Content is available for this article.

[a] Department of Medicine—The Ottawa Hospital, [b] School of Epidemiology, Public Health and Preventive Medicine, University of Ottawa, [c] Clinical Epidemiology Program, Ottawa Hospital Research Institute, Ottawa, [d] Institute for Clinical Evaluative Sciences, Toronto, Ontario, [e] Department of Community Health Sciences, University of Calgary, Calgary, Alberta, [f] Performance Measurement, The Ottawa Hospital, Ottawa, Ontario, Canada.

[*] Correspondence: Daniel M. Kobewka, Division of General Internal Medicine, Department of Medicine, Ottawa Hospital, Civic Campus, 1053 Carling Avenue, Ottawa, ON K1Y 4E9, Canada (e-mail: dkobewka@gmail.com).

Medicine (2017) 96:8(e6162)

Received: 10 August 2016 / Received in final form: 6 January 2017 / Accepted: 27 January 2017

http://dx.doi.org/10.1097/MD.0000000000006162

2. Methods

2.1. Study design and population

This study was conducted at The Ottawa Hospital, a 1065-bed multisite academic teaching hospital with no pediatric department. We included all patients admitted to the hospital during a 3-month period in 2013. We excluded people who died before being admitted. The Ottawa Health Science Network Research Ethics Board approved this study.

2.2. Creation of case abstracts

Every death during the study period was recorded in our hospital's registration database. A physician and nurse independently reviewed each patient's hospital record and wrote a summary of the events leading to the patient's death, documenting instances where they believed care could have been improved. If there was any uncertainty regarding a case, the nurse reviewer interviewed the nurses and physicians who cared for the patient before death. The physician reviewer for each case was from the same specialty as the physician who was responsible for the patient's care at the time of their death. Finally a third reviewer, also a physician, reconciled the 2 summaries with further chart review to create a structured case abstract.

Each structured case abstract contained the patient's age, past medical history, history of presenting illness, physical examination findings, investigations, and a narrative of the course in hospital. Structured case abstracts described all aspects of care that reviewers thought could have been improved, but it did not contain any judgment statements about appropriateness of care.

2.3. Preventability ratings

We recruited physician reviewers by inviting all members of the general internal medicine division at our hospital and several other physicians directly. Physician reviewers were not compensated. Four physicians were randomly selected to review each case.

Each reviewer received standardized training materials that described the purpose of the study and general instructions about how to rate each case. Death preventability was defined as the probability that an act of commission or omission that would be agreed upon by a group of peers as standard of care would have allowed the patient to survive until hospital discharge. Reviewers assigned a preventability rating using a 100-point scale that was anchored at 0 (representing a completely nonpreventable death) and 100 (representing a completely preventable death). The reviewers were also asked: to describe what could have been done differently to prevent death; whether the change in care to prevent death needed to happen in hospital or before hospitalization; and whether the patient would likely be alive in 3 months had they received high quality care. The survey instrument was delivered using Microsoft Access 2007 (Appendix 2, http://links.lww.com/MD/B575).

2.4. Descriptive analysis

We also calculated the median risk of death in hospital using a validated risk score.[24] IRR was assessed using 2 measures of intraclass correlation (ICC) as described by Shrout and Fleiss[25]: first, the reliability of a single reviewer and second, the mean reliability of 4 randomly selected reviewers.

2.5. Latent class analysis

In general, LCA can be used when a variable cannot be measured directly, but only indirectly using 2 or more variables that are themselves measured with error.

LCA requires categorical input variables. An alternative method, called latent profile analysis, allows continuous input variables. Due to the highly skewed nature of observed ratings in our study (and to avoid making distributional assumptions about the observed ratings) LCA was considered more appropriate for our data. We therefore categorized the observed preventability ratings into 4 intervals having equal ranges (less than 25, 26–50, 51–75 and greater than 75) before analysis. Sensitivity analyses were carried out to determine the effect of alternate categorizations of preventability ratings.

