Original Scholarship

Ranking Hospitals Based on Preventable Hospital Death Rates: A Systematic Review With Implications for Both Direct Measurement and Indirect Measurement Through Standardized Mortality Rates

Semira Manaseki-Holland,[*] Richard J. Lilford,[†] An P. Te,[*] Yen-Fu Chen,[†] Keshav K. Gupta,[‡] Peter J. Chilton,[§] Timothy P. Hofer[‖]

[*]*University of Birmingham;* [†]*Warwick Medical School, University of Warwick;* [‡]*Sandwell and West Birmingham Hospitals, NHS Trust;* [§]*Warwick Business School, University of Warwick;* [‖]*Institute for Healthcare Policy & Innovation, University of Michigan*

Policy Points:

- The use of standardized mortality rates (SMRs) to profile hospitals presumes differences in preventable deaths, and at least one health system has suggested measuring preventable death rates of hospitals for comparison across time or in league tables. The influence of reliability on the optimal review number per case note or hospital for such a program has not been explored.

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/1468-0009.12375.

This article is protected by copyright. All rights reserved.

- Estimates for preventable death rates using implicit case note reviews by clinicians are quite low, suggesting that SMRs will not work well to rank hospitals, and any misspecification of the risk-adjustment models will produce a high risk of mislabelling outliers.
- Most studies achieve only fair to moderate reliability of the direct assessment of whether a death is preventable and thus it is likely that substantial numbers of reviews of deaths would be required to distinguish preventable from nonpreventable deaths as part of learning from individual cases, or for profiling hospitals.
- Furthermore, population- and hospital system–specific data on the variation in preventable deaths or adverse events across the hospitals and providers to be compared is required in order to design a measurement procedure and the number of reviews needed to distinguish between the patients or hospitals.

Context: There is interest in monitoring avoidable or preventable deaths measured directly or indirectly through standardized mortality rates (SMRs). While there have been numerous studies in recent years on adverse events, including preventable deaths, using implicit case note reviews by clinicians, no systemic reviews have aimed to summarize the estimates or the variations in methodologies used to derive these estimates. We reviewed studies that use implicit case note reviews to estimate the range of preventable death rates observed, the measurement characteristics of those estimates, and the measurement procedures used to generate them. We comment on the implications for monitoring SMRs and illustrate a way to calculate the number of reviews needed to establish a reliable estimate of the preventability of one death or the hospital preventable death rate.

Author Manuscript

This article is protected by copyright. All rights reserved.

Methods: We conducted a systematic review of the literature supplemented by a reanalysis of authors' previously published and unpublished data and measurement design calculations. We conducted initial searches in PubMed, MEDLINE (OvidSP) and Web of Knowledge in June 2012 and updated them in December 2017. Eligibility criteria included studies of hospitalwide admissions from general and acute medical wards where preventable death rates are provided or can be estimated and which can provide interobserver variations.

Results: Twenty-four studies were included from 1983 to 2017. Recent larger studies suggest consistently low rates of preventable deaths (interquartile range of 3.0-6.0% since 2008). Reliability of a single review for distinguishing between individual cases with regard to the preventability of death had a Kappa statistic of 0.27-0.50 for deaths and 0.24-0.76 for adverse events. A Kappa of 0.35 would require an average of 8-17 reviews of a single case to be precise enough to have confidence in high stakes decisions to change care procedures or impose sanctions within a hospital as a result. No study estimated the variation in preventable deaths across hospitals, although we were able to reanalyze one study to obtain an estimate. Based on this estimate, 200-300 total case note reviews per hospital could be required to reliably distinguish between hospitals.

The studies displayed considerable heterogeneity: 13/24 studies defined preventable death with a threshold of greater than or equal to four in a six-category Likert scale and 11/24 involved a two-stage screening process with nurses at the first stage and physicians at the second. Fifteen studies provided expert clinical review support for reviewer disagreements, advice, and quality control. A "generalist/internist" was the modal physician specialty for reviewers and they received 1-3 days of generic tools orientation and case note review practice. Methods did not consider the influence of human or environmental factors.

Author Manuscript

This article is protected by copyright. All rights reserved.

Conclusions: The literature provides limited information about the measurement characteristics of preventable deaths, suggesting that substantial numbers of reviews may be needed to create reliable estimates of preventable deaths at the individual or hospital level. Any operational program would require population-specific estimates of reliability. Preventable death rates are low, which is likely to make it difficult to use SMRs based on all deaths to validly profile hospitals. The literature provides little information to guide improvements in the measurement procedures.

Keywords: Avoidable, preventable, hospital deaths, hospital mortality, systematic review, variation.

Standardized mortality rates (SMRs) for hospitals are currently used as an indicator of institutional quality and to compare hospitals in order to identify outliers.[1] The rationale for their use is that they are a proxy for excess or preventable deaths, but there are compelling arguments that any signal (preventable death) will be obscured by the noise (all other unavoidable deaths).[2,3] Some policymakers are considering using direct measurements of preventable mortality, rather than trying to infer it indirectly from SMRs, as with the summary hospital-level mortality indicator (SHMI) used in the NHS in England.[4-7] For example, the NHS in England has instituted a system of mandatory physician retrospective case record review (RCRR) of deaths in hospitals in order to establish (and publish) the number of preventable deaths for local Trust use and to learn from mistakes.[8,9] A direct measurement of preventable death is also an obvious way to validate the widespread use of SMRs to measure the quality of care delivered to people prior to their death.

However, preventable death, as well as preventable adverse events (AEs) more broadly, can only be directly measured by the judgement of expert clinical observers who retrospectively review case notes. Although no systematic review has been done for preventable deaths, such judgement-based assessments have generally reported low reliability, meaning that they lack consistency across

Author Manuscript

This article is protected by copyright. All rights reserved.

repeated reviews. Thus, current and future policy and research agendas that propose measuring any preventable AEs, and specifically preventable mortality, should push us to define, and if possible, improve the measurement characteristics of those estimates. Only then can we use case note review measurements in research to validate SMRs, to design operational systems for learning from AEs within hospitals, and to compare preventable deaths between hospitals, possibly augmenting or even replacing comparisons by means of SMRs.

To this end, we conducted a systematic review firstly to summarize data from existing studies reporting avoidable deaths and the measurement characteristics of those estimates and applied these in order to determine the number of reviews needed to establish a reliable preventable death estimate at the individual or hospital level. Secondly, we summarize the heterogeneity between the measurement procedures used in these studies, including reviewer characteristics, selection, and training factors to assess whether there are potential opportunities to improve the reliability of the measurement procedure. This is the first review of methods to measure preventable mortality rates.

Methods

Literature Search

We conducted an initial search in PubMed and ISI Web of Knowledge in 2010. We updated and supplemented this in June 2012 and December 2017 with a broader search in MEDLINE (OvidSP), incorporating a wider range of terms covering preventability and errors, deaths and AEs, hospitals, and case note reviews (Online Appendix 1). After our last search and before finalizing this manuscript, we were made aware of two studies that met our inclusion criteria.[10-12] These studies are included in our review to ensure that our findings remain up-to-date. Reference lists of included studies were also hand searched to find additional articles.

Author Manuscript

This article is protected by copyright. All rights reserved.

Study Selection

The inclusion criteria were studies which (a) evaluated the preventability of hospital deaths (deaths primarily from general and acute medical wards) or preventable AEs contributing to death from a hospital-wide sample or primarily from general and acute medical wards; (b) provided a quantitative estimate of preventability of death or allowed this to be calculated; and (c) incorporated retrospective case record review that elicits the reviewer's own expert judgment, in reaching the conclusion about preventability. Only articles published in English were considered. Two reviewers (Gupta, Chilton, or Te) independently examined titles and abstracts retrieved from literature searches and selected studies for inclusion. Disagreements were resolved by consensus after retrieval of full-text articles and further discussions with a third reviewer (Chen). The review protocol was not submitted to PROSPERO as the review process was initiated before the establishment of PROSPERO.

Data Extraction and Synthesis of Evidence

Two reviewers (Gupta, Chen, Chilton, or Te) extracted data from the selected studies, including all data tabulated in Tables 1-3. The characteristics and findings of included studies were tabulated and summarized in a narrative form. We did not plan to pool results across studies given the underlying differences in settings and methods between the studies. Where data were missing, we wrote to the study authors and obtained details.

Number of Reviewers Required for a Reliable Measurement

Reliability describes the consistency of measurement and can be used to quantify the ability to distinguish between the objects of measurement. Reliability ranges from zero to one and increases

Author Manuscript

This article is protected by copyright. All rights reserved.

with a measurement procedure that makes multiple independent measurements and averages them. Most reports of the reliability of case note review give a number that describes the ability of a single review of any one case note to distinguish between a preventable and a non-preventable death. In Online Appendix 2, we describe one method that makes use of equations that allow you to calculate how reliability improves as the number of measurements is increased.

These commonly reported reliability estimates, which describe the ability to distinguish between case notes of patients who died, can quantify the confidence with which one can act on the presumption that a specific avoidable death had occurred, such as by investing in a root cause analysis to establish proximate causes, or possibly for establishing legal liability or determining compensation *for an individual case*. However, such reliability estimates tell you nothing about determining the performance of *different providers*, such as different hospitals. A key determinant of reliability in any measurement is the variation across the things you want to distinguish between; thus, to distinguish between hospitals requires an estimate of the variation of preventable death rates *across* hospitals.

