# Preventable Deaths at Acute Care Hospitals

Daniel Kobewka

Thesis submitted to the

Faculty of Graduate and Postdoctoral Studies

in partial fulfillment of the requirements

for the Master of Science degree in Epidemiology

Department of Epidemiology and Community Medicine

Faculty of Medicine

University of Ottawa

# Table of Contents

## List of Tables

## List of Figures

## Legend

AE: Adverse Event

AIC: Akaike Information Criteria

BIC:  Bayesian Information Criteria

CI: Confidence Interval

HMPS: Harvard Medical Practice Study

HSMR: Hospital Standardized Mortality Ratio

ICC: Intraclass Correlation

ICU: Intensive Care Unit

IOM: Institute of Medicine

IQR: Inter-Quartile Range

LCA: Latent Class Analysis

OHDW: Ottawa Hospital Data Warehouse

OR: Operating Room

POD: Post Operative Day

SE: Standard Error

SD: Standard Deviation

UTI: Urinary Tract Infection

# Acknowledgements

Thank you to my supervisors Drs. Alan Forster and Carl van Walraven for their guidance in the design, interpretation and writing of this thesis. I learned much from their clear direction and feedback. Thank you to Dr. Monica Taljaard for her help with analyzing the data and interpreting results. I would also like to thank Dr. Paul Ronksley for the helpful discussions about science writing and advice on presentation of results.

Besides my advisors I would like to thank the many members of the division of General Internal Medicine who spent their time reviewing cases to judge preventability. I am grateful for the financial support provided to me by the Department of Medicine. Lastly I want to thank my wife Dr. Carlynne Kobewka for her support and encouragement.

# Abstract

*Background*

Previous measurements of preventable death in hospital do not account for the uncertainty of preventability ratings.

*Objective*

To determine the proportion of deaths in hospital that a have high probability of being prevented with high quality care.

*Methods*

We created summaries for every death at a tertiary care hospital over 4-months. Four reviewers assigned preventability ratings to each death and latent class analysis was used to classify deaths into high and low preventability categories.

*Results*

There were 480 decedents with mean age of 73.9. Inter-rater reliability was poor with an intra-class correlation of 0.14. The best latent class model found that 6.2% (95% CI 0.00 – 15.2%) of deaths had a 31.0% probability of being rated more likely preventable than not by each reviewer. In contrast, 93.8% (95% CI 84.8 - 100.0%) of deaths had a 0.8% probability of being rated more likely preventable than not by each reviewer. The incidence of *truly* preventable deaths is less than the 6.2% that are deemed *possibly* preventable.

*Conclusion*

Very few deaths in hospital are preventable. The low incidence of preventable deaths and low inter-rater reliability means that peer review methodology is only sensitive to large differences in preventable death rate.

# Background

## *1.1 Harm caused by healthcare*

The mission of the publicly funded health system in Canada is "…to improve the lives of all of Canada 's people and make this country's population among the healthiest in the world as measured by longevity…" (1). While longevity is not the only measure of health, it is one of the most tangible and the focus of much medical research. Longevity is a central focus in acute care hospitals in particular where most care is aimed at prolonging life for people who are acutely unwell. Considering the great efforts that go to prolonging life, it is concerning when healthcare fails to achieve that goal and disturbing if it does the opposite.

In 1999, the Institute of Medicine (IOM) published a report titled " To Err is Human: Building a safer health system". The report extrapolated data from several studies, to estimate that 44,000 to 98,000 people die in the United States every year from preventable medical errors (2). A Canadian study in 2004 estimated that the problem is worse per capita in Canada with 9,000 – 24,000 deaths per year from preventable medical errors (3). The IOM report and subsequent studies have been credited with motivating significant efforts to improve patient safety (4). Many hospitals now measure quality of care and engage in quality improvement initiatives. Examples include hand hygiene audits and education programs, surgical site infection monitoring and programs to prevent pressure ulcers in hospital (5-8).

At a system level numerous quality indicators are measured and reported. One such indicator that indirectly measures preventable deaths is the hospital standardized mortality ratio (HSMR) (9, 10). The HSMR is the ratio of observed mortality rate per admission to the expected rate predicted by the patient's comorbidity level. In theory a HSMR > 1 means there are excess deaths that could be due to poor quality care.

Despite efforts to improve patient safety there is limited evidence that we are making progress (11-13). One possible reason for the lack of progress is that many measures of quality have not been derived or validated properly. This is especially true of many system level measures of quality-of-care, including HSMR, that are neither accurate nor precise (14-18). Without accurate and reliable measurements of quality, progress cannot be tracked and successes cannot be studied or expanded upon. Preventable deaths are perhaps the most tangible outcome from poor quality care. Therefore measurement of excess deaths caused by poor quality care should be a central quality measure. Unfortunately measures of preventable deaths including HSMR and labor-intensive chart review methods are inaccurate and unreliable (19, 20). This thesis aims to measure the proportion of deaths that are preventable with high quality care and explore the uncertainty of preventability ratings.

The following section (1.2) explains patient safety terminology and discusses the difficulty in identifying when an AE was directly caused by an error. The next section (1.3) reviews the literature on patient harm focusing on measurements of preventable death in hospital. I have included a detailed summary of the methods used to detect preventable deaths in previous studies and highlighted the methodological differences between the studies. The literature review is important because it puts this thesis in context of previous work and explains shortcomings of previous measurements. The last section of the introduction is a critic of previous preventable death measurements, examining the precise definition of preventable death and some of the measurement problems associated with using peer assessment to detect rare events.

### *1.2 Patient safety definitions*

Patient safety is measured by determining the risk of harm experienced by a patient. This is accomplished through the detection of so-called adverse events (AE), defined as 'a poor health outcome resulting from medical care and not from disease'. AEs are further classified as to their preventability. If an AE is judged to be due to an error, then it is classified as preventable. An error, according to the IOM report, is a "…failure of a planned action to be completed as intended or the use of a wrong plan to achieve an aim." (2) Even though these definitions sound clear cut, they are very difficult to measure.

People admitted to hospitals have illnesses that frequently cause disability and death. This makes it difficult to determine if a particular poor outcome was caused by the patient's underlying disease state or by medical care. Even with the best medical care, some patients will have poor outcomes. It is even more difficult to determine if a particular poor outcome is preventable. For example an intra-cranial hemorrhage in a patient given thrombolytics for a stroke is caused by medical care but probably not preventable if the administration of the thrombolytics was appropriate. On the other hand, death from septic shock when antibiotics were administered late, may or may not be preventable (21).

### *1.3 Preventable deaths in hospital: estimates and methods*

Many studies have attempted to measure the proportion of deaths in hospital caused by medical care using a range of methods **(Table 1).** *Brennan et al*. published the first measure of preventable death in hospital in 1991, based on patients admitted to 31 New York state hospitals in 1984 (22, 23). This study, dubbed the "Harvard Medical Practice Study" (HMPS), used retrospective chart review by clinical experts to judge if an adverse event had occurred. The authors used purposive sampling that under sampled women admitted for normal delivery and people over age 70. They over sampled births with complications and patients admitted to neurosurgery, vascular surgery, cardiac surgery, orthopedics and urology. Next, non-physician research personnel applied 15 screening criteria to further increase the probability of finding an adverse event. The screening criteria included such elements as: death occurred in hospital; patient left hospital with a neurologic deficit; and patient had a myocardial infarction after an invasive procedure. Two physicians then reviewed all cases that passed the screen using a standardized form to document adverse events. Physicians rated their confidence that an adverse event had occurred on a 6-point Likert-like scale. A third physician reviewed cases where an adverse event was found by either of the 2 reviewers to confirm the findings. Only events given a rating greater than 3 out of 6 by the third reviewer were classified as an AE. After adjusting for the purposive sampling frame, this study found that 0.51% of hospital admissions resulted in a preventable death.

The methodology of *Brennan et al.* has subsequently been used for studies in Australia, the United Kingdom, Utah and Colorado United States, Canada, The Netherlands, Denmark, New Zealand and Sweden (24-30). All of these studies were designed to detect adverse events; only a small portion of the AEs in these studies led to preventable deaths. Most of the studies do not directly report *the proportion of AEs causing death per admission,* so the following formula is used to estimate this value:

(% of admissions with an adverse event) X (% of adverse events that were judged preventable) X (% of adverse events where the patient died).

The proportion of preventable deaths per admission in these studies using this calculation was 0.55% (Australia), 0.41% (GB), 0.13% (USA), 0.66% (Canada), 0.28% (Netherlands), 0.18% (Denmark), 0.28% (New Zealand) and 0.26% (Sweden). This calculation makes two assumptions that are likely incorrect: 1) the proportion of adverse events in the equation is independent from the first 2 proportions and 2) the adverse event is the direct cause of each death. The first assumption could bias the estimate in either direction while the second assumption inflates the estimate of preventable deaths. Despite these questionable assumptions, the Institute of Medicine appears to have used this same method in their 1999 report titled "To Err is Human" when they estimated that 44,000-98,000 people die due to medical care per year in the United States (2, 31). The error in the 2 assumptions was confirmed by a study in Australia that used the same methods as the original study by *Brennan et al.* but performed further chart reviews to estimate how many deaths were directly cause by a preventable adverse event (29). The study found that even though 0.28% of admissions had a preventable AE that was related to the patient's death, less than half (0.13% of admissions) had a preventable adverse event that was judged to have directly caused the death.

There have been numerous studies that attempt to directly measure the proportion of preventable deaths. *Dubois et al.* published a study shortly after *Brennan et al.* using a sample of decedents who had a most responsible diagnosis of stroke, pneumonia or myocardial infarction. They sampled 35 deaths from each of 12 hospitals (for a total of 420 deaths). Six American hospital had the highest corporate death rate and 6 had the lowest

(32). A summary of each death was dictated and then evaluated independently by 3 physicians. The physicians used a four point Likert scale to rate preventability, then deaths were classified as preventable if 2 out of 3 or 3 out of 3 physicians gave a rating of greater than 2 out of 4. This method found that **27%** of deaths in hospital were preventable. Using a criterion of unanimity, where a death was classified as preventable if 3 out of 3 physicians rated it of greater than 2 out of 4, 14% of deaths in hospital were classified as preventable.

A decade later in 2000 *Hayward et al.* performed a more rigorous measurement of preventable deaths (33). Once again using a purposive sample over sampling patients who had electrolyte abnormalities or digoxin toxicity. The sample consisted of **111** deaths out of 4198 total deaths that occurred at 7 veteran's administration hospitals in 1995 and 1996. Physicians reviewed cases by reading the patient's chart; approximately 60% of the cases were reviewed multiple times. Each reviewer was asked to rate preventability through better quality care on a 5-point Likert scale. They were also asked, "What do you estimate the likelihood of the prevention of death to be if care had been optimal?" The response to this question was a probability rating from 0-100. The authors then performed a Monte-Carlo simulation to determine the effect of inter-rater reliability on the estimates of preventability of death. There was poor inter-rater reliability in assigning probabilities of preventability and the distribution was highly skewed with many outliers having a large effect on the estimates. After adjusting for sampling frame they found that **6%** of decedents would have left the hospital alive if they had received higher quality care. After adjusting for reliability of reviews and the skewed distribution, only **1.3%** of deaths were preventable. To put this in terms of preventable deaths per admission, 0.11% of admissions resulted in a preventable death before adjustment and 0.025% after adjusting for inter-rater reliability.  These numbers are much smaller than those from the original study by *Brennan et al.* suggesting that either the proportion of preventable deaths per admission was much lower in Heyward's sample or perhaps the rigorous methods gave a more accurate measure of preventable deaths.