The first step in LCA is to specify the number of underlying classes to be estimated. We specified a 2-class model because the dichotomy of "preventable" and "nonpreventable" has been used in all literature on preventable deaths. As a sensitivity analysis, we used Akaike information criterion (AIC) and Bayesian information criterion (BIC) to examine the fit of a 3-class model. In LCA, a statistical model is fit to the frequencies of observed ratings to obtain both the prevalence of each latent class as well as the conditional probability of each rating given membership in a specific latent class. By inspecting the pattern of conditional probabilities, a label can be assigned to each latent class. Interpretation of the latent classes is more challenging when there is a lack of homogeneity in the conditional probabilities associated with each latent class (i.e., when no specific response pattern is highly characteristic within that latent class). Clear interpretation of latent classes also requires separation between latent classes (i.e., sufficient differentiation in the patterns of ratings associated within each latent class). Our latent class model was estimated using maximum likelihood estimation as implemented in SAS 9.3 NC using PROC LCA (Version 1.3.2).[26]

2.6. Classification of preventable deaths

We used the latent class model to produce the Bayes' posterior probability of membership in each latent class for each individual case. Cases were assigned to the class for which they had the highest posterior probability of membership. We reported the model-based sensitivity and specificity of a reviewer to detect a death that was classified in the more preventable class.

As a summary measure of the overall degree of uncertainty in the model's classifications, we calculated the mean posterior probabilities for each class, which is the probability that individuals in each class are correctly classified, as well as the odds of correct classification (which is the improvement in the model's classifications beyond chance).[27] We used bootstrapping with 1000 replications to estimate confidence intervals for odds of correct classification, sensitivity, and specificity.[28,29]

3. Results

3.1. Study population

There were 14,287 hospitalizations during the study period and 480 deaths (3.4%). Patients who died were older, more likely to be male, and had a longer length of stay compared to those who survived (Table 1).

3.2. Reviewer population

Thirteen physician reviewers participated in the study with 12 certified in internal medicine (2 having further training in a

Table 1

Characteristics of all admitted patients and those who died during the study period.

Factor	Level	Vital status at discharge		
		Alive, N=13,787	Dead, N=480	Overall, N=14,267
Mean age (SD)		48.0 (27.2)	73.7 (16.1)	48.9 (27.3)
Gender	Female	7832 (56.8%)	228 (47.5%)	8069 (56.5%)
Median days in hospital (IQR)		3.0 (2.0–7.0)	7.0 (2.5–16.0)	3.0 (2.0–7.0)
Mean number of inpatient encounters in the previous 6 mo (SD)		0.27 (0.71)	0.63 (1.07)	0.28 (0.72)
Median adjusted mortality risk (IQR)*		0.01 (0.00–0.05)	0.28 (0.13–0.47)	0.01 (0.00–0.06)
Service	Obstetrics	1751 (12.7%)	0 (0.0%)	1749 (12.3%)
	General medicine	1507 (10.9%)	160 (33.3%)	1671 (11.7%)
	Nursery	1583 (11.5%)	3 (0.6%)	1584 (11.1%)
	Cardiology	1173 (8.5%)	48 (10.0%)	1219 (8.5%)
	General surgery	994 (7.2%)	20 (4.2%)	1015 (7.1%)
	Orthopedics	984 (7.1%)	11 (2.1%)	995 (7.0%)
	Urology	494 (3.6%)	2 (0.4%)	495 (3.5%)
	Psychiatry	468 (3.4%)	1 (0.2%)	468 (3.3%)
	Neurosurgery	409 (3.0%)	10 (2.1%)	419 (2.9%)
	Gynecology	394 (2.9%)	0 (0.0%)	393 (2.8%)
	Others	4030 (29.2%)	225 (46.9%)	4259 (29.9%)

IQR=interquartile range, SD=standard deviation.
* Calculated using a validated risk score.[24]

medicine subspecialty) and 1 certified in anesthesiology. The median duration in practice of the reviewers was 5 years (range 1–22 years).

3.3. Preventability ratings

Each structured case abstract was reviewed in quadruplicate resulting in 1920 preventability ratings. The mean preventability rating was of 5.7/100 (standard deviation (SD) 16.4) while the mode was 0/100 (1538/1920 (80.1%)). Most ratings (1675/1920 (87.2%)) were less than 10 while few ratings (51/1920 (2.6%)) exceeded 50. The reliability of a single reviewer as measured by the ICC was poor at 0.14, but the mean reliability of 4 reviewers was higher at 0.68.