No study was found to have published an estimate of this quantity despite its critical relevance to any policymaking with respect to preventable deaths. We were able to reanalyze data from one study of 22 hospitals to produce the variance estimates required to make a provisional "best available" calculation of the optimal number of reviews per case and per hospital required to produce a reliable estimate of the hospital preventable death rate (see Online Appendix 2).[13] Only one other study had quantified hospital variation for a more global measurement of preventable AEs that included deaths, and they reported a hospital variance estimate similar in magnitude to the one we estimated.[14]

Author Manuscript

This article is protected by copyright. All rights reserved.

Results

Article Retrieval and Inclusion

Our electronic searches yielded 663 records after duplicates were removed (Figure 1). A citation search of included studies identified six additional articles. In all, 37 articles (representing 23 studies) were included.[10-46] The characteristics of included studies are shown in Table 1. The study selection process and reasons for exclusion are summarized in Online Appendix 3. We were unable to find all the elements we required in the 37 published articles for any of the 23 studies. We wrote to the authors of these studies for more detail and of these, 14 of the authors responded.

{{Figure 1}}

{{TABLE 1}}

Twelve studies[10-13,21,26-33,36,38] focused the reviewers on an assessment of whether a death was preventable. Eleven studies[14-20,22-25,34,35,37,39-46] aimed primarily to identify and evaluate whether AEs were preventable. These AEs could include or accompany the death of a patient. All but two studies were in high-income countries and conducted between 1984 and 2015. They involved a median of 20 hospitals (IQR = 23) and 230 deaths reviewed (range 10 to 7194).

Methods for Assessing Preventable Deaths and Preventable Adverse Events Contributing to Deaths

The majority of the published studies did not present enough details to obtain the information required for this review and unpublished data were obtained by author communications. Through writing to the authors, we obtained additional data on 14 of the 23 studies. These are summarized in Table 2 and Online Appendices 4 and 5.

This article is protected by copyright. All rights reserved.

Author Manuscript

{{TABLE 2}}

Tools and stages of review. A plurality of the studies (9/23) followed the method of the Harvard Medical Practice Study,[22] which in turn was based on an approach called structured implicit physician review developed by the RAND Corporation in the 1980s.[47] This measurement procedure includes an initial screening of patient notes to identify cases in which it is more likely that an adverse event might have occurred. The other studies provided a varied amount information on methodology and therefore we wrote to authors for details. These are summarized in Table 2 and Online Appendices 4 and 5.

In structured implicit case note review the structured component guides the reviewer systematically and more or less temporally through the hospital admission, asking them to focus and rate specific elements of the patient's care in sequence before making an overall judgment about the quality of care.[48] The "implicit" component is inherent in the summary judgements produced by the reviewer about the case, as well as the exercise of professional situational judgement in deciding whether deviations from ideal processes represent an error or are appropriate in the clinical context. This can be contrasted with generating a score based on a checklist where the use of any judgement is much more restricted. A non-structured implicit review has been found to be less reliable in estimating hospital quality of care, presumably owing to the less standardized approach for navigating a record and building up to an overall rating.[49,50]

In our sample, most studies used a kind of structured implicit (or criterion-based implicit) review pro forma. Although the details of the structured component varied, in all cases adopting structured implicit review, the "structured" component required the reviewer to review and make quality judgements over phases of care (such as diagnostic or treatment phases). The reviewer was

Author Manuscript

This article is protected by copyright. All rights reserved.

often asked to write explicit comments about areas of concern (as free-flow text) for each phase, and finally to score quality for each phase of care.

The decision on preventability was made on a scale applying implicit judgement of the physician reviewers. The majority (15/23) of the studies used a six-category grading system (Likert scale) to classify the preventability of deaths and/or AEs.[10,14-20,22-25,32-35,37-46] The categories were inevitably collapsed into a binary outcome. Deaths (and/or AEs) that were considered to have more than a 50/50 chance of being preventable were considered preventable in most studies. Three studies[11-13,31] used a continuous scale (0-100) probability of preventability which was compared with the Likert scale; the 0-100 scale was found to have the same constructs and to impart comparable information to the Likert scale.[13]

Only five studies noted an attempt to anonymise the patient and hospital identifiers in case-notes (13, 23-25, 31-33) to prevent bias during reviews. No study blinded the reviewers to the outcome in these samples selected on the basis of death as the outcome.

Reviewer Selection and Training. In all studies, reviewers were external to the institutions from which case notes were derived to reduce internal institutional bias. For reviewer selection, seven studies did not have a first-stage screening process and deployed only physicians for these reviews.[13,14,18,19,26-29,32,33,36,44-46] Fifteen studies used two stages; a screening process that involved mainly nurses at the first stage and exclusively physician reviewers at the second stage.[10,14-26,31,34,35,38-46] Seven studies used an experienced or supervisor reviewer physician: in six studies they settled disagreements between the physician reviewers[14,16-19,22-25,37,43-46] and in one they were used for quality control purposes (see Table 2 and Online Appendices 4 and 5).[39-42]

The required reviewer experience (where recorded) varied widely across the studies in both nurses and physicians. For physicians, regular handling of case notes, a lengthy period of clinical work (ie, more than 5 years of clinical/reviewing experience), postgraduate education, and independent accreditation were used as criteria. For example, in the US studies, reviewers were board-

This article is protected by copyright. All rights reserved.

Author Manuscript

certified with a general preference for generalists/internists.[10,21,22,43] The UK studies used reviewers from specialties across general medicine and intensive care consultants.[13,32,33,38] Eight studies deployed general physicians,[11-19,22-25,37,43-46] and in seven of these a panel of specialists was available to advise individual reviewers when required.[11,12,14-19,22-25,37,43-46]

Various forms of reviewer training and support were provided. The training duration ranged from 1 to 3 days. Nurses and physicians had the same training in eight studies.[14-20,22-25,35,37,44-46] Eleven studies) were explicit about the exposure to case notes during the training.[10,13,14,18-28,36,37,44-46] Six studies did not disclose reviewer training information. Where enough details were available, training did not define preventability, but rather offered clinicians an opportunity to understand the aims, merits, and some caveats (eg, hind-sight bias[51,52]) of the case note review process, to familiarize them with the pro forma for data extraction, and to exchange views on approaches to difficult cases after practicing the review on one or more case notes.[13,16,17,19,20,22-24,26,31-34,46]

Estimated Preventable Mortality

The proportion of deaths judged to be preventable depends on the cut-off threshold used in the Likert scale. One study chose to estimate preventability at the lowest threshold, namely any probability that the death could have been prevented (eg, 2 or more out of 6),[23,24] whereas most used a threshold of more than 3 out of 6 or 3-4 out of 5. Preventable mortality rates as a proportion of *all admissions* were estimated between 0.07% and 4.62% (Table 3). Most reports were below 0.7%; the 2.27% reported in Brazil[34] and the 4.6% in the Dubois study[26-28] were exceptionally high. Preventability rates as a proportion of *all deaths* were estimated between 0.47% and 29%.[10-13,16-19,20-21,26,28-34,36,38,43-46,53] The studies focusing more broadly on AEs varied in approach when estimating preventable deaths. Their approaches ranged from asking reviewers to rate whether the identified AE contributed to death, to positing that a death is preventable if accompanied by a preventable AE, no matter how minor. The

Author Manuscript

This article is protected by copyright. All rights reserved.

estimates are more direct and consistent when considering the larger studies specifically focused on preventable deaths from only more recent years (2008 to 2017). These have a median preventable death rate of 3% with an interquartile range of 3.0-6.0% (range 0.47%-10%).

{{Table 3}}

The studies that evaluated preventability of any AE as a proportion of *all admissions* reported generally higher but widely variable figures, ranging from 1.02%[22] to 11.65%,[16,17] and preventable AEs as a proportion of *all AEs* ranging from 3.96%[22] to 70.1%.[37]

Interrater Reliability (Kappa Statistic)

The reliability of a single review assessing preventability is reported for 17 of the 23 studies.[10-13,15,18-22,26,28,30-33,35,37,39-46] Fifteen are reported as Cohen's Kappa, a statistic that was developed to measure the agreement between raters taking into consideration the agreement that occurs by chance,[55] although for these ordinal measures the intraclass correlation (reported for the remaining two) is comparable and would probably be preferred.[56,57] The reliability for assessing the preventability of death is reported for nine studies with a median reliability of 0.33 and an interquartile range of 0.27-0.45 (range 0.10-0.50). If limited to the reported reliabilities from five larger studies done in the past 10 years (that included a median of 1,080 deaths) the reliability has a median of 0.27 (range 0.10-0.49). A further eight studies reported the reliability for preventing an AE with a median of 0.36 and an interquartile range of 0.29-0.58 (range 0.21-0.76). No data were found on the effects of reviewer selection, characteristics, or training on the reliability of the judgement of preventability by the reviewers.

Author Manuscript

This article is protected by copyright. All rights reserved.