There are two more recent studies that have directly addressed the question of preventability of death in hospital.  The first study, by *Zegers et al.,* used methods very similar to the original study by *Brennan et al.* but focused only on preventable deaths instead of AEs (34).  They used an initial screening process followed by physician review using a 6-point Likert scale.  Disagreements between the reviewers were resolved with a consensus procedure.  This study found that **4.1**% of deaths in hospital were preventable with better quality of care.  The most recent study by *Hogan et al.* published in 2012 used methods from the original study by *Brennan et al.* and found that **5.2**% of deaths were preventable with better quality medical care (35).  A summary of all these studies can be found in **Table 1.**

**Table 1.** Characteristics and results of previous studies that have measured the proportion of in-hospital deaths classified as preventable

| ***Author*** **Journal. Year** | **Setting and sample** | **Outcome** | **Methods to detect a preventable death** | **Result** |
|---|---|---|---|---|
| Dubois et al. Annals of Internal Medicine 1988(32) | 182 deaths in 12 acute care hospitals in the United States. 6 of the hospitals had death rates 2 standard deviations above expected and 6 had death rates 2 standard deviations below expected. | Preventable deaths | One physician reviewed the chart of patients who died from pneumonia, myocardial infarction and stroke over a 6 month period and dictated a summary using a standard format. Expert panels specific to each of the three diagnoses reviewed the dictated summary and rated preventability on a 4 point scale. Each case was reviewed by 3 physicians. If more than 1 out of 3 of physicians rated individual cases as 3 or 4 out of 4, the death was categorized as "probably preventable". | 27% of in hospital deaths deemed probably preventable.<br>14% preventable by unanimous decision |
| Brennan et al. NEJM 1991(22) | Purposive sample of discharges from hospitals in New York in 1984. A higher proportion of patients were selected from specialties that are prone to adverse events. n=31,426 | Adverse events | A 2-stage process was used. Stage 1: 16 explicit criteria were used to screen cases. Stage 2: Two physicians reviewed the cases that screened positive using a structured adverse event analysis form. The form guides the reviewer to look for injury or disability and judge on a scale of 1-6 whether it was caused by the medical care. Each cases was reviewed by 2 physicians and if there was disagreement a third physician reviewed it to decided on the adverse event and the preventability rating | 3.7% of admissions had an AE<br>***0.51% of hospital admissions result in AE causing death.** |
| Wilson et al. MJA 1994 (24) | Random sample of patients discharged from 28 hospitals in 2 states of Australia in 1992. n=14179 | Adverse events | Same methods used by *Brennan et al.* but 18 screening criteria were used | 16.6% of admissions had an AE<br>***0.55% of hospital admissions result in AE causing death.** |
| Thomas et al. Med Care 2000(26) | Stratified sample from 28 hospitals in Utah and Colorado USA. n=15,000 | Adverse Events | Same methods used by *Brennan et al.* but 18 screening criteria were used and only one reviewer per case. | 2.9% of admissions had an AE<br>***0.13% of hospital** |

| | | | | |
|---|---|---|---|---|
| | | | | **admissions result in AE causing death.** |
| Vincent et al. BMJ 2001(25) | Random sample of admissions to 2 hospitals in London England in 1998. N=1014 | Adverse events | Same methods used by *Brennan et al.* but 18 screening criteria were used and only one reviewer per case. | 10.8 % of admissions had an AE<br>***0.41% of hospital admissions result in AE causing death.** |
| Schioler et al. Ugeskr Laeger 2001 (27) | Sample from 17 hospitals in Denmark in 2001. n=1097 | Adverse Events | Same methods used by *Brennan et al.* but 18 screening criteria were used and only one reviewer per case. | 9.0% of admissions had an AE<br>***0.18% of hospital admissions result in AE causing death.** |
| Hayward et al JAMA. 2001 (33) | Sample of 111 deaths at VA hospitals, | Preventable deaths | An implicit review instrument with 18 questions about specific aspects of care. One questions was "was the death preventable by better quality of care" Also asked the likelihood of prevention of death if care had been optimal. 383 reviews of the 111 deaths. Only 62 had multiple reviews done. Preventability was calculated as the weighted sum of the probability of survival to discharge given by each reviewer. Then did a Monte-Carlo simulation of the effect of inter-rater reliability on estimates of preventability of death. | 22.7% of deaths possibly preventable.<br>6.0% probably or definitely preventable.<br>0.5% of deaths were preventable after adjusting for inter-rater reliability and the probability of leaving the hospital alive |
| Davis et al. N Z Med J. 2003 (28) | Random sample from 13 hospitals in New Zealand in 1998. n=6579 | Adverse Events | Same methods used by *Brennan et al.* but 18 screening criteria were used and only one reviewer per case. | 12.9% of admissions had an AE<br>***0.28% of hospital admissions result in AE causing death.** |
| Baker et al. CMAJ 2004(3) | Random sample of patients discharged from 20 hospitals across Canada. n=4164 | Adverse events | Same methods used by *Brennan et al.* but 18 screening criteria were used and only one reviewer per case. | 7.5% of admissions had an AE<br>**0.66% of hospital admissions result in AE causing death.** |

| | | | | |
|---|---|---|---|---|
| Briant et al. N Z Med J 2006 (29) | Subgroup analysis of data from *Davis et al.* 118 patients from the sample who died | Adverse event causing a preventable death | Same methods used by *Brennan et al.* The patients that died and had AE were re-reviewed to determine if the AE directly caused the death and if the AE was preventable. | **0.23% of admissions resulted in death that were probably preventable** 0.13% were highly preventable. |
| Zegers et al BMJ quality and safety 2008 (34) | Sample of 7926 admissions to Dutch hospitals. | Adverse events | 18 screening questions applied that suggested adverse event. Then 2 physicians independently reviewed those that were screen positive to determine the nature, impact, clinical process and degree of preventability. A 6-point Likert scale was used to determine preventability. If they had differing opinions, a consensus process was used with a 3rd reviewer if no consensus could be reached. | 4.1% of deaths were due to a preventable AE |
| Soop et al nt J Qual Health Care 2009. (30) | Representative sample from 28 hospitals in Sweden in 2003-2004. n=1967 | Adverse Events | Same methods used by *Brennan et al.* but 18 screening criteria were used and only one reviewer per case. | 12.3% of admissions had an AE ***0.26% of hospital admissions result in AE causing death.** |
| Hogan et al BMJ Quality and Safety 2012 (35) | 1000 deaths from 10 different hospitals in the UK | Preventable deaths | Reviewers first judged if there was a problem in care that contributed to the death. Reviewers then rated preventability of the death on a 6 point Likert scale. They were judged preventable if score was 4-6. Used a structured data collection form. Each preventable death was re-reviewed by the PI and an expert reviewer. 25% of all deaths were re-reviewed. | 5.2% of deaths were preventable. 13.1% of deaths had a care problem that contributed to the death. |

* Number estimated and not directly reported in the paper. Assumes independence of AE causing death and preventable AE.

*1.4 Critique of previous measures of preventable deaths*

There have been many criticisms of the methods used to measure preventable deaths. *Macdonald et al.* argue that the results of the Harvard Medical Practice Study (HMPS) published by *Brennan et al.* grossly overestimate preventable deaths because the study had no control group and was therefore unable to adjust for the baseline risk of death (20). There is an assumption in the HMPS, and most of the related studies, that when a patient who experienced a preventable AE subsequently dies, the death was caused by the AE; however, a causative relationship was never assessed even subjectively by the reviewers (31). Even in the few studies that assess whether the AE directly caused the death, only a single study considered the life expectancy of decedents (33). The preventable death of a patient with several months life expectancy, although unacceptable, is not comparable to the preventable death of a patient with decades of life expectancy.

*Leape,* one of the co-authors of the HMPS, argues that the estimates of preventable death are not exaggerated because: retrospective review is more likely to miss AE rather than detect AE's that did not occur; the HMPS did not measure deaths that occurred from medical care outside of the hospital and; prospective AE surveillance studies have found even higher rates of AE than those in the HMPS (19). *Leape* also proposes that peer review may be insensitive because physicians are reluctant to criticize colleagues.

Most studies that use the HMPS methods count an adverse event if 2/3rds of reviewers assign a probability rating greater than or equal to 4 out of 6. These cut offs are arbitrary and do not acknowledge the uncertainty in the judgments on AE or preventability. *Thomas et al.* found that the preventable adverse event rate ranged between 0.8% and 32.3% of admissions when criteria required for adverse event classification (including reviewer confidence and the number of reviewers that had to agree) were changed (36). Aside from this study by *Thomas et al.,* none of the studies of preventable death have sensitivity analyses to determine the effect of this uncertainty on the preventable death rate.

The final concern with previous studies on preventability of death is the poor inter-rater reliability of implicit review. All measures of AEs or preventable death use some form of peer assessment to determine if a death is preventable. Inter-rater reliability measured by the kappa statistic is usually in the range of 0.2 – 0.3 (33, 36-39) with several studies reporting values >0.4 (40, 41). Some authors have proposed that discussion between reviewers will improve reliability; but several studies have shown that, while reviewers have better agreement after discussing a case, their judgment is no closer to the truth (42, 43).

Low inter-rater reliability is important because it introduces a bias that leads to overestimation (44). This is a well-described epidemiologic phenomenon that occurs when an imperfect test is used to detect a rare condition (44). It can be explained with a simple thought experiment. Suppose that a test has 90% sensitivity and 90% specificity to detect a disease with a prevalence of 1% (Table 2). The test will correctly classify 0.9% of the population as having the disease and will ***in***correctly classify 9.9% of the population as having for the disease. This results in 10.8% of the population being test positive; this greatly over-estimates the true prevalence of disease, which is only 1%. The bias is inflated as a test's accuracy decreases and as disease prevalence decreases.

To correct for the bias introduced by a poor test, each patient must have multiple measurements of preventability on a continuous scale. This explicitly acknowledges the uncertainty in assessing preventability and allows measurement of inter-rater reliability so it can be corrected for (44). There are multiple techniques to correct for measurement error, most of which involve modeling the variance from measurement error so that it can be removed (45-48). Among all previous studies only one explicitly acknowledges and corrects for this bias (33).

**Table 2**: Example demonstrating the overestimation of disease prevalence that occurs when an imperfect test with sensitive and specificity of 90% is used to detect a disease with 1% prevalence.

| | | **Disease** | | |
|---|---|---|---|---|
| | | YES | NO | **TOTAL** |
| **Test** | POSITIVE | 9 (0.9%) | 99 (9.9%) | ***108 (10.8%)** |
| | NEGATIVE | 1 (0.1%) | 891 (89.1%) | **892 (89.2%)** |
| | TOTAL | **10 (1%)** | **990 (99%)** | **1000 (100%)** |

* Total number in population who will be test positive

### *1.5 Summary of rational for measuring preventable deaths*

The large range in the proportion of preventable in hospital deaths (0.5% – 27% (32, 33)) is likely due to a combination of bias, differing definitions of preventable death, random error and differences in the true proportion of preventable deaths. Only the paper by *Heyward et al.* addresses the probabilistic nature of preventability ratings, corrects for poor inter-rater reliability and acknowledges the life expectancy of the decadents (33). *Heyward's* study on preventable deaths is 15 years old, from one region of the United States, used a purposive instead of random sample, and only included 111 deaths. Considering the limitations of previous studies and the importance of preventable death as an outcome, reliable and unbiased measurements of preventable death are needed. Methods used for peer assessment need to be developed to increase reliability and precision. Accurate measurements of preventable deaths are needed to direct efforts to improve safety and monitor improvements across the health system.

## Primary Objective

To determine the proportion of deaths at The Ottawa Hospital judged to be potentially avoidable.

## Secondary Objective

To describe the characteristics of the patients and the hospitalization associated with a death judged to be preventable.

## Methods

### *2.1 Study design and setting*

We used an observational cohort design with retrospective and prospective data collection. The study took place at The Ottawa Hospital, a 1064 bed academic tertiary care hospital (with no pediatric department) in Ottawa, Canada. The hospital serves a catchment area of 1.2 million people. All three campuses including the Civic, General and Heart Institute were included in the study. Rehabilitation centers were excluded.

### *2.2 Population*

Our study included every death that occurred in patients admitted to The Ottawa Hospital between 5 September and 16 December 2013. Patients who died prior to admission (i.e. in the emergency room) and stillbirths were excluded.

### *2.3 Overview of study*

A case summary was created for each death based on chart reviews and interviews with the medical staff involved in the case. Each summary included detailed information about any possible improvements in patient care. Four independent clinical experts were selected at random from a pool of 13 to review the case summary and judge the preventability of each death if the patient had received optimal care. The reviewers also judged whether the patient would be alive in 3-months had they received optimal care. Latent class analysis was then used to determine the probability that each death was truly preventable. The characteristics of the decedent and hospitalization associated with deaths deemed to be preventable death were determined.