3.4. Preventable deaths

The latent class model found that 8.4% (95% CI: 5.2–11.6%) of deaths fell into a class that we labeled "possibly preventable deaths" the other 91.6% (95% CI: 88.4–94.8%) of death fell into a class we labeled "nonpreventable deaths." Figure 1 displays the probability of each preventability rating range as a function of the cases assigned latent class. Individuals in the "possibly preventable death" group had a moderate (50.0%) probability of being rated 0% to 25% preventable and low (12.4% and 12.6%) probabilities of being rated 51% to 75% or 76 to 100% preventable respectively. Individuals in the "nonpreventable death" group had a very high probability (97%) of being rated 0 to 25% preventable, a low probability (2%) of being rated 26 to 50% preventable and a very low (0.6%) probability of being rated more than 50% preventable. The model-based sensitivity and specificity to detect a "possibly preventable death" was 25.0% (95% CI: 0.0–32.1%) and 99.4% (95% CI: 54.6–99.7%) respectively. The odds of correct classification, which is the odds of the model assigning an individual to the correct class compared to random chance was 3.2 (95% CI: 1.7–6.3) for the "nonpreventable death" class and 95.8 (95% CI: 3.0–455.3) for the "possibly preventable death" class. For our population the positive predictive value was 79.3% and the negative predictive value was 93.5%.

The mean posterior probability of correct classification was 97.2% (SD 6.5%) for the "nonpreventable death" class and 90.0% (SD 11.6%) for the "possible preventable death" class indicating a greater homogeneity of ratings for the former. Sensitivity analysis using different categorizations of the preventability ratings resulted in lower mean posterior probabilities and had little effect on the prevalence or interpretation of the preventability classes (Appendix 3, http://links.lww.com/MD/B575). Sensitivity analysis found that the 2-class model fit better than a 3-class model based on AIC and BIC.

Thirty-one deaths (6.5%) had posterior probabilities exceeding 50% probability of membership in the "possibly preventable" class. The remaining 449 deaths (93.5%) were best classified as "nonpreventable." Table 2 contains the characteristics of the "possibly preventable" and "nonpreventable" deaths. Patients with "possibly preventable" deaths had longer lengths of stay, were more likely to be admitted to a surgical service, and had a notably lower mean adjusted risk of death in hospital.[24] Twenty-one "possibly preventable" deaths (67.7%) were judged by at least half of reviewers to be preventable in part by actions that could have been taken in hospital as opposed to actions that needed to be taken before hospital admission. Five people having a "possibly preventable" death (16.1%) were judged by at least half of reviewers to have likely been alive in 3-months had they received error free care. Only 4 patients having a "possibly preventable" death (12.9%) were judged to be preventable at least in part by actions taken in hospital and to likely be alive in 3 months had they received the highest quality medical care. Therefore, we found that a death possibly preventable through hospital activity involving a patient likely to be alive 3 months hence occurred only in 4 of 14,267 admissions (0.03%).

Descriptions of each case in the "possibly preventable" death group can be found in Appendix 4, http://links.lww.com/MD/B575. Actions that may have prevented death were diverse and included better communication between care settings, correct or appropriate medication administration, and avoidance of procedure errors.

4. Discussion

Physicians reviewed the medical records of 480 patients who died at a tertiary care academic teaching hospital in quadruplicate to rate the likelihood that each death could be prevented with error