Calculating the Optimal Number of Reviews and Reviewers Per Case Note to Estimate Preventable Death Per Case Note and Per Hospital

The interquartile range of reliability reported for the ability of a single review to distinguish between cases with respect to whether death was preventable was 0.27 to 0.45. At a representative level of reliability of 0.35 for a single review, we can estimate that an average of 8 reviews per case note would be required to achieve a reliability of 0.8 when distinguishing between cases. Seventeen reviews per case would be required to achieve a reliability of 0.9, a level often recommended for testing with high-stakes consequences. If the reliability of a single review was as high as 0.5, then only 4 or 9 reviews per case note would be needed for a reliability of 0.8 or 0.9, respectively. However, any given operational program would have to determine the reliability of their measurement procedure in their population to figure out the number of cases needed to review.

About 200 to 300 total reviews per hospital would be required to reach a reliability of 0.8 for distinguishing between hospitals, based on the limited evidence available about the between-hospital variance and other components of variance (see Online Appendix 2 for the estimates used and methods to project sample size). However, given 300 reviews in total, better reliability is achieved with more reviews per patient and fewer patients overall. Holding the total number of reviews constant, increasing the number of reviews per case increases reliability (eg, 10 reviews per case for 30 cases) more than selecting more cases per hospital (150 cases per hospital with 2 reviews per case). A strategy of only one review per case would provide at best fair reliability (0.20-0.40) no matter how many total reviews were done per hospital. Figure 2 illustrates how the reliability changes as the numbers of reviews and reviewers per hospital vary.

{{Figure 2}}

Author Manuscript

This article is protected by copyright. All rights reserved.

It is important to emphasize that more extensive and particularly population-specific data about the sources of variability in the review procedure could substantially change the projected number of reviews needed in either direction. In general, more heterogeneity across hospitals, more consistent reviewers, evaluating change over time within hospital, and a focus on relative as opposed to absolute probability of preventable death would result in a more modest and feasible number of reviews needed to produce a reliable estimate.

Discussion

We set out to review the literature on measuring preventable deaths and to determine if it would allow us to project how many reviews and reviewers would be required for hospitals to learn lessons from reviewing preventable deaths and for a hospital system to profile hospitals based on their preventable death rates. Secondarily, we looked at whether the literature contained any information on how the reliability of physician retrospective case record review to identify preventable deaths could be improved by refining the measurement procedure. To this end we conducted a review of studies of preventable hospital deaths published from 1980 to 2015.

The first important finding is that the preventability of death was consistently low in the reviewed studies and remarkably consistent across the more recent large studies. After our review was completed ,one additional study from Norway of 1,000 deaths was published online ahead of print, reporting a preventable death rate of 4.2%, consistent with the interquartile range of 3-6% from the larger studies of the past decade that we describe (reliability was not estimated).[72] While some studies did vary the probability thresholds and Likert scale anchors for defining preventability as described earlier, most studies used a similar operational definition of more than a 50/50 chance on balance of probability for defining that a death was preventable. However, the difficulty of establishing how

Author Manuscript

This article is protected by copyright. All rights reserved.

representative the deaths reviewed were for many studies, as well as the heterogeneity of the measurement procedures employed made it impossible in our mind to develop a generalizable summary estimate.

Nevertheless, a low prevalence of preventable death should substantially heighten concern about using SMRs calculated from discharge data to profile hospitals. If 95% of deaths are nonpreventable, detection of outlier hospitals has an extremely low positive predictive value[3] and any misspecification of risk adjustment models will also necessarily introduce substantial bias in any judgement using SMRs about which hospitals have higher or lower rates of preventable deaths.

Another important finding is the lack of any published estimates in the literature of how much variation there is in preventable death rates across hospitals. Without this it is impossible to estimate the reliability for distinguishing between hospitals with respect to their preventable death rates or to design an operational program to do so. Using direct measurement, we estimated that as many as 300 or more total reviews could be required per hospital to distinguish between hospitals in a league table with high stakes relegation and promotion consequences. Additionally, holding the total number of reviews per hospital constant, the optimal number of cases per hospital and reviews per case would require trade-offs to ensure the maximum generalizability and precision.

Furthermore, recall that the explicit purpose of comparing SMRs is to identify differences in preventable or avoidable death rates for which the SMR is just a proxy. The only study to look at this found little correlation between SMRs and preventable deaths across hospitals.[71] If it is found more broadly that the rates are not correlated, or that the variation in SMRs across hospitals is substantially larger than the variation in preventable death rates as directly measured, it would add substantial support to the concerns voiced by a number of critics that SMRs are measuring something else, most likely unmeasured case-mix differences. Yet, profiling hospitals based on SMRs remains ubiquitous and in the United States is tied to significant and increasing financial risk to hospitals in the absence

This article is protected by copyright. All rights reserved.

Author Manuscript

of this critical piece of information that could further support or call into question the validity of SMRs.

The literature does provide more data about the reliability of a single measurement to distinguish individual cases with respect to whether a preventable death or preventable adverse event more generally occurred. This reliability estimate is relevant for quality reviews of sentinel cases by hospitals to learn from possible mistakes or for reviews by licensing boards or for cases subject to litigation. It is clear that high reliability is desirable before possible sanctions or major changes in work flows or procedures are contemplated on the basis of a judgement that a preventable death has occurred. For a typical reliability of 0.35 from the fairly wide range observed, between 8 to 17 reviewers could be required to reliably distinguish between patients with respect to whether a preventable death occurred. This is far larger than is commonly used for credentialing, legal cases, and sentinel case and root cause analysis reviews.

However, providing these specific calculations as examples should not obscure the more important point that different measurement questions and different patient and hospital populations will each require their own estimates of reliability. These reliability estimates can then in turn be used to develop question and population-specific calculations of the number of reviewers and reviews per record required so that an estimate with the required precision can be obtained. The numbers may vary substantially based on the setting and question.

We also summarise variation in the measurement procedures across studies (Online Appendix 5). We provide previously unpublished and summary data about many aspects of the procedures used as it was often not reported in the published papers. While the assessment methods had areas in common across the studies, on the whole they were quite heterogeneous. We found no empirical assessment of how single vs two-stage assessments, pro forma tools, reviewer selection or training, reviewer characteristics, and environmental influences effect consistency of measurement. Formal reliability or generalizability studies to evaluate different aspects of training and measurement

Author Manuscript

This article is protected by copyright. All rights reserved.

procedures could be built into an operational program to facilitate improvements in the reliability of measurement. Details of these criteria and methodological issues as related to existing literature are discussed in Online Appendix 5.

Finally, It is worth reiterating that the structured implicit case note review method was originally designed to measure quality, not preventable death, and has a large literature describing its use for this purpose.[58] We should perhaps abandon attempts to measure the absolute proportion of deaths that are preventable as an impossible quest.[13] Physicians are not good at estimating prognostic survival probabilities much less the even more challenging counterfactual probabilities such as "what is the probability of survival *if an event had not occurred*," which raises concern about the validity of such estimates.[59-61] Rather, structured implicit review could be used to directly measure the quality of care in the period before a patient's death, in keeping with how these methods were originally designed when developed 30-50 years ago.[47,62-64] This might be particularly useful if it was successfully demonstrated that quality problems were more common in those who eventually died during a hospitalization than in randomly selected cases.

The systematic review component of this study has several limitations. Because of practical reasons, we excluded studies not published in English. We found a large variation in the reported preventable mortality, but with only a limited number of studies we are unable to confirm the exact source of the observed heterogeneity. We have focused on overall hospital mortality and acute general medicine cases in this review.

Conclusions

Based on available information, preventable deaths comprise a relatively small fraction of all deaths, raising concerns about the feasibility of using SMRs as a proxy for preventable deaths. Structured implicit review is a challenging measurement task and it is likely that relatively large numbers of

Author Manuscript

This article is protected by copyright. All rights reserved.

reviews are needed either to allow for learning from individual cases or to compare hospitals. Furthermore, there is a critical lack of any reported estimates of hospital variance in preventable death rates which is required to design systems in a responsible way that profile hospitals based on preventable death rates, whether measured directly or indirectly. There is little evidence on factors that affect the consistency of case note reviews other than reviewer experience and agreement between reviewers remain fair to moderate.

Any operational system assessing hospital quality around deaths will need to invest in a substantial ongoing effort to quantify the variation across hospitals and reviewers, although the cost of this would still be small relative to the cost of the operational system itself. It is also important to evaluate how the selection and training of the reviewers and measurement procedures can make the reliability more consistent (see Online Appendix 5 for an expanded discussion).[65] Attempting to measure preventable deaths on an absolute scale would require engagement with the behavioral science and cognitive psychology literature, pertinent to human and system-wide errors[66] in health care,[67] that best locate the bounded rationality of human decision-making,[68] and the biases that plague it.[69,70] However, whether measuring preventable deaths, or quality more generally as we would recommend, those who want to profile providers must recognize that no program can be designed to distinguish between providers without stable estimates of the amount of variation that exists across those providers.

Author Manuscript

This article is protected by copyright. All rights reserved.

REFERENCES

1. *The Dr Foster Hospital Guide 2009: How Safe is Your Hospital?* London: Dr Foster Intelligence; 2009.

2. Lilford R, Pronovost P. Using hospital mortality rates to judge hospital performance: a bad idea that just won't go away. *BMJ*. 2010;340:c2016.

3. Girling AJ, Hofer TP, Wu J, et al. Case-mix adjusted hospital mortality is a poor proxy for preventable mortality: a modelling study. *BMJ Quality Saf*. 2012;21(12):1052-1056.