### *2.4 Creation of Case Summaries*

Each case summary was created from three independent reviews. When a death occurred, the hospital's computer system automatically generated a record in our database that included the patient's time of death, their demographic information, and free text fields to describe their care in hospital. This blank record was then sent to a nurse and a physician of the same specialty as the most responsible physician who cared for the patient in hospital. The nurse and physician both wrote a summary of the patient's care in hospital that described any ways that the patient's care could have been improved. To complete this task they independently read the chart and, if needed, contacted the physicians and nurses who cared for the patient prior to the death to get further details. In select cases where the care was highly specialized, the physician reviewer documented their expert opinion on whether any gaps in the process of care existed. For example in neonatal deaths the case summary included a judgment statement such as "the resuscitation was appropriate". These opinion statements were only used in cases where the subsequent physician reviewers would be unlikely to have the medical expertise to evaluate the care (e.g. when the error related to a technical aspect of a surgery or neonatal resuscitation). Next, a third reviewer, also a physician, verified the information and reconciled the case summaries from the nurse and physician with further chart review. This third reviewer created a structured case form, explicitly describing any ways that the patient's care could have been improved and summarized the final days of the patient's life. The fields included in this form are shown in table 3 and the forms used for case creation and review can be found in Appendix 1. The only exception to this process occurred in patients admitted to the Heart Institute. There, a single physician created case summaries for these patients because cardiologists and nurses from cardiology could not be recruited to perform data collection.

**Table 3:** Information included in the case summaries that were given to expert reviewers.

| Variable | Description |
|---|---|
| Age | Age in years at death |
| Past Medical History | Previous diagnoses with relevant information about severity or stage of each diagnosis. |
| Admission history | Summary of the events leading up the admission to hospital. Summary of symptoms and time course that lead to admission. |
| Admission physical exam | All pertinent physical exam findings from the exam documented on admission. |
| Course in hospital | A fact based narrative of the diagnoses that were made and how they were treated. For patients with long admissions, new diagnoses that occurred during the admission were described along with treatments received. The final days of each patient's life was described in detail along with any decisions to limit life-prolonging |

| | |
|---|---|
| | treatment at the end of life. |

*2.5 Creation of cases for reviewer calibration testing*

We created 5 cases that were thought to be highly preventable and 5 with very low preventability to test reviewer's ability to determine preventability. These same 10 cases were given to each reviewer (Appendix 3).

*2.6 Selection of reviewers*

An email was sent to all 30 members of the division of General internal medicine at The Ottawa Hospital inviting them to be reviewers for the study. The email explained the purpose of the study, and the time commitment required for participation. No monetary reward or incentive was offered. One physician from critical care, one from anesthesia and 2 from general internal medicine were directly asked to participate in the study.

*2.7 Randomization of Cases*

We made four copies of the case summary database. Each copy was randomized with computer generated random numbers and then divided into 4 equal parts. The 10 calibration cases were dispersed randomly through out each part.

*2.8 Review by clinical experts*

Each clinical expert received instructions on how to judge preventability and data on the proportion of preventable deaths from previous studies. For each case, the reviewers were asked to answer four questions using the form in appendix 1. The first question was: what could have been done differently, that would be considered within the standard of care, to reduce the risk of this patient dying during the current hospitalization? When a reviewer identified an action that may have reduced the chance of the patient dying in hospital, they were then asked to describe the action and give a probability that the action would have prevented the death. The reviewer was then asked to identify if the actions that may have prevented the death needed to be taken during the admission, prior to the admission, or both. Finally, the reviewer was asked to judge whether the patient would be alive in 3 months time using a 4-point Likert-type scale with the ratings: "Almost certainly would be alive"; "Uncertain but probably would be alive"; "Uncertain but probably would be dead"; "Almost certainly would be dead".

The reviewers were instructed that a death should be judged as possibly preventable "when there were gaps in quality of care resulting in an adverse event that caused death". The following definitions were given to reviewers to interpret this statement:

<u>High quality care:</u> is the best care you can expect in the average Canadian hospital. High quality care minimizes the risk of complications, maximizes the likelihood of a good outcome and is respectful of patient's wishes.

<u>Standard of care:</u> The standard should be care processes that would be agreed upon by a group of your peers. Do not hold other physicians to standards that are idiosyncratic to you or controversial within the profession.

An adverse event: is any unintended physical or mental injury that occurs as the result of medical management rather than from the patient's underlying disease process. An adverse event could result from acts of commission (incorrect diagnosis, treatment or poor performance) or acts of omission (a failure to diagnose or treat).

### *2.9 Data sources*

Clinical data were extracted by chart review and from interviews with the physicians and nurses who were caring for the patients when they died. Patient demographic information and admission information was taken from the Ottawa Hospital Data Warehouse (OHDW). The OHDW is a relational database that collects clinical and administrative data from various systems in the hospital. Table 4 contains variables that were obtained from the OHDW.

**Table 4**: Variables obtained from The Ottawa Hospital Data Warehouse

| Variable | Description |
|---|---|
| Age | Age at time of death |
| Gender | Male / Female |
| Marital Status | Single/married/widowed |
| Place of residence | Home/Nursing home |
| Language | Patient's preferred language |

| | |
|---|---|
| Family doctor | Does the patient have a family doctor listed Y/N |
| Risk of Death in Hospital | *The probability of death in hospital on |
| Elixhauser comorbidity score | A summary score of comorbidities – Patient level |
| Number of hospital admissions in the last year | Number of hospital admissions in the last 365 days |
| Number of hospital admissions in the last 6 months | - |
| Most responsible diagnosis | The diagnosis that consumed the most resources during the hospitalization. Grouped by ICD 10-CA chapter. |
| Admitting diagnosis | The diagnosis at the time of admission |
| Length of stay | Number of days in hospital before death |
| Primary service caring for the patient | The medical specialty that is primarily responsible for the patient at the time of death. |

* Risk of death determined from the Escobar plus score. This score has been validated in The Ottawa Hospital Population. (49, 50)

### *2.10 Analysis*

All analysis was performed with SAS 9.3© (Cary, N.C.).

### *2.11 Description of reviewers*

We described the gender, years in practice, and medical specialty of each reviewer. The distribution of each reviewer's ratings was presented. Reviewer reliability was summarized using intra-class correlation (ICC) for a single reviewer and for any random set of 4 reviewers using the methodology described by *Shrout and Fleiss* (51, 52). ICC is the preferred statistic when measuring agreement between reviewers who are rating on a continuous scale (53, 54). The ICC could be interpreted as the proportion of total variance that is due to differences in the subjects and not the reviewers. Using the best fitting latent class model, we calculated the sensitivity and specificity of each reviewer.

### *2.12 Descriptive statistics*

We used the variables listed in **Table 4** to describe the patient's that died during the study period. Continuous variables were summarized using means and standard deviations. Variables that are highly skewed were described with medians and interquartile range or transformed into categorical variables to better communicate the variability. Categorical variables were summarized using proportions.

### *2.13 Methods for First Objective*

#### 2.13.1 Explanation of Latent Class Analysis

We used latent class analysis (LCA) to determine the proportion of deaths that were judged to be preventable if patients received high quality care. A latent variable is one that cannot be directly observed but is instead inferred from other observed variables. For example, a movie's quality cannot be directly measured but can be gauged by: the proportion of reviews that are positive; the quality of previous movies completed by its director; the number of awards for which the movie is nominated or garners.

LCA is a latent variable model that postulates an error free latent classification for each individual. In this study, we postulated that every death was either preventable or not but there was no way to directly measure this variable. LCA uses observed indicator variables that are subject to error (i.e. the individual physicians' reviews) to **determine the probability that** each case resides in a particular latent class. For this thesis, the latent variable was the preventability of hospital deaths; the observed variables were the ratings of reviewers; LCA was used to determine the likelihood that each case resided in a particular preventability class based on the physician ratings.

When one builds a latent class model, the meaning of each latent class is undefined. This means that we cannot predetermine if one latent class will contain deaths that are truly preventable or non-preventable. We can only specify the number of latent classes and then use maximum likelihood estimation methods to estimate the following parameters: the prevalence of each latent class within the study sample; and the probability of receiving a preventability rating conditional on being in each latent class. The meaning of each latent class can, only then, be subjectively described following this analysis by examining the probability that individuals in that particular latent class received a high or low preventability rating. For example, if individuals in a latent class A have greater than 90% chance of receiving a rating of *more than 50% preventable* from each reviewer and only a 10% of receiving a rating of *less than 50% preventable*, then latent class A could be labeled as "Likely preventable deaths".

### 2.13.2 Categorization of Preventability ratings

LCA requires that both the indicator variable and the latent variable be categorical. Transforming the continuous observed, or indicator, variables (the reviewers' perceived probability of preventability) into a categorical variable results in loss of information; however, it has the advantage of requiring no assumptions about the *distribution* of the indicator variable (which is necessary for Latent Profile Analysis, in which the observed variable is continuous). We categorized the indicator variables into varying numbers of intervals with equal ranges and assessed its impact on model parameters and model fit. We used 2, 3, 4, 5 and 6 intervals. We used intervals with equal ranges because it required no assumptions about the meaning of a probability rating. A post-hoc analysis was performed using categories of 0-10% *not preventable,* 11-89% *possibly preventable* and 90-100% *preventable.* We used these categories to capture cases where reviewers were quite certain of preventability or non-preventability and then lumped all uncertain ratings (11-89%) together. Even though reviewers were asked to rate preventability on a 100-point scale, this may not reflect how reviewers think about probability-of-preventability. For example how does a reviewer reliably decide if a case is 45% versus 55% preventable? While these 2 ratings may seem qualitatively the same to the reviewer they will result in different categories of preventability and therefore different models.

### 2.13.3 Model assumptions

The primary assumption of LCA is that the indicator variables are independent of each other within each latent class. This assumption is reasonable because each case was reviewed by a different set of randomly assigned reviewers. For our study, we also assumed a common sensitivity and specificity for each reviewer. While this is unlikely to be true, it is an acceptable assumption because reviewers were randomly selected and assigned to each case. Consequently there will be no bias created by the assignment of reviewers to cases, even if reviewers have variable sensitivity and specificity.

### 2.13.4 Model Selection

The parameters in the LCA were estimated using an iterative approach that sought to maximize the likelihood function. The likelihood function is defined as the likelihood of achieving the observed data given a set of parameter estimates. Iterations of parameter estimates continued until the change in the likelihood function was below a predetermined cut off or a certain number of iterations had been reached.

### 2.13.5 Evaluation of model fit and model selection

To identify the best fitting model we calculated, for each model, the following statistics were used using the methodology described by *Colins et al* (55): absolute model fit; homogeneity; latent class separation; AIC; BIC; and mean posterior probability of class membership for each model. Absolute model fit was tested using the $G^2$ statistic and the Chi-squared distribution. As the ratio of individuals in the sample (N) divided by the number of possible response patterns drops below 5, the Chi squared distribution no longer approximates the $G^2$ distribution (56, 57). In this case, parametric bootstrapping or a procedure called posterior predictive checks can be used to obtain the test statistic distribution (58, 59). These procedures were attempted but the models had too many parameters to be solved.

To evaluate relative goodness of fit of nested models, the likelihood-ratio difference test can be used. Unfortunately when the models being compared differ in the number of latent classes, they are *not* nested and the distribution of the test statistic is unknown. The procedures mentioned above, namely parametric bootstrapping or posterior predictive checks, must then be used to estimate the test statistic distribution. Other measures of relative fit, such as the Akaike Information Criterion (AIC) or Baysian Information Criteria (BIC), that penalize the $G^2$ for added parameters can be used to compare models even if they are not nested (55).

For analysis, the variable "probability-of-preventability" was categorized to use as a categorical indicator variable for latent class analysis. When the probability-of-preventability is categorized, there is loss of information; therefore models with more categories are preferred if they fit well because they are based on more information. I explored how model fit and latent class prevalence changes when probability-of-preventability is divided into different numbers of intervals with equal ranges. There is no test statistic to compare latent class models when the observed variables are categorized differently.

For a latent class model to fit well, there must be *separation* between latent classes and *homogeneity* within each latent class. A latent class is homogeneous when the probability of a certain response, conditional on latent class membership, is close to 1 or 0. Put another way, a latent class is homogenous if the probability of a particular response category (such as 0-50% preventable) is close to 0 or 1. Figure 1 is an example of a latent class model. Each line represents a latent class. The X-axis is probability-of-preventability assigned by reviewers categorized into 2 categories, preventable and non-preventable. The Y-Axis is the probability of an individual receiving a particular preventability rating conditional on being in each class. Both classes in this example have good homogeneity because the probabilities of being rated preventable or non-preventable are close to 0 or 1. The model in the example also has good separation because the probability of receiving a rating of preventable is very different depending on which class an individual is in. The good homogeneity and separation in the example will allow individuals to be placed into a particular latent class with a high degree of certainty.