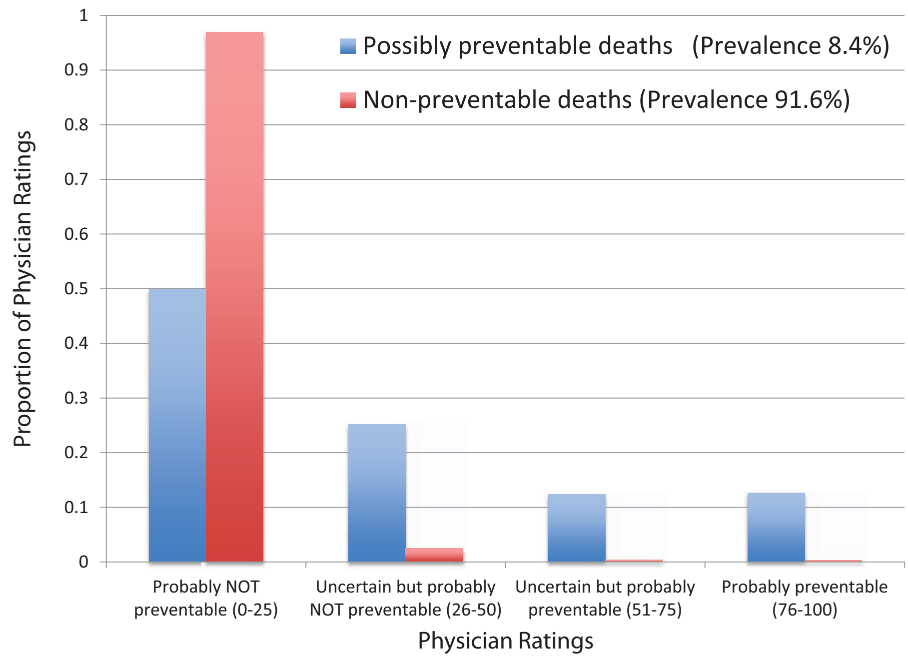

Figure 1. Results from latent class analysis showing physician ratings for possibly preventable and nonpreventable deaths. The latent class model has 2 categories: the first was labeled "possibly preventable deaths" and the second "nonpreventable deaths." This figure presents the distribution of physician preventability ratings (horizontal axis) for deaths in each the "possibly preventable deaths" (blue bars) and "nonpreventable deaths" (red bars) categories.

free medical care. We were able to clearly classify deaths into 2 groups but were unable to interpret the groups as a simple dichotomy of "preventable" and "nonpreventable" deaths. Deaths classified as nonpreventable were clearly not preventable but there was considerable uncertainty that deaths classified as "possibly preventable" were truly preventable. One in 12 deaths were possibly preventable, but these cases still had a 50% probability of receiving the lowest preventability rating (0–25%

Table 2

Characteristics of nonpreventable and possibly preventable deaths based on model posterior probabilities.

	Nonpreventable death, n=449	Possibly preventable death, n=31	Overall, n=480
Mean age (SD)	74.0 (16.1)	72.1 (15.6)	73.9 (16.0)
Female	214 (47.7)	17 (54.8)	231 (48.1)
Mean Elixhauser score (SD)	12.8 (9.0)	7.4 (6.6)	12.4 (9.0)
Number of inpatient encounters in the previous 6 mo			
0	282 (62.8)	25 (80.6)	307 (64.0)
1	94 (20.9)	5 (16.1)	99 (20.6)
≥2	73 (16.3)	1 (3.2)	74 (15.4)
Total length of stay, d			
1–3	137 (30.5)	8 (25.8)	145 (30.2)
4–7	97 (21.6)	7 (22.6)	104 (21.7)
8–14	82 (18.3)	5 (16.1)	87 (18.1)
≥15	133 (29.6)	11 (35.5)	144 (30.0)
Admitting hospital service			
General medicine	155 (34.5)	6 (19.4)	161 (33.5)
Intensive care	68 (15.1)	3 (9.7)	71 (14.8)
Cardiology	42 (9.4)	5 (16.1)	47 (9.8)
Medical oncology	38 (8.5)	1 (3.2)	39 (8.1)
Hematology	18 (4.0)	0 (0)	18 (3.8)
Neurology	16 (3.6)	0 (0)	16 (3.3)
Radiation oncology	15 (3.3)	0 (0)	15 (3.1)
Family medicine	14 (3.1)	0 (0)	14 (2.9)
General surgery	14 (3.1)	7 (22.6)	21 (4.4)
Orthopedics	8 (1.8)	4 (12.9)	12 (2.5)
Other	61 (13.6)	5 (16.1)	66 (13.8)
Patient has family physician	407 (90.6)	27 (87.1)	434 (90.4)
Palliative care consult	134 (29.8)	5 (16.1)	139 (29.0)
Median adjusted risk of death in hospital (IQR)[24]	0.29 (0.15–0.47)	0.07 (003–0.19)	0.28 (0.13–0.47)

Entries are frequencies and percentages unless otherwise specified.
IQR=interquartile range, SD=standard deviation.

preventable) from each reviewer. This is seen in the low sensitivity (25%) of each reviewer to detect a "possibly preventable" death. While we cannot be certain which deaths would have actually been prevented with better quality care, it would only be a fraction of those classified as possibly preventable. Another important finding is that most people with possibly preventable deaths had a very poor prognosis.

Our methodology is unique and therefore our results are not directly comparable to previous studies. Despite the differences, the prevalence of possibly preventable deaths in our study is similar to recent estimates of preventable deaths from the UK and the Netherlands.[11,13] We are only aware of one other study that corrected for uncertainty in reviewer ratings. This study, by Hayward and Hofer used a preventability rating of greater than 50%, to classify deaths as preventable[7]. This was different from the data driven approach we used in our study. Similar to our study however, Hayward and Hofer found that adjusting for IRR and then excluding patients who would likely not be alive in 3 months resulted in a drastically lower prevalence of preventable deaths.