4. Scott IA, Brand CA, Phelps GE, Barker AL, Cameron PA. Using hospital standardised mortality ratios to assess quality of care—proceed with extreme caution. *Med J Aust*. 2011;194(12):645-648.

5. Nicholl J. Case-mix adjustment in non-randomised observational evaluations: the constant risk fallacy. *J Epidemiol Comm Health*. 2007;61(11):1010-1013.

6. Lilford R, Edwards A, Girling A, et al. Inter-rater reliability of case-note audit: a systematic review. *J Health Serv Res Pol*. 2007;12(3):173-180.

7. Tender for Retrospective case record review programme - Lot 2 Tender Documentation for a National Programme to use retrospective record review to estimate the proportion of hospital deaths due to problems in care in English acute NHS trusts [press release]. 2015. [No longer available online but program description at https://www.rcplondon.ac.uk/projects/national-mortality-case-record-review-programme.]

8. National Quality Board, UK. National Guidance on Learning from Deaths: A Framework for NHS Trusts and NHS Foundation Trusts on Identifying, Reporting,

Author Manuscript

This article is protected by copyright. All rights reserved.

Investigating and Learning from Deaths in Care. March 2017. https://www.england.nhs.uk/wp-content/uploads/2017/03/nqb-national-guidance-learning-from-deaths.pdf. Accessed January 4, 2019.

9. The Royal College of Physicians. RCP hosts pioneering national Mortality Case Record Review Programme. February 15, 2016. https://www.rcplondon.ac.uk/news/rcp-hosts-pioneering-national-mortality-case-record-review-programme. Accessed January 4, 2019.

10. Hayward RA, McMahon Jr LF, Bernard AM. Evaluating the care of general medicine inpatients: how good is implicit review? *Ann Intern Med*. 1993;118(7):550-556.

11. Kobewka DM, Van Walraven C, Taljaard M, Ronksley P, Forster AJ. The prevalence of potentially preventable deaths in an acute care hospital: A retrospective cohort. Medicine (United States). 2017;96(8):e6162.

12. Kobewka DM, Van Walraven C, Taljaard M, Ronksley P, Forster AJ. The prevalence of potentially preventable deaths in an acute care hospital: a retrospective cohort [Appendices]. *Medicine (Baltimore)*. 2017;96(8):e6162.

13. Manaseki-Holland S, Lilford RJ, Bishop JR, et al. Reviewing deaths in British and US hospitals: a study of two scales for assessing preventability. *BMJ Qual Saf*. 2017;26:408-416.

14. Zegers M, de Bruijne MC, Spreeuwenberg P, Wagner C, van der Wal G, Groenewegen PP. Variation in the rates of adverse events between hospitals and hospital departments. *I J Qual Health Care*. 2011;23(2):126-133.

15. Aranaz-Andrés JM, Aibar-Remón C, Limón-Ramírez R, et al. Prevalence of adverse events in the hospitals of five Latin American countries: results of the 'Iberoamerican study of adverse events' (IBEAS). *BMJ Qual Saf*. 2011;20(12):1043-1051.

This article is protected by copyright. All rights reserved.

Author Manuscript

16. Aranaz-Andrés JM, Aibar-Remón C, Vitaller-Burillo J, et al. Impact and preventability of adverse events in Spanish public hospitals: results of the Spanish national study of adverse events (ENEAS). *I J Qual Health Care*. 2009;21(6):408-44.

17. Aranaz-Andrés JM, Aibar-Remón C, Vitaller-Murillo J, Ruiz-López P, Limón-Ramírez R, Terol-García E. Incidence of adverse events related to health care in Spain: results of the Spanish National Study of Adverse Events. *J Epidemiol Community Health*. 2008;62(12):1022-1029.

18. Baines RJ, Langelaan M, De Bruijne MC, et al. Changes in adverse event rates in hospitals over time: a longitudinal retrospective patient record review study. *BMJ Qual Saf*. 2013;22(4):290-298.

19. Baines RJ, Langelaan M, De Bruijne MC, Wagner C. Is researching adverse events in hospital deaths a good way to describe patient safety in hospitals: a retrospective patient record review study. *BMJ Open*. 2015;5(7):e007380.

20. Baker GR, Norton PG, Flintoft V, et al. The Canadian Adverse Events Study: the incidence of adverse events among hospital patients in Canada. *Can Medl Assoc J*. 2004;170(11):1678-1686.

21. Best WR, Cowper DC. The ratio of observed-to-expected mortality as a quality of care indicator in non-surgical VA patients. *Medical Care*. 1994;32(4):390-400.

22. 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 Medicine*. 1991;324(6):370-376.

Author Manuscript

This article is protected by copyright. All rights reserved.

23. Briant R, Buchanan J, Lay-Yee R, Davis P. Representative case series from New Zealand public hospital admissions in 1998 - III: Adverse events and death. *N Z Med J*. 2006;119(1231).

24. Davis P, Lay-Yee R, Briant R, Ali W, Scott A, Schug S. Adverse events in New Zealand public hospitals II: preventability and clinical context. *N Z Med J*. 2003;116(1183):U624.

25. Davis P, Lay-Yee R, Briant R, et al. Adverse events in New Zealand public hospitals: principal findings from a national survey. Wellington, New Zealand: Ministry of Health; 2001.

26. Dubois RW, Brook RH. Preventable deaths: who, how often, and why? *Ann Intern Med*. 1988;109(7):582-589.

27. Dubois RW, Rogers WH, Draper D, Brook RH. Author's reply: does hospital mortality predict quality? *N Engl J Med*. 1988;318(24):1622.

28. Dubois RW, Rogers WH, Moxley III JH, Draper D, Brook RH. Hospital inpatient mortality. *N Engl J Med*. 1987;317(26):1674-1680.

29. Flaatten H, Brattebø G, Alme B, et al. Adverse events and in-hospital mortality: an analysis of all deaths in a Norwegian health trust during 2011. *BMC Health Serv Res*. 2017;17(1):465.

30. Gupta M, Fuchs B, Cutilli C, et al. Preventable mortality: does the perspective matter when determining preventability? *J Surg Res*. 2013;184(1):54-60.

31. Hayward RA, Hofer TP. Estimating hospital deaths due to medical errors: preventability is in the eye of the reviewer. *JAMA*. 2001;286(4):415-420.

Author Manuscript

This article is protected by copyright. All rights reserved.

32. 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;21(9):737-745.

33. Hogan H, Zipfel R, Neuburger J, Hutchings A, Darzi A, Black N. Avoidability of hospital deaths and association with hospital-wide mortality ratios: Retrospective case record review and regression analysis. *BMJ*. 2015;351:h3239

34. Martins M, Travassos C, Mendes W, Pavão ALB. Hospital deaths and adverse events in Brazil. *BMC Health Serv Res*. 2011;11.

35. Michel P, Quenon JL, Djihoud A, Tricaud-Vialle S, De Sarasqueta AM. French national survey of inpatient adverse events prospectively assessed with ward staff. *Qual Saf Health Care*. 2007;16(5):369-377.

36. Roberts AP, Morrow G, Walkley M, et al. From research to practice: results of 7,300 mortality retrospective case record reviews in four acute hospitals in the North-East of England. *BMJ Open Qual*. 2017;6(2).

37. Soop M, Fryksmark U, Köster M, Haglund B. The incidence of adverse events in Swedish hospitals: A retrospective medical record review study. *Int J Qual Health Care*. 2009;21(4):285-291.

38. Sorinola OO, Weerasinghe C, Brown R. Preventable hospital mortality: learning from retrospective case record review. *JRSM Short Rep*. 2012;3(11):77.

39. Thomas EJ, Brennan TA. Incidence and types of preventable adverse events in elderly patients: Population based review of medical records. *BMJ*. 2000;320(7237):741-744.

40. Thomas EJ, Studdert DM, Brennan TA. The reliability of medical record review for estimating adverse event rates. *Ann Intern Med*. 2002;136(11):812-816.

Author Manuscript

This article is protected by copyright. All rights reserved.

41. Thomas EJ, Studdert DM, Burstin HR, et al. Incidence and types of adverse events and negligent care in Utah and Colorado. *Medical Care*. 2000;38(3):261-271.

42. Thomas EJ, Studdert DM, Newhouse JP, et al. Costs of medical injuries in Utah and Colorado. *Inquiry*. 1999;36(3):255-264.

43. Wilson McL R, Runciman WB, Gibberd RW, Harrison BT, Newby L, Hamilton JD. The quality in Australian health care study. *Med J Aust*. 1995;163(9):458-471.

44. Zegers M, Bruijne MC, Wagner C, Groenewegen PP, Waaijman R, Wal G. Design of a retrospective patient record study on the occurrence of adverse events among patients in Dutch hospitals. *BMC Health Serv Res*. 2007;7(1):1.

45. Zegers M, de Bruijne MC, de Keizer B, et al. The incidence, root-causes, and outcomes of adverse events in surgical units: Implication for potential prevention strategies. *Patient Saf Surg*. 2011;5:13.

46. 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 Health Care*. 2009;18(4):297-302.

47. Rubenstein LV, Kahn KL, Harrison E, Sherwood MJ, Rogers WH, Brook RH. *Structured Implicit Review of the Medical Record*. Santa Monica, CA: Rand Corporation; 1991.