Latent class homogeneity and separation can be summarized with the *mean posterior probability of correct classification*. Once a latent class model has been created, it can be used to calculate the probability of class membership for each individual. Table 5 contains an illustration of this concept. Each line is an individual in the sample. The next set of columns in the table contains the preventability ratings assigned by the 4 reviewers. The next set of columns contains the probability that the individual is a member of class 1, 2 and 3. The last column is the class that best describes the individual based on the probabilities of class membership. The *mean posterior probability of correct classification* is calculated as the mean probability of class membership for all individuals who are best placed in a particular class. For example the *mean posterior probability of correct classification* for class 1 is calculated as the mean of the values denoted by the * in Table 5. A value close to 1 means that the observed variables predict class membership with a high degree of certainty.

For this thesis the *mean posterior probability of correct classification* was calculated and is presented using color-coding of tables. I used red to represent *mean posterior probability of correct classification* less than 80%, yellow for 80-90% and green for greater than 90%. This classification was used because it represents the range of mean posterior probabilities.

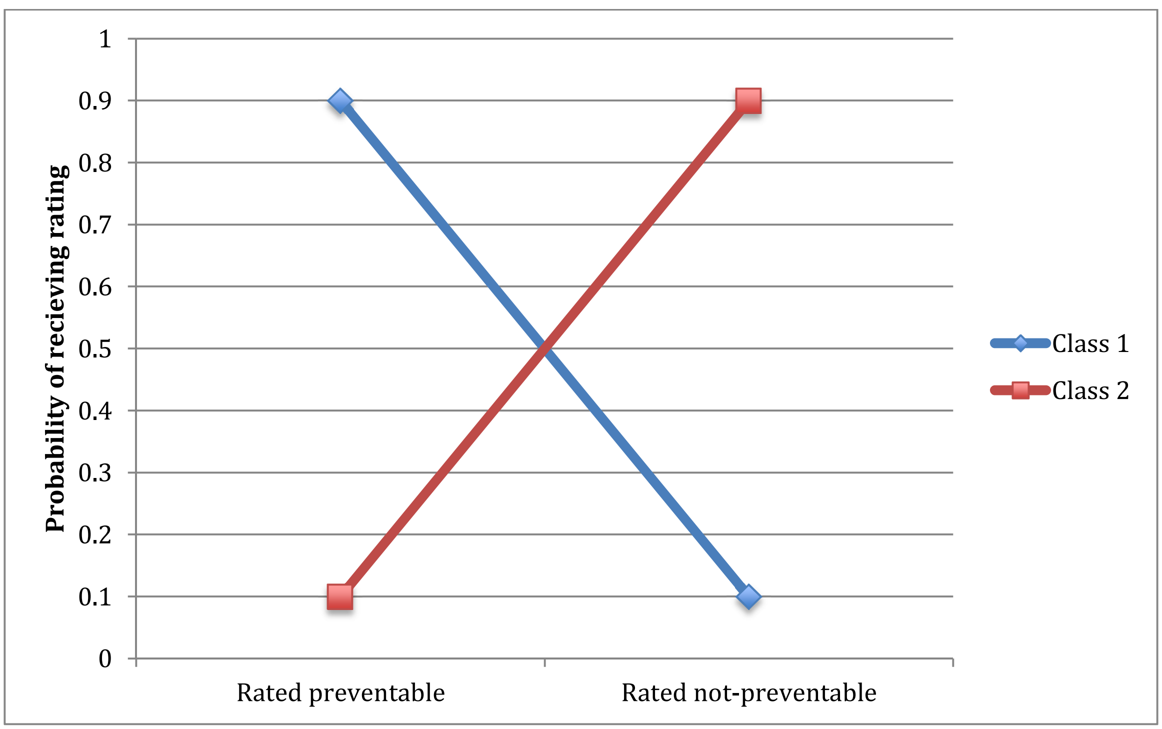



**Figure 1**: Probability of receiving rating conditional on latent class membership.

**Table 5**: Illustration of the probability of class membership for 8 cases.

| Case # | Preventability Rating from Reviewer number: | | | | Probability of fitting into class # | | | Best Class |
|---|---|---|---|---|---|---|---|---|
| | **1** | **2** | **3** | **4** | **1** | **2** | **3** | |
| 1 | >75% | 25-50% | 25-50% | 25-50% | 0.89* | 0.00 | 0.11 | 1 |
| 2 | >75% | 25-50% | 25-50% | >75% | 0.95* | 0.00 | 0.05 | 1 |
| 3 | >75% | >75% | 50-74% | 25-50% | 0.92* | 0.00 | 0.08 | 1 |
| 4 | 50-74% | 25-50% | 25-50% | 25-50% | 0.64* | 0.00 | 0.36 | 1 |
| 5 | 50-74% | >75% | >75% | >75% | 0.77* | 0.00 | 0.23 | 1 |
| 6 | 0-24% | 0-24% | 0-24% | 0-24% | 0.00 | 0.96 | 0.04 | 2 |
| 7 | 0-24% | 25-50% | 0-24% | 0-24% | 0.00 | 0.66 | 0.04 | 2 |
| 8 | 0-24% | 0-24% | 0-24% | 25-50% | 0.00 | 0.66 | 0.04 | 2 |

The mean probability of correct classification for class 1 is calculated as the mean of all probabilities denoted by *

### 2.13.6 Sensitivity Analysis

A post-hoc analysis was performed excluding ratings from one outlying reviewer who rated more deaths as preventable than any other reviewer. The ratings from this reviewer were changed to missing and the analysis was repeated. Missing values were treated as missing at random. This is a reasonable assumption because the cases were randomly assigned to each reviewer. The model-based prevalence of each class and the meaning of each class were compared to the models created using all data from all reviewers.

### *2.14 Methods for Second Objective*

To determine patient characteristics associated with a preventable death, we determined the posterior probability of class membership for each death using the most parsimonious model. Deaths that had a probability of fitting into the preventable death group that exceeded 50% were analyzed together to determine associations. Brief case descriptions and descriptions of what could have been done differently to reduce the chance of each patient dying in hospital are presented. We did not perform tests of statistical significance on the 2 groups because of the low number of preventable deaths and the multiple comparisons. We present the patient and encounter level characteristics associated with preventable deaths.

### *2.15 Ethics approval*

This project was approved by the Ottawa Hospital research ethics board.

## Results

### *3.1 Description of population*

The study population included all admissions occurring over a 3-month period in 2013. During the study period there were 14,290 admissions at The Ottawa Hospital. The mean age was 48.9 years (SD 27.3 years), 43.5% were male. The mean length of stay was 7.1 days (SD 13.8 days) with a median length of stay of 3 days (IQR 2-7 days). Lastly the mean risk of death during the hospitalization calculated using the methods described by *Escobar et al.* was 0.06 (SD 0.12) (49).

There were 480 deaths that occurred during the study period. The mortality rate was 4.9 deaths per 1000 patient days. The mean age of decedents was 73.7 years (SD 16.1 years) 52.6% were male. The mean length of stay was 14.2 days (SD 23.5 days) with a median length of stay of 7 days (IQR 3-17 days). The mean risk of death during the hospitalization calculated using the methods described by *Escobar et al.* was 0.30 (SD 0.20) (49).

### *3.2 Description of reviewers*

We recruited 13 reviewers. Twelve of them rated 120 cases plus the 10 fictitious cases each and 1 reviewer rated all 480 cases plus the 10 fictitious cases. Eleven of the reviewers were certified in internal medicine, two of these reviewers had training in a sub specialty of internal medicine including one intensivist and one respirologist. One reviewer was a certified anesthesiologist. Four reviewers had been in practice for 2 years; 3 reviewers had been in practice for more than 5 years; 6 reviewers had been in practice for more than 10 years. Seven reviewers were male and 6 were female.

### *3.3 Description of reviewer's ratings*

There were 1920 preventability ratings assigned to 480 deaths. The mean preventability rating (0 meaning not preventable, 100 meaning preventable) was 5.7 out of 100 (SD 16.3) with a median rating of 0 out of 100 (IQR 0 – 0). The most common rating assigned was 0 out of 100 (1538/1920 = 80.1%) followed by 10 (62/1920 = 3.2%) and 50 (56/1920 = 2.9%). The proportion of ratings assigned to each probability by each reviewer, is shown in **Figure 2** and **Table 6**. Ratings less than 10 were common (1675 (79.4%)) while ratings greater than 50 were uncommon (51 (2.6%)). **Figure 2** shows that there was more variation in preventability ratings less than 50 between reviewers than for ratings greater than 50. There was one outlying reviewer (number 4 in table 6, represented by the purple line in figure 2) who gave higher ratings of preventability more frequently than all other reviewers. This reviewer rated 29 out of 120 (24.3%) deaths as $\geq$ 50% preventable.

**Table 6:** Number of cases reviewed and preventability ratings assigned by each reviewer.

| Preventability rating assigned by the reviewer on a scale from 0 - 100 | Reviewer (Number of cases rated) | | | | | | | | | | | | | TOTAL |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **1** (480) | **2** (120) | **3** (120) | **4** (120) | **5** (120) | **6** (120) | **7** (120) | **8** (120) | **9** (120) | **10** (120) | **11** (120) | **12** (120) | **13** (120) | **1920** |
| | **Number of cases assigned rating (%)** | | | | | | | | | | | | | |
| **0-10** | 455 (94.8) | 108 (90.0) | 108 (90.0) | 77 (64.2) | 102 (85.0) | 97 (80.8) | 110 (91.7) | 93 (77.5) | 112 (93.3) | 114 (95.0) | 95 (79.2) | 107 (89.2) | 97 (80.8) | 1675 (79.4) |
| **11-50** | 11 (2.3) | 11 (9.2) | 11 (9.2) | 35 (29.2) | 18 (15.0) | 16 (13.3) | 10 (8.3) | 27 (22.5) | 3 (2.5) | 3 (2.5) | 22 (18.3) | 12 (10.0) | 15 (12.5) | 194 (11.0) |
| **51-90** | 11 (2.3) | 1 (0.8) | 0 (0.0) | 4 (3.3) | 0 (0.0) | 7 (5.8) | 0 (0.0) | 0 (0.0) | 3 (2.5) | 3 (2.5) | 3 (2.5) | 0 (0.0) | 8 (6.7) | 40 (1.9) |
| **>90** | 3 (0.6) | 0 (0.0) | 1 (0.8) | 4 (3.3) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 0 (0.0) | 2 (1.7) | 0 (0.0) | 0 (0.0) | 1 (s0.8) | 0 (0.0) | 11 (0.5) |

The intra-class correlation coefficient (ICC) was low at 0.14 for a single reviewer and much better at 0.68 for any random set of 4 reviewers. Reviewers' ratings of the 10 fictitious cases created to be highly preventable or non-preventable are presented in table 7. The cases are presented in appendix 3. There was more agreement between reviewers about non-preventable deaths than about preventable deaths. Only case 3 in the table was an act of commission resulting in death and there was excellent agreement between reviewers. Agreement was not a good for cases where an act of omission resulted in death.