4.1. Implications

Our study found that avoidable deaths were rare and difficult to identify with certainty. A key finding of our work is that there is uncertainty when classifying a death as preventable or not, this is true of previous studies of preventable deaths but is rarely explicitly acknowledged when results are interpreted. Our finding that preventability of death is uncertain is important because it suggests that previous studies of preventable deaths may overestimate prevalence by not accounting for poor IRR. The IRR in our study was similar to that reported by others and resulted in uncertainty that deaths in the "possibly preventable death" class were truly preventable.[30–32] From our results, we cannot be certain what proportion of deaths would truly be prevented with error-free care but we know it would only be a subset of the 8.4% that were "possibly preventable." This low number of preventable deaths calls into question the hypothesized association between quality of care and preventable deaths. This association is the basis of a common system wide measure of preventable deaths, the Hospital Standardized Mortality Ratio (HSMR).[33,34] In a modeling study, Girling et al found that the HSMR will be poorly predictive of preventable deaths if the true preventable death rate is less than 15%.[33,35] We found that the true preventable death rate is far less than 15%. If our results are generalizable to other hospitals then HSMR is not a useful metric of quality of care. On the other hand HSMR may be useful to detect hospitals that have very elevated rates of preventable deaths.

4.2. Strengths and limitations

Our work is unique because we used a probabilistic model to estimate the prevalence of preventable deaths and measure the uncertainty in classification. Other strengths of our study include the large sample size, multiple reviewers per case and the high certainty of classification.

There are several limitations to our study. First off it is a single centre study and therefore it is unknown if our results are generalizable. Despite this limitation our methods are generalizable and highlight the importance of measuring and correcting for IRR. Another limitation is that reviewers were prone to hindsight bias because they knew that the outcome for every case was death and therefore they may have been more likely to think that a particular error in care directly caused a death when in fact it did not. This bias is not unique to our study. However, hindsight bias would likely inflate the estimate of preventable deaths and therefore only strengthens our conclusion that truly preventable deaths are rare. As with other adverse events studies reviewers did not have autopsy data that may have ruled in or out some deaths as preventable. Another limitation in our work is the poor separation and homogeneity of the latent classes, meaning that there was disagreement among reviewer's ratings, resulting in uncertainty that deaths in the "possibly preventable death" class were truly preventable. While this inhibited us from obtaining an estimate of truly preventable deaths it was an interesting and important finding. Lastly, nearly all reviewers were internists, which may have introduced bias. We guarded against this by ensuring that a physician from the appropriate specialty was involved in the creation of each case abstract.

5. Conclusion

While medical errors certainly occur, it is rare for them to be the certain cause of death. Attempts to count the number of preventable deaths are a misguided approach for quality improvement because it is difficult to know for certain if a particular death was preventable and truly preventable deaths are very rare. It is likely more fruitful to focus on individual processes that are related to the root causes of harm, regardless of event severity. Less subjective outcomes, such as hospital acquired infections or medication administration errors are easier to measure and the measurement process itself can directly inform the path to improvement.

References

[1] Brennan TA, Leape LL, Laird NM, et al. Incidence of adverse events and negligence in hospitalized patients. Results of the Harvard Medical Practice Study I. N Engl J Med 1991;324:370–6.
[2] Baker GR, Norton PG, Flintoft V, et al. The Canadian Adverse Events Study: the incidence of adverse events among hospital patients in Canada. CMAJ 2004;170:1678–86.
[3] Dubois RW, Brook RH. Preventable deaths: who, how often, and why? Ann Intern Med 1988;109:582–9.
[4] Thomas EJ, Brennan TA. Incidence and types of preventable adverse events in elderly patients: population based review of medical records. BMJ 2000;320:741–4.
[5] Wilson RM, Runciman WB, Gibberd RW, et al. The Quality in Australian Health Care Study. Med J Aust 1995;163:458–71.
[6] Davis P, Lay-Yee R, Briant R, et al. Adverse events in New Zealand public hospitals II: preventability and clinical context. N Z Med J 2003;116:U624.
[7] Hayward RA, Hofer TP. Estimating hospital deaths due to medical errors: preventability is in the eye of the reviewer. JAMA 2001;286: 415–20.
[8] Schioler T, Lipczak H, Pedersen BL, et al. [Incidence of adverse events in hospitals. A retrospective study of medical records]. Ugeskr Laeger 2001;163:5370–8.
[9] Vincent C, Neale G, Woloshynowych M. Adverse events in British hospitals: preliminary retrospective record review. BMJ 2001;322:517–9.
[10] Briant R, Buchanan J, Lay-Yee R, et al. Representative case series from New Zealand public hospital admissions in 1998—III: adverse events and death. N Z Med J 2006;119:U1909.
[11] Hogan H, Healey F, Neale G, et al. Preventable deaths due to problems in care in English acute hospitals: a retrospective case record review study. BMJ Qual Saf 2012;21:737–45.
[12] Soop M, Fryksmark U, Koster M, et al. The incidence of adverse events in Swedish hospitals: a retrospective medical record review study. Int J Qual Health Care 2009;21:285–91.
[13] Zegers M, de Bruijne MC, Wagner C, et al. Adverse events and potentially preventable deaths in Dutch hospitals: results of a retrospective patient record review study. Qual Saf Hhealth Care 2009;18:297–302.