48. Pearson ML, Lee JL, Chang BL, Elliott M, Kahn KL, Rubenstein LV. Structured implicit review: a new method for monitoring nursing care quality. *Medical Care*. 2000:1074-1091.

Author Manuscript

This article is protected by copyright. All rights reserved.

49. Gibbs J, Clark K, Khuri S, Henderson W, Hur K, Daley J. Validating risk-adjusted surgical outcomes: chart review of process of care. *Int J Qual Health Care*. 2001;13(3):187-196.

50. Rubin HR, Rogers WH, Kahn KL, Rubenstein LV, Brook RH. Watching the doctor-watchers. How well do peer review organization methods detect hospital care quality problems? *JAMA*. 1992;267(17):2349-2354.

51. Fischhoff B. Hindsight is not equal to foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance*. 1975;1(3):288.

52. Fischhoff B. Hindsight ≠ foresight: the effect of outcome knowledge on judgment under uncertainty. *Qual Saf Health Care*. 2003;12(4):304-311.

53. Brett AS, Roberts M, Ruderman AJ, et al. Does hospital mortality predict quality? *N Engl J Med*. 1988;318(24):1622-1624.

54. Hanskamp-Sebregts M, Zegers M, Vincent C, van Gurp PJ, de Vet HCW, Wollersheim H. Measurement of patient safety: a systematic review of the reliability and validity of adverse event detection with record review. *BMJ Open*. 2016;6:e011078.

55. Cohen J. A coefficient of agreement for nominal scales. *Educational and Psychosocial Measurement*. 1960;20:37-46.

56. Maclure M, Willett WC. Misinterpretation and misuse of the kappa statistic. *Am J Epidemiol*. 1987;126(2):161-169.

57. Chmura Kraemer H, Periyakoil VS, Noda A. Kappa coefficients in medical research. *Statistics in Medicine*. 2002;21(14):2109-2129.

Author Manuscript

This article is protected by copyright. All rights reserved.

58. Kahn KL, Rubenstein LV, Sherwood MJ, Brook RH. Structured Implicit Review for Physician Implicit Measurement of Quality of Care. Santa Monica, CA: RAND Corporation; 1989. http://www.rand.org/pubs/notes/N3016/. Accessed January 9, 2018.

59. Chow E, Harth T, Hruby G, Finkelstein J, Wu J, Danjoux C. How accurate are physicians' clinical predictions of survival and the available prognostic tools in estimating survival times in terminally ill cancer patients? A systematic review. *Clin Oncol (R Coll Radiol).* 2001;13(3):209-218.

60. Christakis NA, Lamont EB. Extent and determinants of error in doctors' prognoses in terminally ill patients: prospective cohort study. *BMJ*. 2000;320(7233):469-473.

61. Glare P, Virik K, Jones M, et al. A systematic review of physicians' survival predictions in terminally ill cancer patients. *BMJ*. 2003;327(7408):195.

62. Rochester region perinatal study. Medical review project, Empire State Medical, Scientific and Educational Foundation, Inc. *N Y State J Med*. 1967;67(9):1205-1210.

63. Richardson FM. Peer review of medical care. *Med Care*. 1972;10(1):29-39.

64. Sheps MC. Approaches to the quality of hospital care. *Public Health Rep* 1955;70(9):877-886.

65. Hogan H. The problem with preventable deaths. *BMJ Qual Saf*. 2016;25:320-323.

66. Reason J. Human Error. New York: Cambridge University Press; 1990.

67. Reason J. Beyond the organisational accident: the need for "error wisdom" on the frontline. *Qual Saf Health Care*. 2004;13(suppl 2):ii28-ii33.

68. Kahneman D. A perspective on judgment and choice: mapping bounded rationality. *Am Psychol*. 2003;58(9):697.

Author Manuscript

This article is protected by copyright. All rights reserved.

69. Saposnik G, Redelmeier D, Ruff CC, Tobler PN. Cognitive biases associated with medical decisions: a systematic review. *BMC Med Inform Decis Mak*. 2016;16(1):138.

70. Blumenthal-Barby J, Krieger H. Cognitive biases and heuristics in medical decision making: a critical review using a systematic search strategy. *Med Decis Making*. 2015;35(4):539-557.

71. Hogan H, Zipfel R, Neuburger J, Hutchings A, Darzi A, Black N. Avoidability of hospital deaths and association with hospital-wide mortality ratios: retrospective case record review and regression analysis. *BMJ*. 2015;351:h3239.

72. Rogne T, Nordseth T, Marhaug G, et al. Rate of avoidable deaths in a Norwegian hospital trust as judged by retrospective chart review. *BMJ Qual Saf*. 2018;pii:bmjqs-2018-008053. doi: 10.1136/bmjqs-2018-008053. [Epub ahead of print]

Funding/Support: PJC and RJL were supported by the National Institute for Health Research (NIHR) Collaboration for Leadership in Applied Health Research and Care (CLAHRC) West Midlands. This paper presents independent research and the views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR, or the Department of Health and Social Care.

Author Manuscript

This article is protected by copyright. All rights reserved.

Acknowledgements: We would like to politely thank the study authors (14/23 of whom we contacted) who provided further information for this systematic review. This has helped to improve the systematic review content and quality.

Conflict of Interest Disclosures: All authors completed the ICMJE Form for Disclosure of Potential Conflicts of Interest. No conflicts were reported.

Address Correspondence To: Timothy Hofer, VA Center for Practice Management and Outcomes Research, PO Box 130170, Ann Arbor, MI 48113 (email: thofer@umich.edu).

Supplementary Material

Additional supporting information may be found in the online version of this article at http://onlinelibrary.wiley.com/journal/10.1111/(ISSN)1468-0009:

Appendix 1. Search Strategies and Results

Appendix 2. Methods and Results for Estimating the Number of Reviews Needed

Appendix 3. List of Excluded Studies

Appendix 4. Details of Mortality Review Process by Study

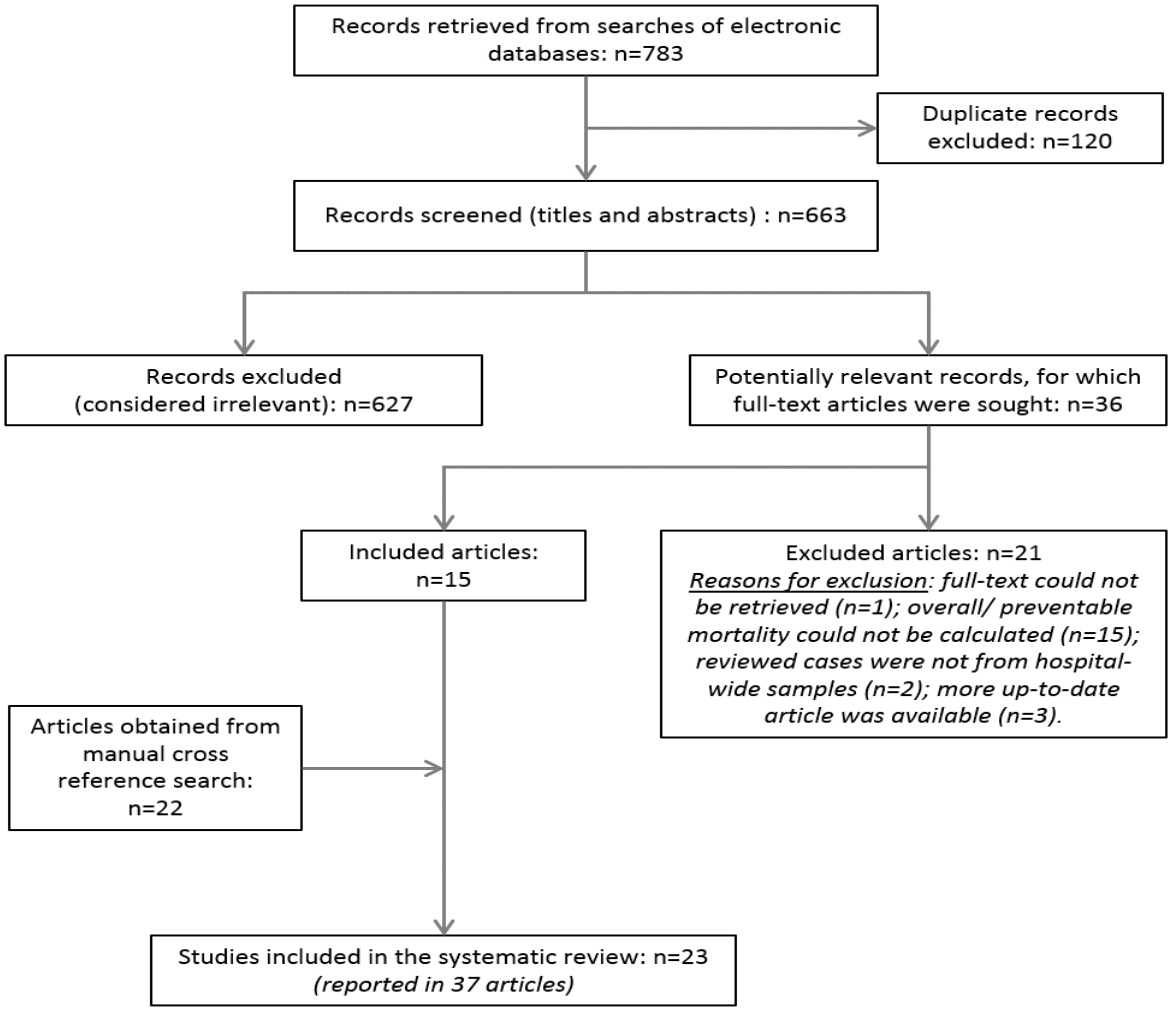
Appendix 5. Discussion of the Findings of Methods Used for Reviewing Case Notes In Our Included Studies

Author Manuscript

This article is protected by copyright. All rights reserved.