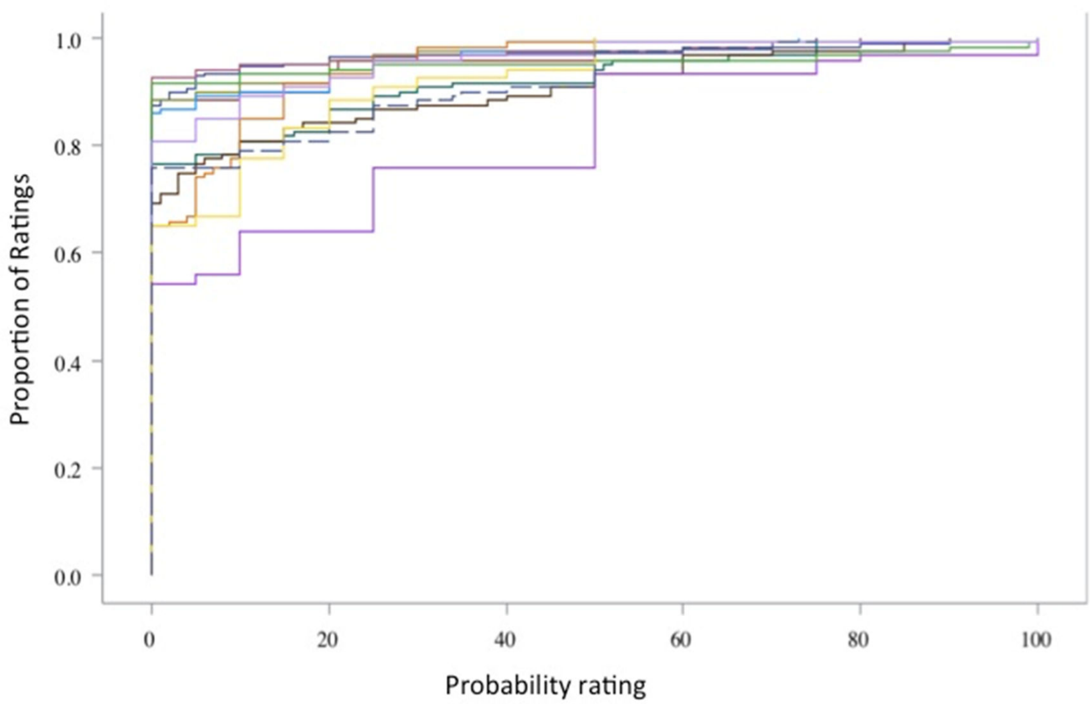



**Figure 2:** Proportion of preventability ratings assigned to each probability by 13 reviewers

**Table 7**: Preventability ratings for 10 fictitious cases created to be highly preventable or highly non-preventable

| Reviewer | Preventable Death | Reviewer | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
| 1 | Yes | 86 | 50 | 50 | 75 | 95 | 95 | 50 | 50 | 1 | 60 | 60 | 50 | 0 |
| 2 | Yes | 50 | 50 | 25 | 50 | 90 | 50 | 30 | 80 | 90 | 30 | 50 | 35 | 5 |
| 3 | Yes | 100 | 100 | 100 | 100 | 95 | 100 | 90 | 100 | 100 | 100 | 100 | 100 | 95 |
| 4 | Yes | 68 | 40 | 25 | 50 | 70 | 80 | 50 | 80 | 0 | 60 | 75 | 50 | 35 |
| 5 | Yes | 60 | 35 | 10 | 75 | 75 | 0 | 0 | 10 | 50 | 10 | 50 | 5 | 30 |
| 6 | No | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 7 | No | 0 | 0 | 0 | 25 | 15 | 0 | 0 | 0 | 0 | 0 | 0 | 15 | 0 |
| 8 | No | 0 | 0 | 0 | 0 | 50 | 0 | 0 | 0 | 0 | 0 | 25 | 5 | 0 |
| 9 | No | 0 | 0 | 0 | 10 | 0 | 20 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 10 | No | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

### *3.4 Latent class Models and discussion of model fit*

The observed variable, probability-of-preventability is continuous but for analysis was changed into a categorical variable. Probability-of-preventability was categorized using 2, 3, 4, 5 and 6 intervals with equal ranges. This resulted in 5 different indicator variable datasets that differed in how the indicator variable was categorized. Each indicator variable dataset was used to create a 2-class model and a 3-class model. This resulted in 10 different models that are summarized in tables 8 and 9.

Arguably the most important characteristic of a good latent class model is clear delineation of latent classes that occurs when each class is homogeneous and there is separation between classes (55). Homogeneity and separation are summarized by mean posterior probability of class membership, with values close to 1 indicating good separation and homogeneity. Refer to section 2.16: Evaluation *of model fit and model selection* for an in depth explanation of this concept. Color-coding in tables 8 and 9 indicates the mean posterior probability of class membership. Green means that mean probability of class membership for those in the class is >90%, yellow means 80-90% and red means <80%.

Tables 8 and 9 can be difficult to interpret so I will explain in detail. Table 8 contains the results from 5 latent class models with each model having 2 latent classes. The indicator variable for each model is the probability-of-preventability ratings divided into different numbers of intervals with equal ranges. The first row in Table 8 is a model made with two-intervals (0-50% preventable and 51-100% preventable). The § in column 1 denotes that there was no significant difference between the model and the sample as explained in section 2.16; this means that the model fits the data well. Columns 3 and 4 present the prevalence of each latent class. The 2 classes in each model have been labeled “Non-preventable death” and “Possibly preventable death”. The label “Non-preventable death” was chosen because, as explained in section 3.6 and appendix 4, people in the “non-preventable death” class had a near 100% chance of being rated highly non-preventable by any single reviewer. In contrast, people in the “Possibly preventable death” class had only a 20-30% chance of being rated highly preventable by any single reviewer. A more detailed explanation of the models and labels can be found in section 3.6.

The meaning of each latent class was slightly different in each model and is explained in detail in subsequent text and in figures 3, 4 and appendix 4. The first row of table 8 contains the prevalence of each latent class for model 1. Reading from the left, 93.8% of the sample fit into the “non-preventable death” class while

6.2% of the sample fit into the possibly preventable death class. The standard error was 4.6%. The standard error is the standard deviation of the sampling distribution of the prevalence of each class. Both cells in the first line are shaded green meaning that individuals placed in a class have greater than 90% probability of being correctly classified. Using mean posterior probability of correct classification, the best 2-class models were number 1 and 4 while the best 3-class model (table 9) is number 8.

Based on a chi-squared test of absolute model fit, models 1, 2 and 7 fit the data well, while model 6 was rejected. Models 3, 4, 5, 8, 9 and 10 could not be evaluated with a test of absolute model fit because the distribution of the test statistic is unknown. Using AIC and BIC to evaluate relative fit, the 2-class models fit the data better than the 3-class models when 2, 3 or 4 intervals were used for probability-of-preventability (model 1-3). When the probability-of-preventability was divided into 5 or 6 intervals (models 9 and 10), AIC favored the 3-class model and BIC favored the 2-class model. Model fit statistics can be found in Appendix 4. In summary multiple models fit the data well. Model 1 was the most parsimonious 2-class model while models 8 or 9 were the most parsimonious 3-class models depending on the criteria being used. Models 1 and 8 will be explained in detail and used for further analysis.

**Table 8**: Prevalence of each latent class when the probability of preventability is divided into different numbers of intervals with equal ranges. All models have two latent classes.

| | | Proportion of entire sample in each class | |
|---|---|---|---|
| **Model** | **Number of intervals used for the indicator variable** | **Prevalence of NON-PREVENTABLE DEATH class (SE)** | **Prevalence of POSSIBLY PREVENTABLE DEATH class (SE)** |
| 1 § | Two | 0.9383 (0.0461) | 0.0617 (0.0461) |
| 2 § | Three | 0.9208 (0.0180) | 0.0792 (0.0180) |
| 3 ¶ | Four | 0.9164 (0.0163) | 0.0836 (0.0163) |
| 4 ¶ | Five | 0.9194 (0.0157) | 0.0805 (0.0157) |
| 5 ¶ | Six | 0.8539 (0.0208) | 0.1461 (0.0208) |

§ Test of absolute model fit finds no significant ($P < 0.05$) difference between the study population and the model. ¶ No test of absolute model fit is possible. Mean probability of class membership for individuals in the class is: < 80% → [red], 80-90% → [yellow], >90% → [green]

**Table 9:** Prevalence of each latent class when the probability of preventability is divided into different numbers of intervals wit equal ranges. All models have three latent classes.

| | | Proportion of entire sample in each class | | |
|---|---|---|---|---|
| **Model** | **Number of Intervals used for indicator variable** | **Non-Preventable (SE)** | **Probably not preventable (SE)** | **Possibly preventable (SE)** |
| 6 | Two* | 0.7201 (**) | 0.2183 (**) | 0.0616 (**) |
| 7 | Three§ | 0.6358 (0.0497) | 0.3158 (0.0612) | 0.0484 (0.0212) |
| 8 | Four¶ | 0.8462 (0.0239) | 0.1424 (0.0244) | 0.0114 (0.0063) |
| 9 | Five¶ | 0.8828 (0.0199) | 0.0900 (0.0195) | 0.0272 (0.0094) |
| 10 | Six¶ | 0.6677 (0.0317) | 0.2831 (0.0333) | 0.0492 (0.0129) |

* Test of absolute model fit shows a significant difference (p <0.05) between the study population and the model, model is therefore is rejected. § Test of absolute model fit finds no significant difference between the study population and the model (P >0.05). ¶ No test of absolute model fit is possible. Standard error in (). ** Standard errors could not be calculated. Mean probability of class membership for individuals in the class is: < 80% → [red], 80-90% → [yellow], >90% → [green]

*3.5 Comparison of Class Prevalence between models*

In each model, the label "possibly preventable death" has been applied to the latent class with the highest probability of preventability; however, the probability of a patient receiving a high rating is somewhat different in each model. For example in model 1 individuals in the "possible preventable death" class had a **31%** probability of receiving a rating of "greater than 50% preventable"; in contrast, individuals with model 2 who were in the "possibly preventable death" class had a

**22%** probability of being rated greater than 67% preventable (Figure 3 and appendix 4). Despite these differences in the meaning of the "possibly preventable death" class in each model, they all shared the same label. Because of the differences, the prevalence of the "possibly preventable death" class cannot be directly compared between models. Although the meaning of being in the "possibly preventable death" class was not the same in each model, the prevalence was similar among all 2-class models except model 5, where individuals have less than 10% probability of receiving a rating greater than 50 % (Figures 3, 4 and appendix 4). Model 5 also had a much higher prevalence of the possibly preventable death class that the other models. In the 2 and 3-class models, the prevalence of the "possibly preventable death" class ranged from 6.2% - 14.6% and 1.1% - 4.9% respectively, excluding models that were rejected because of poor fit.

*3.6 Interpretation of models*

The 2-class model and 3-class model (models 1 and 8) with the best separation and homogeneity are presented in figures 3 and 4 respectively and are explained in the following text. These models will be referred to as the 2-class model and the 3-class model, respectively. Figures for the other 8 models and fit statistics for each model can be found in Appendix 4.

I will go through figure 3 in detail to explain how it should be interpreted. Each line represents a different latent class. The X-axis is the probability category assigned by the reviewer. The Y-axis is the probability that an individual reviewer will assign a patient to a probability category. For example, a person in the latent class represented by the red line in figure 3 has a 99.2% probability of being rated

0-50% preventable by each reviewer and, conversely, a 0.8% probability of being rated 51-100% preventable. Because individuals in this latent class have probabilities of being assigned each probability category close to 0 or 1, the class is very homogeneous. Also, “non-preventable deaths” is an appropriate label for these individuals because they almost always receive preventability ratings of 0-50% and almost never 50-100%.

In the 3-class model, individuals in the latent class labeled “non-preventable death” had a 98.0% probability of receiving the lowest probability rating and a 0.1% probability of receiving the highest preventability rating (figure 4). These values were very close to 1 and 0 meaning that, similar to the 2-class model, the “non-preventable death” class is highly homogeneous and these cases were very likely to be non-preventable deaths.

The latent class labeled “possibly preventable death” is not as homogeneous in either the 2 or 3-class models. In the 2-class model, individuals in the “possibly preventable death” class had a 31.0% chance of being rated greater than 50% preventable and a 69% chance of being rated less than 50% preventable by any single reviewer. The 3-class model was slightly more homogeneous with individuals in the “possibly preventable death” class having only an 11.2% chance of being rated less than 25% preventable by any single reviewer and a 36.0% of receiving the highest preventability rating (greater than 75% preventable) from each reviewer.

In the 3-class model the red and green lines have very similar shapes. The class represented by the green line has been labeled “probably non-preventable” because individuals in the class are likely to receive low preventability ratings (0-25%) but not to the same degree as individuals in the non-preventable death class.