[14] Michel P, Quenon JL, de Sarasqueta AM, et al. Comparison of three methods for estimating rates of adverse events and rates of preventable adverse events in acute care hospitals. BMJ 2004;328:199.
[15] Thomas EJ, Studdert DM, Burstin HR, et al. Incidence and types of adverse events and negligent care in Utah and Colorado. Med Care 2000;38:261–71.
[16] Forster AJ, Taljaard M, Bennett C, et al. Reliability of the peer-review process for adverse event rating. PLoS ONE 2012;7:e41239.
[17] Hayward RA, McMahon LFJr, Bernard AM. Evaluating the care of general medicine inpatients: how good is implicit review? Ann Intern Med 1993;118:550–6.
[18] Hofer TP, Bernstein SJ, DeMonner S, et al. Discussion between reviewers does not improve reliability of peer review of hospital quality. Med Care 2000;38:152–61.
[19] Zegers M, de Bruijne MC, Wagner C, et al. The inter-rater agreement of retrospective assessments of adverse events does not improve with two reviewers per patient record. J Clin Epidemiol 2010;63:94–102.
[20] Hayward RA, Heisler M, Adams J, et al. Overestimating outcome rates: statistical estimation when reliability is suboptimal. Health Serv Res 2007;42:1718–38.
[21] Oppenheimer L, Kher U. The impact of measurement error on the comparison of two treatments using a responder analysis. Stat Med 1999;18:2177–88.
[22] Formann AK, Kohlmann T. Latent class analysis in medical research. Stat Methods Med Res 1996;5:179–211.
[23] Collins LM and Lanza ST. Latent Class and Latent Transition Analysis. Hoboken, NJ: John Wiley & Sons; 2010:3–22.
[24] Escobar GJ, Greene JD, Scheirer P, et al. Risk-adjusting hospital inpatient mortality using automated inpatient, outpatient, and laboratory databases. Med Care 2008;46:232–9.
[25] Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull 1979;86:420–8.
[26] Gandhi PU, Szymonifka J, Motiwala SR, et al. Characterization and prediction of adverse events from intensive chronic heart failure management and effect on quality of life: results from the pro-B-type natriuretic peptide outpatient-tailored chronic heart failure therapy (PROTECT) study. J Card Fail 2015;21:9–15.
[27] Colins LM, Lanza ST. Latent Class and Latent Transition Analysis: With Applications in the Social, Behavioral and Health Sciences. Wiley, Hoboken New Jersey:2010.
[28] Cameron AC and Trivedi PK. Microeconomics Using Stata. College Station: Stata Press; 2009.
[29] How can I bootstrap in SAS. Available at: http://www.ats.ucla.edu/stat/sas/faq/bootstrap.htm. Accessed January 5, 2017, 2017.
[30] Goldman RL. The reliability of peer assessments of quality of care. JAMA 1992;267:958–60.
[31] Localio AR, Weaver SL, Landis JR, et al. Identifying adverse events caused by medical care: degree of physician agreement in a retrospective chart review. Ann Intern Med 1996;125:457–64.
[32] Thomas EJ, Lipsitz SR, Studdert DM, et al. The reliability of medical record review for estimating adverse event rates. Ann Intern Med 2002; 136:812–6.
[33] Shahian DM, Wolf RE, Iezzoni LI, et al. Variability in the measurement of hospital-wide mortality rates. N Engl J Med 2010;363:2530–9.
[34] Freemantle N, Richardson M, Wood J, et al. Can we update the Summary Hospital Mortality Index (SHMI) to make a useful measure of the quality of hospital care? An observational study. BMJ Open 2013;3.
[35] Girling AJ, Hofer TP, Wu J, et al. Case-mix adjusted hospital mortality is a poor proxy for preventable mortality: a modelling study. BMJ Qual Saf 2012;21:1052–6.