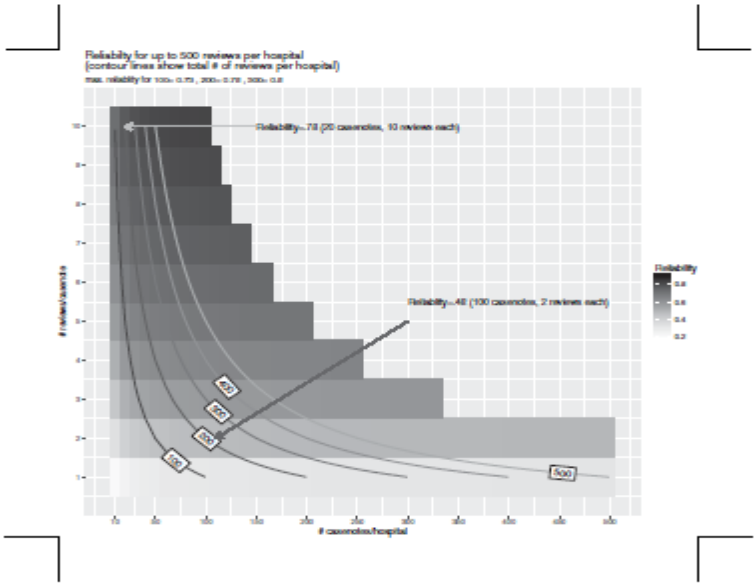
Figure Titles

Figure 1: Review Flow Diagram of Article Retrieval and Inclusion



This article is protected by copyright. All rights reserved.

Figure 2: Reliability for Up To 300 Reviews per Hospital



Filename: art/milq12375-fig-0002.eps
Height: 21.61371pc Width: 24.15639pc
Scale = 100

Table 1. Characteristics of Included Studies and Methods Used for Assessing the Preventability of Deaths or Adverse Events (AEs)

Author	Location; Date of Study	Target Group / Type of Hospital	Grading of Preventability[a]	Threshold for Defining a Preventable Case	Kappa (ICC) for Preventability	Interhospital Variance/ICC	Comments
Dubois and	United States;	12 / private	1-4[†]	≥3[b] Death as	κ =0.4, 0.3 and 0.2[c]	Not reported	• Hospital-wide

Author Ma

This article is protected by copyright. All rights reserved.

Brook, 1988[26-28]	1985	hospitals		"probably preventable"	preventability of **death** (182 charts, each reviewed by three physicians)		medical wards with conditions specific to cerebrovascular accident, pneumonia and myocardial infarction • Acute-care hospita

Author Manuscript

This article is protected by copyright. All rights reserved.

							ls that were considered outliers with higher and lower than expected mortality • Preventable mortality estimated from data

Author Manuscript

This article is protected by copyright. All rights reserved.

							• 14% of deaths (of all deaths) were prevent able
Brenna n et al, 1991[22]	New York, United States; 1984	51 private and nonfede ral acute care hospital s	1-6	≥ 4 negligence is more likely than not	κ =0.24 / preventabil ity of **AE** (based on duplicated review of 318 cases (2/51 hospitals)	Not reported	• Hospital-wide, excluding psychiatric patients • Nonfederal , acute-care hospitals • Preventabl e mortality estimated from data • Weighted

Author Manuscript

This article is protected by copyright. All rights reserved.

							figures based on events discovered during index hospitalization only • 13.6% of patients with AEs died
Haywar d et al, 1993[10]	United States; 1988-1990	1 teachin g hospital	1-6	≥ 5 better quality care could have prevented the death	κ = 0.5 Death preventabl e by better quality of care (based on dual reviews of 79 deaths)	N/A (Insufficien t denominato r)	• Hospital-wide medical wards with no single diagnostic-related group contributin g ≥ 5% of

Author Manuscript

This article is protected by copyright. All rights reserved.

							patient admissions • Acute-care university teaching hospital • 9% of patient deaths preventable
Best and Cowper, 1994[21]	United States; 1986	16 Veteran Affair Medical Centers	1-4	≥ 3 Somewhat likely that better management in the hospital might have prevented patient's death	κ = 0.33 "agreement = ≤ 2 positions on 9-point scale" (111 match-pairs from high and low	Not reported	• Veteran Affairs Medical Centers (Small, med/large and Psychiatric/long-term types) • 21.6% of

Author Manuscript

This article is protected by copyright. All rights reserved.

					mortality risk Veteran Affairs Medical Centers)		patient with better care management might have prevented death (or near the time of death)
Wilson et al, 1995[43]	New South Wales and South Australia; 1992	28 private and public acute care hospitals	1-6	≥ 4 "Preventability more likely than not, more than 50/50 but close call"	κ = 0.33 preventability of **AE** (based on duplicated review of 6,200 cases [all cases positive for screening criteria])	Not reported	• Hospital-wide excluding day-only admissions and admissions to psychiatric wards • Preventabl

Author Manuscript

This article is protected by copyright. All rights reserved.

							e AEs and PM estimated from data • 4.9% of patients with AEs died
Thomas et al, 1999; 2000a; 2000b; 2002.[39-42]	Utah and Colorado, United States; 1992	28 private and public hospitals	1-6	≥ 4 "More likely than not, > 50:50 but close call"	κ = 0.19 to 0.23 (95% CI, 0.05 to 0.37) preventability of **AE** (based on 3 independent reviews of 500 records)	Not reported	• Hospital-wide (13 in Utah and 15 in Colorado), excluding psychiatric and veteran hospitals and patients < 16 • Number of

Author Manuscript

This article is protected by copyright. All rights reserved.

							patients with AEs not specified, only total number of AEs • Based on events discovered during index hospitalization only • 6.6% of patients with AEs died
Haywar d and Hofer, 2001[31]	United States; 1994-1995	7 Veteran s Affairs	1-5[d]	≥ 4[d] "probably" - was the death	ICC = 0.34 preventabil ity of **Death**	N/A (Insufficien t denominato	• Hospital-wide, excluding data of

Author Manuscript

This article is protected by copyright. All rights reserved.

		hospitals		preventable by optimal care	(based on 383 review of 111 cases)	r)	patients receiving comfort care and nonveterans • Public hospitals • Patients with hospital-acquired laboratory abnormality over-sampled • Reviewed deceased patients only

Author Manuscript

This article is protected by copyright. All rights reserved.

Davis et al, 2001; 2003[24, 25] Briant et al, 2006[23]	New Zealand; 1998	13 public acute care hospitals	1-6	≥ 4 "Close call, > 50:50"	Not reported	Not reported	• Hospital-wide excluding specialist institutions • Public hospitals • Over all hospitals there were: 850 AEs; 315 avoidable AEs ≥ 4; 531 ≥ 2 • 4.5% of patients with AEs died • 6.1% of avoidable AEs;

Author Manuscript

This article is protected by copyright. All rights reserved.

							unclear concerning disability/death status
Baker et al, 2004[20]	Canada; 2000	20 public acute care hospitals	1-6	≥ 4 "Preventability more than likely (more than 50/50, but close call)"	κ = 0.69, (95% CI, 0.55-0.83) / preventability of **AE** (based on duplicated review of a random sample of 10% cases)	Not reported (Hospital size groupings preclude *de novo* calculation)	• Hospital-wide, excluding psychiatric and obstetric hospitals, day-only admission and patients < 18 • Acute-care hospitals • Weighted percentages to account

Author Manuscript

This article is protected by copyright. All rights reserved.

							for total charts per hospital and hospitals per type per province • 15.7% of patients with AEs died
Michel et al, 2007[35]	France; 2004	71 private and public hospitals	1-6	≥ 4 "more likely than not"	κ = 0.31 (95% CI, 0.05-0.57) / preventability of **AE** (based on 58 cases judged to have AE	Not reported	• Hospital-wide, excluding obstetric hospitals • Retrospective case-note review and 7-day

Author Manuscript

This article is protected by copyright. All rights reserved.

					by both reviewers)		observation with data collection across 294 wards • Patients with (preventable) AEs not noted • 8.2% of patients with AEs died
Soop et al, 2009[37]	Sweden; 2003-2004	28 public acute care hospitals	1-6	≥ 4 "more than 50% likelihood"	κ = 0.76 / preventability of **AE** (based on duplicated review of 642 cases [all cases	Not reported	• Hospital-wide, excluding psychiatric, rehabilitation, and palliative

Author Manuscript

This article is protected by copyright. All rights reserved.

					positive for screening criteria])		hospitals and day-only admission • Acute-care hospitals with high proportion of elderly patients; all deaths occurred in elderly/critically ill patients • Preventable mortality estimated from data • 4.1% of patients

Author Manuscript

This article is protected by copyright. All rights reserved.

							with AEs died
Aranaz-Andres et al, 2008; 2009[16,17]	Spain; 2005	24 public hospitals	1-6	≥ 4 "positive" - not defined	Not reported	Not reported	• Hospital-wide • Retrospective cohort study • Patients had 655 AEs; 278 preventable AEs (with at least moderate evidence) • Patients with preventable AEs estimated based on

Author Manuscript

This article is protected by copyright. All rights reserved.