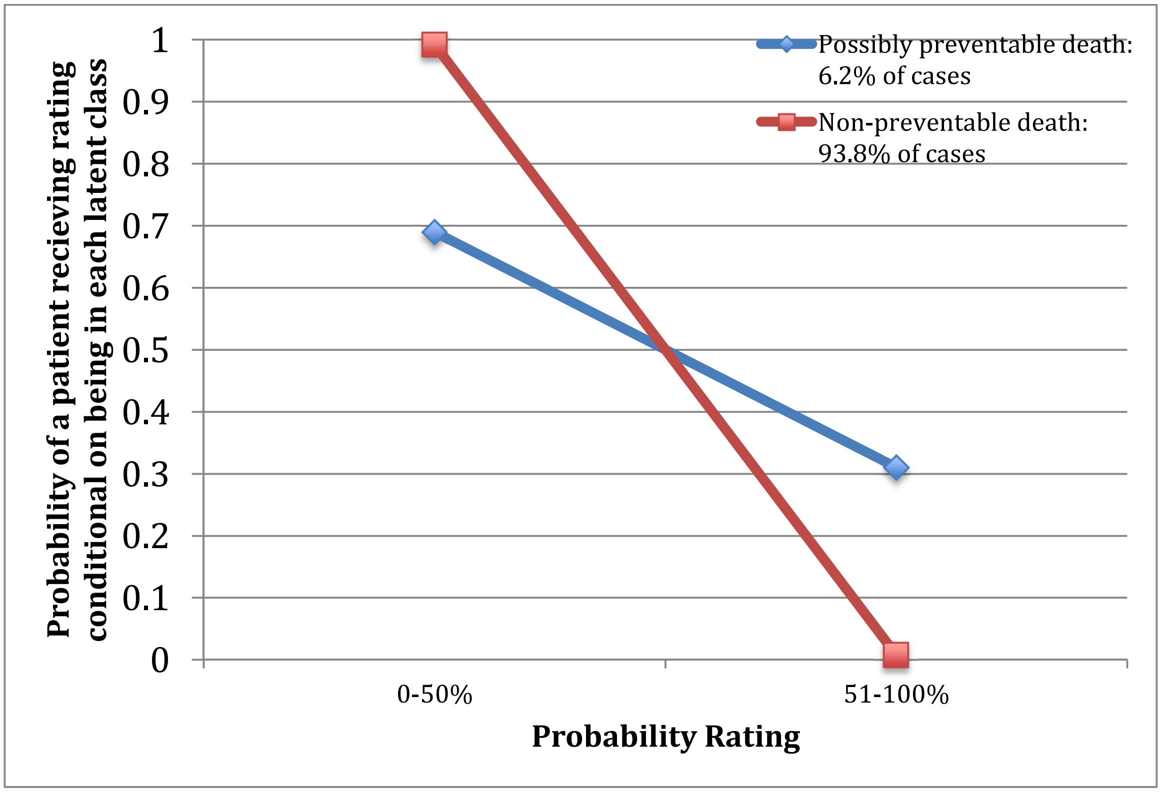



**Figure 3:** Two-class model with preventability ratings in 2-intervals with equal ranges. Probability of receiving rating conditional on being in a latent class. Legend contains class label and class prevalence.

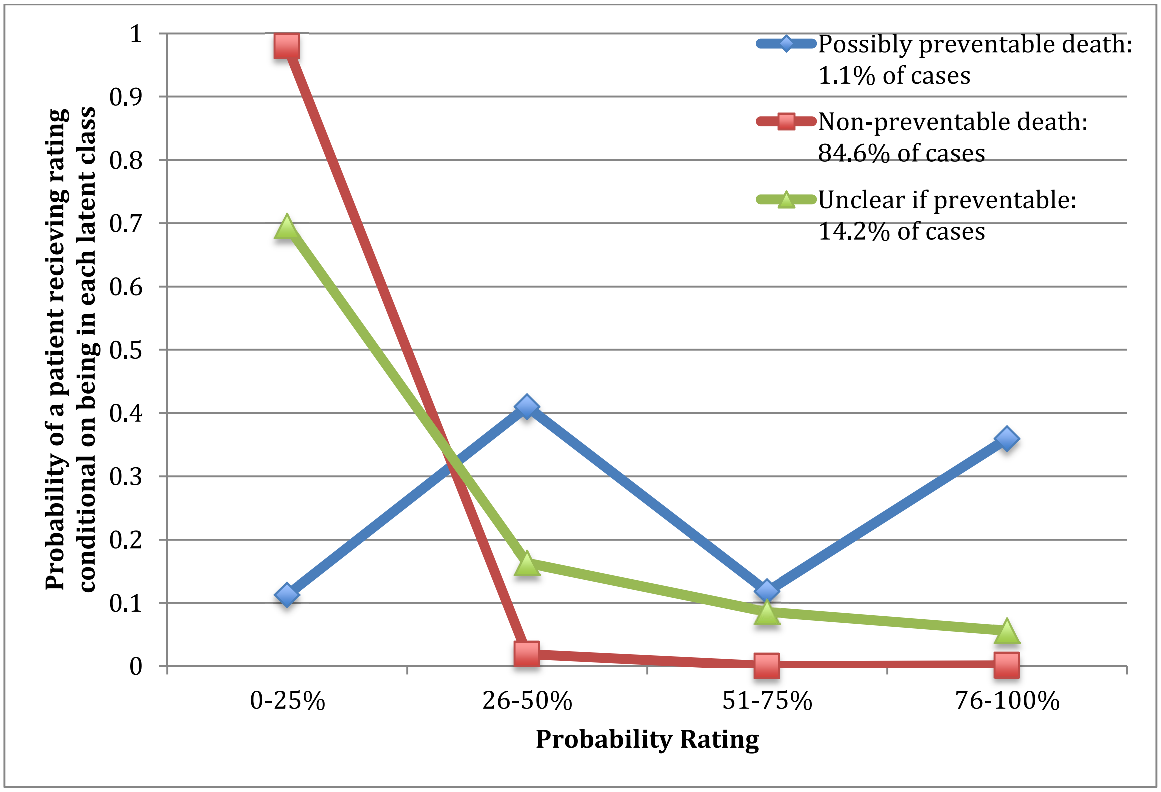



**Figure 4:** Probability of receiving rating conditional on being in each latent class. Three-class-model with probabilities-of-preventability in 4 intervals with equal ranges. Legend contains class labels and class prevalence.

### *3.7 Sensitivity Analysis*

To test if the outlying reviewer, described in section 3.3, affected the results meaningfully, this person's preventability ratings were changed to missing and the analysis was repeated (**tables 10 and 11**). The LCA procedure in SAS assumes that any missing data is missing at random, which is a reasonable assumption for this analysis because each reviewer was assigned a random set of cases. Most models fit better when the outlier was excluded and the mean posterior probability of class membership was generally higher, especially in the 3-class models. The exception to this was model 1, the parsimonious model from the main analysis, which did not fit as well. The prevalence of the "possibly preventable death" class was lower in all models when the outlier was excluded. The range in prevalence of the "possibly preventable death" class in the 2 and 3-class models is 4.1-12.0% and 3.0 – 3.7% respectively when models with poor fit are excluded.

**Table 10:** Prevalence of each latent class when the probability of preventability is divided into different numbers of intervals with equal range. Data excludes the one reviewer with anomalous ratings. All models have two latent classes.

| | **Proportion of entire sample in each class** | |
|---|---|---|
| **Number of Intervals used for indicator variable** | **Non-Preventable (SE)** | **Possibly preventable (SE)** |
| Two* | 0.9259 (0.0767) | 0.0741 (0.0767) |
| Three§ | 0.9540 (0.0131) | 0.0460 (0.0131) |
| Four¶ | 0.9398 (0.0141) | 0.0602 (0.0141) |
| Five¶ | 0.9417 (0.0134) | 0.0583 (0.0134) |
| Six¶ | 0.8796 (0.0190) | 0.1204 (0.0190) |

* Test of absolute model fit shows a significant difference ($p < 0.05$) between the study population and the model, model is therefore is rejected. § Test of absolute model fit finds no significant difference between the study population and the model ($P > 0.05$). ¶ No test of absolute model fit is possible. Standard error in (). ** Standard errors could not be calculated. Mean probability of class membership for individuals in the class is: < 80% → [red], 80-90% → [yellow], >90% → [green].

**Table 11**: Prevalence of each latent class when the probability of preventability is divided into different numbers of intervals with equal range. Data excludes the one reviewer with anomalous ratings. All models have three latent classes.

| | Proportion of entire sample in each class | | |
|---|---|---|---|
| **Number of intervals used for indicator variable** | **Non-Preventable (SE)** | **Probably not preventable (SE)** | **Possibly preventable death (SE)** |
| Two* | 0.5021 (**) | 0.4152 (**) | 0.0827 (**) |
| Three§ | 0.7135 (**) | 0.2567 (**) | 0.0298 (**) |
| Four¶ | 0.7316 (0.0365) | 0.2317 (0.0373) | 0.0367 (0.0108) |
| Five¶ | 0.9411 (0.0134) | 0.0309 (0.0103) | 0.0280 (0.0090) |
| Six¶ | 0.7872 (0.0275) | 0.1771 (0.0273) | 0.0357 (0.0096) |

* Test of absolute model fit shows a significant difference (p <0.05) between the study population and the model, model is therefore is rejected. § Test of absolute model fit finds no significant difference between the study population and the model (P >0.05). ¶ No test of absolute model fit is possible. Standard error in (). ** Standard errors could not be calculated. Mean probability of class membership for individuals in the class is: < 80% → [red], 80-90% → [yellow], >90% → [green].

*3.8 Post-hoc analysis with different cut offs*

The probability of preventability was categorized into 3 groups, ≤10%, 11-89% and ≥ 90%. The 2-class model made with these cut offs had poor homogeneity and separation and is not shown. In the 3-class model, the "preventable death" class had better separation and homogeneity than any previous model but the "unclear if preventable" class had very poor separation and homogeneity (Figure 5). It is notable that this model had better separation of the preventable class than any previous model and also had the lowest prevalence of this class (0.8%). Put another way, the more certain we are that the "preventable death" class contains only preventable deaths, the lower the prevalence of the preventable death class. It is also interesting that in this model, there is only a 41.3% chance that a person in the "possibly preventable death" class will receive a preventability rating of 90-100%.

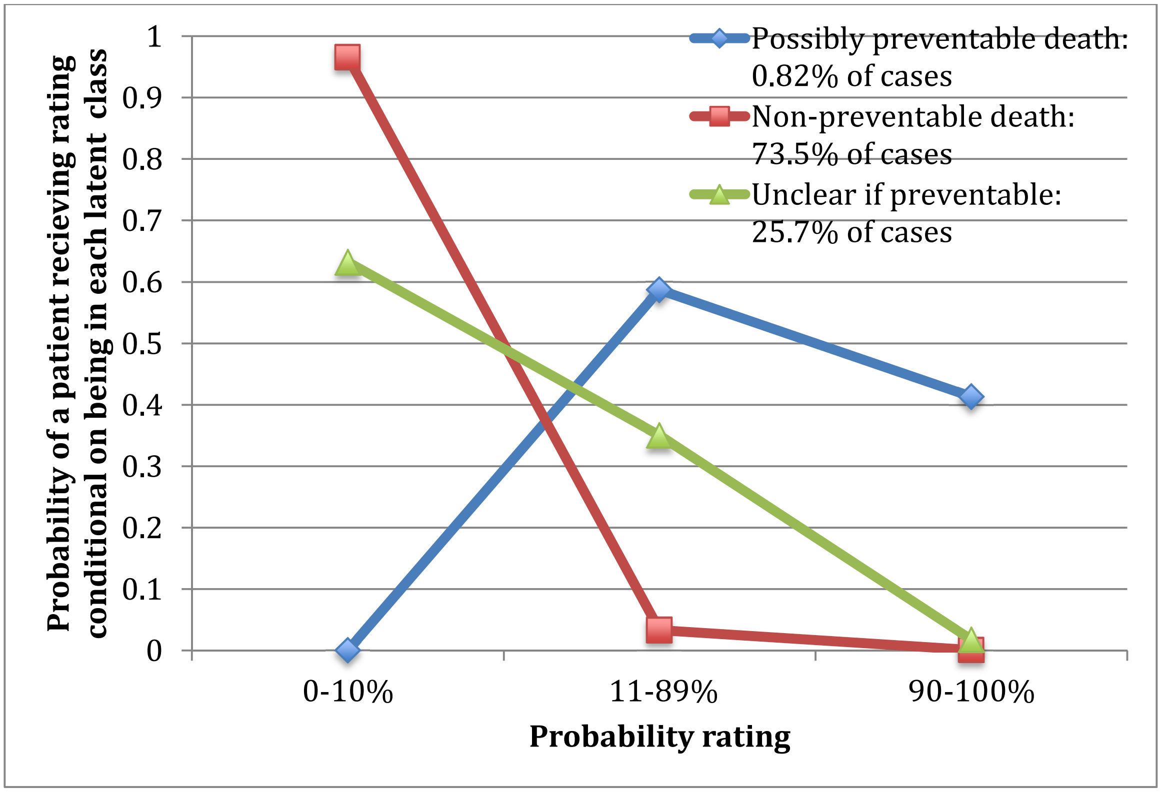



**Figure 5:** Probability of receiving rating conditional on being in each latent class. Three-class-model with probabilities-of-preventability in 3 intervals. Legend contains class labels and class prevalence.