							42.6% of AEs were preventable • Retrospective cohort study • 4.4% of patients with AEs died; Kappa was reported only for the identification of AEs between reviewers and "gold standards"
Aranaz-	Argentin	58	1-6	≥ 4	κ ranged	Not	• Hospital-

Author Manuscript

This article is protected by copyright. All rights reserved.

Andres et al, 2011[15]	a, Colombia, Costa Rica, Mexico and Peru; 2005	public hospitals		"positive" - not defined	from 0.27 to 0.74 between countries / preventability of **AE** (sample size not stated)	reported	wide • Retrospective case-note review and prospective data collection • Preventable mortality estimated from data • 5.8% of patients with AEs died
Martins et al, 2011[34]	Brazil; 2003	3 teaching hospitals	1-6	≥ 4 (wording not described)	Not reported	Not reported	• Hospital-wide, including obstetric wards. • 38% of

Author Manuscript

This article is protected by copyright. All rights reserved.

							patients with AEs died.
Hogan et al, 2012[32]	England; 2009	10 acute hospitals	1-6	≥ 4 "Probably preventable, more than 50/50 but close call"	κ = 0.49 (95% CI, 0.2-0.8) / preventability of **Death**, based on duplicated review of 250 cases (25% of sample)	"There were no significant differences between proportions of preventable deaths found at each hospital."[32(p740)]	• Hospital-wide, excluding obstetric and psychiatric wards, pediatric patients, and palliative care • 100 cases randomly selected from each acute hospital • Reviewed

Author Manuscript

This article is protected by copyright. All rights reserved.

							deceased patients only
Sorinola et al, 2012[38]	England. 2009	1 acute hospital	1-6	≥ 4 "Preventable death"	None given for preventability of death. Reported κ = 0.75 (from sample of 400 notes) only for "determination of a problem in care" (more equivalent to presence of an AE)	N/A (Insufficient denominator)	• Hospital-wide, excluding obstetric and psychiatric wards, pediatric patients, and palliative care • 400 death cases selected consecutively in 2009 • Preventable mortality

Author Manuscript

This article is protected by copyright. All rights reserved.

							estimated from data
Gupta et al, 2013[30]	United States; 2009-2012	1 acute hospital	1-5	≥ 4 "Possibly preventable"	κ = 0.10 Preventability of **Death** agreement between provider classification and a mortality review committee (15 cases only)	N/A (Insufficient denominator)	• Hospital-wide • 2,483 died, 1,683 had surveys completed • Preventable mortality estimate provided
Baines et al, 2013; 2015[18, 19]	The Netherlands; 2004 and 2008	33 acute hospitals	1-6	≥ 4 AE was found to be preventable when the care did	κ = 0.4 for preventability of adverse events[46]	Preventable AEs ICC = 3.7% (hospital-level)	• Hospitals including palliative care and excluding psychiatric

Author Manuscript

This article is protected by copyright. All rights reserved.

Zegers et al, 2007; 2009; 2011a; 2011b [14, 44-46]				not comply with existing professional standards and/or due to shortcomings of a healthcare practitioner, management or system			, obstetric, and pediatric patients • Hospitals were randomly selected on location • Reviewed patients discharged alive and deceased patients • Higher proportion of preventable AEs in deceased than

Author Manuscript

This article is protected by copyright. All rights reserved.

							patients discharged alive
Hogan et al, 2015 [33] [e]	England; 2012-2013	24 acute hospitals	1-6	≥ 4 Probably avoidable, more than 50-50	κ = 0.45 (95% CI, 0.24-0.66) / based on random sample of 486 avoidable **Death** cases (grade 4-6)	Not reported	• Hospitals, excluding obstetric, psychiatric, and pediatric patients • 100 cases randomly selected from each acute hospital • Reviewed only deceased patients
Manase ki-	England and	22 hospital	1-5	≥ 3 On the	κ = 0.27 (95% CI,	Not reported	• Hospitals with

Author Manuscript

This article is protected by copyright. All rights reserved.

Holland et al, 2016[13] [d,f]	Wales; 2003-2009	s		balance of probability (ie, > 50% chance)	0.19-0.39) intra-class correlation across a single review		inclusion of only respiratory conditions from medical wards • 191 case notes for those admitted with respiratory complaints and those 65 years and over • Case notes randomly assigned to 2-7 reviewers.

Author Manuscript

This article is protected by copyright. All rights reserved.

							(Total of 653 reviews)
Flaatan et al, 2017[29]	Norway; 2011	3 acute hospitals	1-5	≥ 4 "Possibly preventable"	Not reported	Not reported	• All hospital deaths across 3 hospitals in 2011 (including emergency departments) • 1,185 death notes reviewed across one-year period • Case notes assigned to

Author Manuscript

This article is protected by copyright. All rights reserved.

							six consultant reviewers each from different specialties
Kobewka et al, 2017[11, 12]	Canada; 2013	1 acute hospital	0-100	> 50 "Possibly preventable"	ICC = 0.14 (480 deaths each reviewed by 4 reviewers; reliability for average of four reviewers reported as 0.68)	N/A (Insufficient denominator)	• Hospital, excluding pediatrics • 480 deceased case notes (structured case abstracts) produced across 3-month admission period • Case notes randomly

Author Manuscript

This article is protected by copyright. All rights reserved.

							assigned to 4 physician reviewers
Roberts et al, 2017[36]	United Kingdom; 2012-2015	4 Northeast England, UK acute care Trusts (~23 hospitals)	1-6 (PRISM) 1-5 (NCEPOD)	≥ 4 (PRISM) ≥ 3 (NCEPOD)	κ = N/A Not reported for this study, authors cited a reliability estimate of κ = 0.45 from PRISM[32-33]	N/A	• All hospital deaths across 4 Trusts • 7,370 medical records reviewed • Case notes reviewed predominantly by consultants, some by nurses.

[a] Scale of degree of preventability. This tends to range from "6, (virtually) certain evidence of preventability" to "1 (virtually) no evidence for preventability."

Author Manuscript

This article is protected by copyright. All rights reserved.

[b] We have reversed the scale to facilitate comparisons with other studies. The original scale ranged from 1, *(definitely) preventable death* to 4, *(definitely not) preventable death.* Cases with a grade of 2 or lower *(probably or definitely),* on the original scale, were considered as preventable.

[c] For cerebrovascular accident, myocardial infarction and pneumonia, respectively.

[d] We have reversed the scale to facilitate comparisons with other studies. The original scale ranged from 1, *(definitely) preventable death* to 5, *(definitely not) preventable death.* Cases with a grade of 2 or lower *(probably or definitely),* on the original scale, were considered as preventable.

[e] "In your judgment, is there some evidence that the patient's death was avoidable if the problem/s in health care had not occurred?"

[f] The "England study has been extracted from the 2016 paper as the US data has been included in Hofer and Hayward.[31]

Table 2. Summary of Study Processes and Review Methods

Category		No.	References
Inclusion of a screening stage	No screening stage	4	32, 33, 36, 37
	Yes (16-18), criteria	15	10, 14-26, 31, 34, 35, 38-46
	Trigger tool	4	15, 26, 34, 38
Scale used for implicit judgement	Binary	0	
	4 point Likert	2	21, 26
	5 point Likert	3	13, 31, 36
	6 point Likert	16	10, 14-20, 22-25, 32-46
	Continuous	2	11-13
Reviewer screening stage 1	Physician	7	13, 14, 18, 19, 27-29, 32, 33, 36, 44-46
	Nurse	11	14-19, 21-25, 34, 35, 37-42, 44-46

Author Manuscript

This article is protected by copyright. All rights reserved.

	Pharmacist	1	38
Reviewer review stage 2	Physician expert Advice available	15	14-25, 27, 28, 34-46
	Pharmacist support	0	
	Nurse support	0	
Duration of expert advice	Indefinite duration	3	10, 33, 36
	Temporary duration	3	16, 17, 21, 23-25
	No stated duration	2	13, 33
Reviewer affiliations	External to the institution being reviewed	20	10-26, 31-35, 37-46[a]
	Internal	2	21, 36*
Hospital anonymization	Undertaken	5	13, 23-25, 31-33
	NOT undertaken	17	10-12, 14-22, 26-28, 34-46
Clinical experience of physicians	< 5 years	0	
	5-10 years	4	11, 12, 15-17, 20
	> 10 years	7	21, 32-34, 36, 37, 43
	Previous experience not mentioned	2	10, 39-42
	No mention of experience	5	22-28, 35
Speciality of physicians	General medicine/internal medicine (alone)	13	10, 15-17, 20-25, 32, 34, 35, 37, 38, 43
	Internal medicine and specialists	9	11-14, 18, 19, 21, 26, 31, 33, 36, 39-42, 44-46
Review discrepancies	Physicians	3	14, 18, 19, 36, 43-46

Author Manuscript

This article is protected by copyright. All rights reserved.

and disagreements	Nurses	0	
reconciled	Medical health	1	22
	Analysts/records analyst		
	Executive board	2	16, 17, 37
	Information not available	6	20, 21, 23-28, 39-42
Physician reviewer	≤ 1 day	7	14, 18, 19, 21, 23-25, 27, 28, 32, 33, 38,
Training duration			44-46
	1-3 days	7	13, 20, 31, 34, 36, 39-43
	≥ 3 days	3	16, 17, 35, 37
	Not stated	4	10-12, 15, 26
Training content	Case note Exposure	12	10, 13, 14, 18-28, 31, 36, 37, 44-46
	Specialist advice Provided	8	14, 16-19, 21, 23-25, 27, 28, 31, 32, 36,
			44-46
	Absence of preventability	18	10, 13-20, 22-26, 31-35, 37-46
	definition		
	Familiarity with study	14	10, 13, 14, 18-25, 27, 28, 33, 34, 36-42,
	tools		44-46

[a] Best et al.[21] was half external and half internal.

Author Manuscript

This article is protected by copyright. All rights reserved.

Table 3. Preventable Mortality and/or Adverse Events Reported in the Included Studies

Author, Year (Country)	No. of Admitted Patient Case Notes Sampled for Review	No. of Deceased Patient Case Notes Reviewed	No. of Admission Case Notes Selected After Screening for Review By Physicians	Preventable AEs (% of Admissions)	Preventable AEs (% of all AEs)	Preventable mortality (% of admissions)	Preventable mortality (% of deceased)	Threshold for preventability & comments[a,b]
Dubois and Brook, 1988 (United States)[26-28]	1,946	182	1,946	NR	NR	4.6% (weighted estimate, calculated n = [90]/1,946)	26.9% 49/182 14% 25/182	Preventability score ≥ 3 out of 4[c] (majority decision) Preventability score ≥ 3 out of 4[c] (unanimous decision)

Author Manuscript

This article is protected by copyright. All rights reserved.

Brennan et al, 1991 (United States)[22]	30,121	NR	7,743	306 (1.02% weighted)	3.96% 306/7743	0.30% 89/30,121	Not reported	Causation score ≥ 1 on a 0-6 scale; preventability score ≥ 4 out of 6
Hayward et al, 1993 (United States)[10]	675	135 (calculated, reported as 20% of sample)	675	NR	NR	0.44% [3]/675 Weighted for over-sample of deaths	9% [12]/135 (n = 12 calculated from rate reported)	Preventability score ≥ 4 out of 6
Best and Cowper, 1994 (United States)[21]	NR	222[d]	NA	NR	NR	NR	21.6% median	Preventability score ≥ 3 out of 4
Wilson et al, 1995 (Austral	14,179	114	1,718	1,205 (8.50%)[e]	NR	0.55% 78/14,179	29.00%	Causation score ≥ 2 out of 6; preventability score ≥4 out of 6

Author Manuscript

This article is protected by copyright. All rights reserved.

ia)[43]								
Thomas et al, 1999; 2000a; 2000b; 2002 (United States)[39-42]	14,700	NR	448	3.00% 448/14,700	NR	0.265% 39/14,700	NR	Causation score ≥ 4 out of 6; preventability: "an adverse event was considered preventable if it was avoidable by any means currently available unless that means was not considered standard care." The implicit judgement methods are similar to those used in Bates et al.[70]

Author Manuscript

This article is protected by copyright. All rights reserved.

Haywar d and Hofer, 200131 (United States)[31]	NA	111	NA	NA	NR	0.23%-0.61% (at least possibly preventa ble) (95% CI)	22.7%; 6.0% (weighte d for samplin g design)	Preventability score ≥ 3 out of 5[f] Preventability score ≥ 4 out of 5[f]
Davis et al, 2001; 2003. Briant et al, 2006 (New Zealand) [23,24]	6,579	118	850	6.28% 413/6,5 79	48.6% 413/85 0	0.36% 24/6579	19.8%-20.7%	Causation score ≥ 2 out of 6 Preventability score ≥ 2 out of 6
Baker et al, 2004 (Canada)[20]	3,692	236	1,512	2.8% (95% CI, 2.0% to 3.6%)[h]	7.01% 106/1, 512[g]	0.66% (95% CI 0.37% - 0.95%)[h] (death from preventa ble AE)	16.9% 40/236[f]	Causation score ≥4 out of 6 Preventability score ≥ 4 out of 6[f]

Author Manuscript

This article is protected by copyright. All rights reserved.

Michel et al, 2007 (France) [35]	8,754	NR	NR	1.08% 95/8,754	NR	0.09% 8/8,754	NA	Causation score ≥ 4 out of 6 Preventability score ≥ 4 out of 6
Soop et al, 2009 (Sweden) [37]	1,967	10	241	8.6% 169/1,967	70.1% 169/241	0.25% 5/1,967	NR	Causation score ≥ 4 out of 6
Aranaz-Andres et al, 2008; 2009 (Spain) [16,17]	5,624	225	1,755	11.65% 655/5,624	37.3% 655/1,755	0.07% 5/5,624[i]	4.5%	Causation score ≥ 4 out of 6 Preventability score ≥ 4 out of 6
Aranaz-Andres et al, 2011 (Argentina, Colombia, Costa	11,379	NR	1,754	10.47% 1,191/11,379	59% 674/1144	NR	NR	Causation score ≥ 4 out of 6 Preventability score ≥ 4 out of 6

Author Manuscript

This article is protected by copyright. All rights reserved.

Rica, Mexico, Peru)[15]								
Martins et al, 2011 (Brazil)[3,4]	1,103	94	1,103	5.07% 56/1,103	5.07% 56/1,103	2.3% 25/1,103 (coexisting previous AE and death)	26.6% 25/94	Causation score ≥ 4 out of 6 Preventability score ≥ 4 out of 6
Hogan et al, 2012 (England)[32,j]	NR	1,000	NA	NR	NR	NR	5.2% 52/1,000	Preventability score ≥ 4 out of 6 (reporting 1 of 3)
Sorinola et al, 2012 (England)[38]	NR	400	NA	NR	NR	NR	3.5% 14/400	Preventability score ≥ 4 out of 6
Gupta et al, 2013 (USA)[30]	NR	1,683	NR	NR	NR	NR	2.50% 42/1,683	Preventability score ≥ 4 out of 5

Author Manuscript

This article is protected by copyright. All rights reserved.

Baines et al, 2013; 2015. Zegers et al 2007; 2009; 2011a; 2011b (The Netherlands)[14, 18, 19, 44-46]	11,949	762	1,130	NR	NR	NR	4.5%	Preventability score ≥ 4 out of 6
Hogan et al, 2015 (England)[33,k]	NR	2,400	NA	NR	NR	NR	3% 101/2,400	Preventability score ≥ 4 out of 6
Manaseki-Holland et al,	NR	191[l]	NA	NR	NR	NR	10% (median) Q1 3% Q3 28%	Preventability score ≤ 2 out of 5

Author Manuscript

This article is protected by copyright. All rights reserved.

2016 (England)[13]								
Flaatten et al, 2017 (Norway)[29]	59,605	1,167	NR	NR	NR	0.057% 34/59,605	2.91% 34/1,167	Preventability score ≥ 50 out of 100
Kobekwa et al, 2017 (Canada)[11,12]	14,267	480[j]	NR	NR	NR	0.22% 31/14,267	6.46% 31/480[m]	Preventability score ≥ 50 out of 100[m]
Roberts et al, 2017 (UK)[36]	NR	7,194	NR	NR	NR	NR	0.47% 34/7,194	Preventability score ≥ 50 out of 100

Abbreviations: NA, not assessed; NR, not reported.

[a] Causation score is the score given to the likelihood of the adverse event being caused by medical care/management. A causation score of ≥ 2 out of 6 corresponds to "at least slight to modest evidence of management causation"; a causation score of ≥ 4 out of 6 corresponds to "management causation more likely – more than 50/50."

Author Manuscript

This article is protected by copyright. All rights reserved.

[b] A preventability score of ≥ 2 out of 6 corresponds to "at least slight to modest evidence of preventability"; a preventability score of ≥ 4 out of 6 corresponds to "preventability more than likely – more than 50/50."

[c] We have reversed the scale to facilitate comparisons with other studies. The original scale ranged from 1, *(definitely) preventable death* to 4, *(definitely not) preventable death.* Cases with a grade of 2 or lower *(probably or definitely),* on the original scale, were considered as preventable.

[d] Pairs were matched across high observed-to-expected mortality (OTEM) and low OTEM Veteran affairs hospitals

[e] This indicator is for deaths considered with a high level of preventability.

[f] Figures are taken from direct author response rather than published data.

[g] Of 255 patients with iatrogenic adverse events, 106 had > 50% probability of preventability.

[h]Adjusted for sampling frame.

[i] Associated with preventable AE.

[j] "Was the patient's death due to problems in the healthcare or did problems in healthcare contribute to the death?"

[k] "In your judgment, is there some evidence that the patient's death was avoidable if the problem/s in health care had not occurred?"

[l]Multiple reviews were undertaken with the case notes.

[m] > 50% probability of membership in the "possibly preventable" class.

Author Manuscript

This article is protected by copyright. All rights reserved.