### *3.9 Characteristics of the preventable deaths*

The 2-class model created with dichotomized preventability ratings (model 1) was used to categorize individuals in the population into non-preventable deaths and possibly preventable deaths. The characteristics of individuals who best fit into each category are presented in **Table 12**. Even though model 1 estimates the prevalence the "possibly preventable death" latent class to be 6.2%, only 11 of the 480 deaths (2.3%) are best classified as "possibly preventable deaths". This incongruence occurred because the model could not classify some individuals with certainty. When there is uncertainty about which class an individual belongs to the model will add a fraction to each class. When the model is later used to classify individuals into their "best" class they will be placed into the class that they most likely belong even if it is a near tie. Many individuals fell into the "non-preventable death" class by a small margin. Consequently only 2.3% of the sample was **best** classified as "possibly preventable deaths" even though the prevalence of the "possibly preventable death" class was 6.2%. Interestingly 2.3% is very similar to the prevalence of the "possibly preventable" class in the 3-class models in the sensitivity analysis. Out of the 11 patients with possibly preventable deaths, more than half the reviewers judged that the actions required to prevent the death had to occur exclusively before hospitalization in 4 cases; this means that only 7/480 (1.5%) of deaths were preventable by actions taken during the hospitalization. In 8 of 11 deaths deemed "possibly preventable", at least 2 out of 4 reviewers judged that the patient would probably or certainly be dead in 3 months even with the highest quality care. There were only 2/480 (0.42%) cases where the actions necessary to prevent death needed to occur in hospital and where the patient was judged likely to be alive in 3 months if they received the highest quality care. Considering the entire population at risk, the probability of a possibly preventable death of someone who is likely to be alive in 3 months by actions taken in hospital is 0.014%.

There are several notable differences between the non-preventable deaths and possibly preventable deaths presented in **Table 12**. Overall, individuals with potentially preventable deaths appeared to be healthier; they had fewer inpatient encounters and inpatient days in the previous 6-months; they had lower probability of death in hospital as calculated by their Escobar score (49); and had lower Elixhauser comorbidity scores than those with non-preventable deaths. Individuals with possibly preventable deaths had longer lengths of stay during their index admission and were more often admitted under general surgery.

**Table 13** summarizes the 11 preventable deaths and the actions that reviewers thought would have prevented them. The actions needed to prevent the deaths were diverse; they included better assessment of operative risk, failure of communication on discharge, errors in technical aspects of surgery and delayed surgery. The actions needed to prevent death were a combination of pre-hospital and during hospitalization acts of commission and omission.

**Table 12:** Characteristics of patients with preventable and non-preventable deaths

| Variable | **Non-preventable Death n=469** | **Possibly Preventable Death n=11** | **Total n=480** |
|---|---|---|---|
| Mean Age at admission ± SD | 73.6± 10.5 | 73.9±16.2 | 73.9±16.1 |
| **Campus** | | | |
| General | 247 (52.7) | 9 (81.8) | 256 (53.3) |
| Civic | 174 (37.1) | 2 (18.2) | 176 (36.7) |
| Heart Institute | 48 (10.2) | 0 (0) | 48 (10.0) |
| **Language** | | | |
| English | 399 (85.1) | 8 (72.7) | 407 (84.4) |
| French | 48 (10.2) | 3 (27.3) | 52 (10.8) |
| Other | 22 (4.7) | 0 (0) | 22 (4.6) |
| | | | |
| **Number of inpatient encounters in the previous 6 months** | | | |
| 0 | 298 (63.5) | 9 (81.8) | 307 (64.0) |
| 1 | 98 (20.9) | 1 (9.1) | 99 (20.6) |
| 2 | 39 (8.3) | 0 (0) | 39 (8.1) |
| 3 | 22(4.7) | 0 (0) | 33 (6.9) |
| 4 | 6 (1.3) | 1 (9.1) | 7 (1.5) |
| ≥ 5 | 6 (1.3) | 0 (0) | 6 (1.3) |
| **Inpatient days in the last 6 months** | | | |
| 0 | 298 (63.5) | 9 (81.8) | 307 (64.0) |
| 1-5 | 34 (7.2) | 0 (0) | 34 (7.1) |
| 6-10 | 29 (6.2) | 1 (9.1) | 30 (6.3) |
| 11-15 | 28 (6.0) | 1 (9.1) | 29 (6.0) |
| 16-20 | 20 (4.3) | 0 (0.0) | 20 (4.2) |
| 21-25 | 11 (2.3) | 0 (0.0) | 11 (2.3) |
| >25 | 49 (10.4) | 0 (0.0) | 49 (10.2) |
| **Total length of stay (days)** | | | |
| 1-5 | 204 (43.5) | 3 (27.3) | 207 (43.1) |
| 6-10 | 84 (17.9) | 2 (18.2) | 86 (17.9) |
| 11-15 | 53 (11.3) | 2 (18.2) | 55 (11.5) |
| 16-20 | 41 (8.7) | 1 (9.1) | 42 (8.8) |
| 21-25 | 21 (4.5) | 0 (0.0) | 21 (4.4) |
| >25 | 66 (14.1) | 3 (27.3) | 69 (14.4) |
| **Admitting Hospital** | | | |

| | | | |
|---|---|---|---|
| **Service** | | | |
| General medicine | 159 (33.9) | 2 (18.2) | 161 (33.5) |
| Intensive care | 69 (14.7) | 2 (18.2) | 71 (14.8) |
| Cardiology | 47 (10.2) | 0 (0.0) | 47 (9.8) |
| Medical Oncology | 38 (8.1) | 1 (9.1) | 39 (8.1) |
| General Surgery | 17 (3.6) | 4 (36.4) | 21 (4.4) |
| Neurology | 16 (3.4) | 0 (0.0) | 16 (3.3) |
| Radiation oncology | 15 (3.2) | 0 (0.0) | 15 (3.1) |
| Family Medicine | 14 (3.0) | 0 (0.0) | 14 (2.9) |
| Malignant Hematology | 14 (3.0) | 0 (0.0) | 14 (2.9) |
| Orthopedics | 11 (2.4) | 1 (9.1) | 12 (2.5) |
| Other | 69 | 1 | 70 |
| **Discharge Hospital Service** | | | |
| General medicine | 134 (28.6) | 2 (18.2) | 136 (28.3) |
| Intensive care | 119 (25.4) | 7 (63.6) | 126 (26.3) |
| Cardiology | 42 (9.0) | 0 (0.0) | 42 (8.8) |
| Medical Oncology | 36 (7.7) | 0 (0.0) | 36 (7.5) |
| Neurology | 14 (3.0) | 0 (0.0) | 14 (2.9) |
| Radiation oncology | 12 (2.6) | 0 (0.0) | 12 (2.5) |
| General Surgery | 11 (2.3) | 1 (9.1) | 12 (2.5) |
| Malignant Hematology | 12 (2.6) | 0 (0.0) | 12 (2.5) |
| Family Medicine | 10 (2.1) | 0 (0.0) | 10 (2.1) |
| Other | 79 (16.8) | 1 (9.1) | 80 (16.7) |
| | | | |
| **Patient has family physician** | 423 (90.2) | 11 (100) | 434 (90.4) |
| **Patient seen by palliative care during the admission** | 137 (29.2) | 2 (18.2) | 139 (29.0) |
| Mean Adjusted Risk of death in hospital ± SD (49) | 0.31± 0.20 | 0.13± 0.16 | 0.31± 0.20 |
| Mean Elixhauser score ± SD | 12.56 ± 11.74 | 6.64 ± 3.96 | 12.42± 8.98 |
| ≥ 2/4 reviewers judged that the patient would likely be dead in 3 months | 467 (99.6) | 8 (72.7) | 475 (99.0) |

**Table 13**: Description of the cases that model 1 classified as "possibly preventable"

| Case | Probability of preventability ratings | | | | Admitting Service | Age | Total Length of stay (days) | Description of case | Action that might have prevented the death |
|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | | | | | |
| **1§** | 5 | 75 | 90 | 15 | INTENSIV E CARE | ≥ 65 | 1 | Patient was transferred from a community hospital where they had a hemi-colectomy 1 month ago to resect a colon cancer. An anastomotic leak resulted in multiple abscesses. Patient was transferred for interventional radiology to place drains. During the procedure the patient had decreased LOC and became hypoxemic. It was thought to be due to sedation given for the procedure but level of consciousness did not improve with Narcan and flumazenil. The patient was intubated and transferred to ICU. A family meeting was held and the family asked for comfort care only. | There should be standard training and procedures for the use of sedation during interventional radiology procedures. An anaesthetist should be involved when needed. |
| **2*** | 70 | 20 | 38 | 75 | ONCOLO GY | ≥ 65 | 3 | A patient with small cell lung cancer who received a first round of chemotherapy 11 days ago, presented to a peripheral hospital with febrile neutropenia, hypotension and tachycardia. They received a bolus of IV fluids and were transferred to our hospital for management. They were hypotensive on arrival (BP 84/60) but waited 6 hours after their arrival before a physician assessed them. The first antibiotics were administered 6 hours after arrival. The patient continued to be hypotensive for the next 48 hours and received only small fluid boluses. 100cc/hour of NS. ICU was then consulted emergently after 48 hours. The patient was transferred to ICU and died shortly after with no further resuscitation attempted because of very poor prognosis. | Earlier antibiotics and IV fluids may have prevented death. |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **3** | 10 | 100 | 85 | 0 | GENERAL MEDICINE | > 65 | 14 | The patient was admitted with falls, hyponatremia, 50lbs of weight loss and generalized weakness. The patient then developed pneumonia and was treated. They aspirated and became hypoxemic but recovered with in 24 hours. They then developed Clostridium difficile colitis and were given Flagyl. On the final day of the hospitalization he began complaining of R shoulder pain. He had rapid atrial fibrilation, hypotension decreased level of consciousness and died with in 24 hours. | The neck pain, hypotension and tachycardia should have been investigated and treated. The story is concerning for a perforated viscous. It is unlikely that proper treatment was administered considering no investigations were done. |
| **4** | 100 | 73 | 70 | 10 | GEN. SURGER Y | > 65 | 34 | The patient was seen in cancer assessment clinic after being found to have a large rectal mass with rectal bleeding. A biopsy was done via sigmoidoscopy to confirm a diagnosis of colon cancer. The patient went to surgery to have the mass resected. A "misfire" of the stapler occurred during surgery. The staple went through the rectal wall. The patient had an anastomotic leak and was brought back to OR for repair. The pathology report for the resected specimen came back showing no evidence of cancer. The patient then developed a wound infection. They then deteriorated further and goals of care changed to comfort only. | If the staple had not misfired the patient may not have died. The surgery failed to remove the cancer even though the surgeon thought they had. |
| **5§** | 80 | 0 | 20 | 75 | NEUROS URGERY | > 65 | 11 | Patient with NASH cirrhosis with severe portal hypertension that had been treated with a TIPPS procedure previously. Now admitted for elective carotid endarterectomy because of recurrent TIAs and severe R carotid stenosis. Extensive pre-operative consultation occurred with the conclusion that she was low risk for complications. ICU was consulted urgently post-operatively for decreased LOC and increased troponins. Patient was treated for SBP and NSTEMI. The patient required intubation. Abdominal XR showed intestinal pneumatosis. General surgery thought the patient was too high risk to operate on and patient died shortly after. | The patient was labeled as low risk but was in fact high risk. None of the pre-operative assessments mentioned cirrhosis as a factor that increased surgical risk. Hepatic encephalopathy was not managed appropriately. |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **6*§** | 100 | 50 | 50 | 100 | GENERAL MEDICINE | < 65 | 2 | The patient was admitted to hospital for hypercapnia secondary to narcotic administration and UTI. They were discharge back to their nursing home but returned to hospital 24 hours later with hypercapnic respiratory failure similar to the previous admission. The patient had received opiates at her nursing home despite not being on any opiates while in hospital, the cause for the previous admission was thought to be opiate use also. The patient was placed on BiPAP but this was discontinued because of decreased level of consciousness making it unsafe. The patient had previously expressed they did not want intubation and so care was directed at comfort only. | Communication with the nursing home telling them not administer opioids may have prevented the death. |
| **7*** | 80 | 100 | 60 | 50 | GEN. SURGERY | ≥ 65 | 44 | The patient was admitted with a high-grade small bowel obstruction. Multiple trials of conservative treatment with fluids and food was not successful. On day 8 of the admission the patient's bowels perforated. They were was brought to OR emergently. Post op they were admitted to ICU for peritonitis and septic shock. They were in ICU for several weeks with ongoing sepsis and acute kidney injury requiring dialysis. Eventually family asked for comfort care only. | Earlier surgical intervention may have prevented the death. |
| **8§** | 1 | 25 | 60 | 52 | INTENSIVE CARE | ≥ 65 | 10 | The patient was seen 2 weeks ago at a different hospital and had a foley catheter inserted for urinary retention. They were sent home with the foley. The patient presented to hospital shortly after with abdominal pain, urosepsis and shock. They were treated with IV fluids, antibiotics, vasopressors and required intubation. A CT revealed a ureteric stone and a stent was placed. Vancomycin was given on admission but ongoing dosing was not started until 48 hours later when urine cultures were reported as growing enterococcus. The intensive care team discussed a one-way extubation with the family. After extubation the patient did well initially but then had worsening respiratory distress and goals of care were changed to comfort only. | Investigation of the urinary retention on initial presentation with ultrasound. Continuation of broad-spectrum antibiotics until urine culture results were complete may have prevented the death. |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **9** | 0 | 30 | 70 | 80 | ORTHOP AEDICS | ≥ 65 | 6 | The patient was transferred from another hospital for closed reduction of a chronic anterior shoulder dislocation. There was no documentation of assessment of pre-operative risk. Post operatively the patient had decrease oxygen saturation, fever and elevated INR. They were diagnosed with pneumonia, UTI, and decompensated heart failure with pulmonary edema. Over the next 5 days they had decreased oxygen saturation. ICU was consulted urgently the following day for respiratory failure. The patient decided against intubation so treatment was directed at symptoms only. | The operation should have been delayed until the patient was better. The operative risk was not adequately assessed. |
| **10** | 40 | 80 | 75 | 0 | GEN. SURGER Y | < 65 | 19 | The patient was admitted to ICU post resection of retroperitoneal sarcoma including L nephrectomy, partial gastrectomy, splenectomy , partial L diaphragm resection , L adrenalectomy. On POD 9 while still in ICU the chest tube was removed. Later that day the patient complained on feeling unwell and their heart rate slowed and stopped. Code-blue activated and return of spontaneous circulation was achieved. They Patient then went to OR urgently because CXR revealed a white out of the L hemi thorax. In the OR a rupture of the thoracic aorta of unknown cause was found and repaired. The patient returned to ICU but did not recover neurologically because of anoxic brain injury sustained during the cardiac arrest. The goals of care were changed to comfort only. | Daily CXR to confirm correct placement of chest tube with no migration and CXR immediately after removal. Unclear if a competent staff supervised the removal. |
| **11** | 20 | 100 | 100 | 99 | GEN. SURGER Y | ≥ 65 | 30 | The patient sustained a colon perforation during a colonoscopy and went to the operating room for a right hemi-colectomy. On year later they were admitted for elective loop ileostomy reversal. On POD 4 the patient developed pneumonia and ileus. POD 8 the patient developed AKI and an anastomotic leak. Multiple drains were placed and the patient was started on micafungin and meropenem. A one-way extubation was done on POD 30 and the patient died later that day. | The elective procedure should not have been performed. The risk-benefit ratio was not adequately assessed. |

* Denotes cases where ≥ 2/4 reviewers thought the patient would be dead in 3-months even with the highest quality of care.

§ Denotes cases where ≥ 2/4 thought that actions necessary to prevent the death were entirely prior to hospitalization

*3.10 Reviewer Sensitivity and Specificity*

Using the classifications from model 1, the sensitivity and specificity of each reviewer to detect a "possibly preventable death" was calculated using a rating of greater than 50% to determine if a reviewer detected a preventable death (**Table 14).** Mean reviewer specificity was 0.95. Reviewer 4 had the lowest specificity (0.78) and was the outlier excluded in the sensitivity analysis. Mean reviewer sensitivity was 0.63 with large variability between reviewers. The variability is due to low numbers of true positives in each reviewer's sample.

**Table 14**: Sensitivity and specificity to detect a death classified as possibly preventable by the model.

| Reviewer | True Positives assigned | True negatives assigned | Number of true positives in reviewer's sample | Number of true negatives in reviewer's sample | Sensitivity | Specificity |
|---|---|---|---|---|---|---|
| **1** | 5 | 460 | 11 | 469 | 0.45 | 0.98 |
| **2** | 1 | 115 | 3 | 117 | 0.33 | 0.98 |
| **3** | 1 | 114 | 2 | 118 | 0.50 | 0.97 |
| **4** | 4 | 91 | 4 | 116 | 1.00 | 0.78 |
| **5** | 1 | 139 | 2 | 139 | 0.50 | 1.00 |
| **6** | 4 | 110 | 4 | 116 | 1.00 | 0.95 |
| **7** | 1 | 116 | 2 | 118 | 0.00 | 0.98 |
| **8** | 0 | 112 | 1 | 119 | 0.00 | 0.94 |
| **9** | 2 | 112 | 4 | 116 | 0.50 | 0.97 |
| **10** | 1 | 117 | 1 | 119 | 1.00 | 0.98 |
| **11** | 2 | 108 | 3 | 117 | 0.67 | 0.92 |
| **12** | 2 | 116 | 2 | 118 | 1.00 | 0.98 |
| **13** | 4 | 108 | 5 | 115 | 0.80 | 0.94 |
| | | | | **Mean** | **0.63** | **0.95** |

# Discussion

## *4.1 Summary of Findings*

Thirteen reviewers gave 1920 preventability ratings to 480 consecutive deaths. We then used latent class analysis to determine the natural groupings of patients according to preventability ratings. When deaths were categorized into 2-classes, 6.2% of deaths were "possibly preventable". The precise meaning of "possibly preventable" in this model was that they had a 31.0% probability of being rated greater than 50% preventable *by any one reviewer*. The other 93.9% of the deaths were "non-preventable" and had only a 0.8% chance of being rated greater than 50% preventable by any one reviewer. When deaths were categorized into 3-classes 1.1% of deaths were classified as "possibly preventable". The other 98.9% of deaths fell into categories best described as "probably not preventable deaths" (14.2%) and "non-preventable deaths" (84.6%). Sensitivity analysis that excluded one anomalous reviewer did not substantially change the results from the 2-class models but resulted in a better fit of all 3-class models and a consistent finding that 3% of deaths were possibly preventable. When the 2-class model was applied to the data, 11 of 480 (2.29%) deaths were best classified as "possibly preventable" and only 3 out of 480 (0.63%) were predicted to be alive in 3-months had they received the highest quality medical care. The main conclusion from these analyses is that between 0.6% and 6.2% of deaths are possibly preventable with high quality care. The precise number depends on the meaning of "preventable".

The analysis identified non-preventable deaths with certainty but could not be as definitive about preventable deaths. This is not surprising considering the poor inter-rater reliability, the poor sensitivity of reviewers, and the low number of preventability ratings greater than 50%. We can confidently say that the prevalence of *truly* preventable deaths is definitely less than the 6% that are deemed *possibly* preventable. Models in which individuals in the "possibly preventable death" class had a higher probability of receiving a high preventability rating, had a lower prevalence of the "possibly preventable death" class. Put another way, the more confident we were that people in the "possibly preventable death" class were truly preventable deaths, the fewer preventable deaths we detect. One other notable finding is that very few people (0.4%) had preventable deaths and were expected to be alive in 3-months.

Putting these results in terms of preventable deaths per admission as reported in most previous studies on this topic, we found that 0.21% (30/14,290) of admissions resulted in a possibly preventable death while only 0.021% (3/14,290) of the sample could be classified as preventable deaths after adjusting for 3-month survival.

### *4.2 Comparison to previous measures of preventability*

The study by *Heyward et al.* in 2000 is the best measure of preventable death for the reasons expressed in the introduction (33). Interestingly the prevalence of 6.2% that we found is very close to the unadjusted prevalence of 6.0% that *Heyward et al.* found 15 years ago at a different center in a different country. Similarly our

0.625% prevalence of preventability in the sample after adjusting for 3-month survival was also very similar to the 0.5% found by *Hayward et al.* Our study's finding that 0.21% of admissions result in a "possibly preventable death" is at the low end, but in line with the multitude of studies that used the HMPS methodology (22, 24-26, 28, 29, 32). The major difference in our study is that we were explicit about the meaning of "possibly preventable deaths". In reality, all previous measurements of preventable deaths are actually "possibly preventable deaths" but the meaning in terms of probability of being rated preventable is not explicit in earlier works as it is in this thesis.

Compared to the study by *Dubois et al.* who found that 27% of deaths in hospital were possibly preventable, our measure is drastically lower (32). The reasons for the difference are unclear but it may have to do with bias in *Dubois'* study because they did not correct for poor inter-rater reliability and because they only sample certain diagnoses instead of all deaths. It is also possible that the true incidence of preventable deaths is lower at our institution than it was in the hospitals in *Dubois'* study.

### *4.3 Reviewer reliability*

Inter-rater reliability was poor with an intra-class correlation coefficient (ICC) of 0.14 for a single reviewer and 0.68 for any set of 4 reviewers. The reliability of reviewers was somewhat lower than that reported in other studies 0.2-0.4 (51). This may be because we used a 100-point scale instead of the 6 point scale used in most other studies or because of the outlying reviewer. A contributing factor is that

ICC, the preferred measure of reliability when a continuous measure is used, gives a less inflated measurement than kappa that is used in many other studies (52, 54, 60). The poor reliability is partially responsible for patients in the "probably preventable death" class having only a 31% chance of being rated preventable in model 1.

The specificity of reviewers was excellent (mean 95%), with exception of the outlier. From clinical experience, this is not surprising. There are many deaths that are clearly not preventable and reviewers can easily pick these out with certainty. Sensitivity on the other hand was poor (mean 63%). This highlights the difficulty in determining with certainty that a death is preventable. In our study, the uncertainty in judging preventability ultimately led to uncertainty in the "probably preventable death" class.

### *4.4 Characteristics of preventable deaths*

When the model was applied back to the data, only 11 out of 480 (2.3%) deaths fit best into the "possibly preventable" class despite the class prevalence being 6.2%. This incongruence occurred because a large number of deaths fit into neither class very well but fell into the "non-preventable" class by a small margin. This resulted in the estimated *class prevalence* being much larger than the number of individual deaths that fit the class best.

The model classified too few deaths as "possibly preventable" to draw strong conclusions about their characteristics. A few interesting associations were noted though. People with preventable death had fewer previous admissions, longer length of stay, lower baseline risk of death and lower Elixhauser comorbidity scores. It is somewhat intuitive that people who die unexpectedly will be healthier prior to the admission and then have a long admission with many opportunities for adverse events to occur. Length of stay is associated with more adverse events in nearly every study on the topic (3, 22, 28). The preventable deaths were most often admitted under general surgery. This is finding common to many AE studies and is probably because surgery has many risks that are at least partially attributable to a specific intervention (12, 26, 61).

### *4.5 Actions needed to prevent death*

The actions needed to prevent death were diverse but there are a couple repeated errors. There were 2 cases of inadequate assessment of pre-operative risk and 2 cases of inadequate recognition of sepsis. Both quality problems are well known with evidence-based practice guidelines (62-65). Root cause analysis is needed to understand precisely why best practice was not followed in these cases. Adverse events are common but only a small part of the population (11/480) was classified as "possibly preventable". This means there are many gaps in quality of care that were not picked up by this analysis. Although quality gaps that lead to preventable deaths are very important, focusing only on issues that lead to death misses important quality issues.

4.6 *Strengths*

The analysis in this thesis acknowledges the probabilistic nature of preventability ratings. Preventability is a continuous latent variable that can only be measured with error prone peer assessment methods. It is a false dichotomy to classify preventable deaths as preventable or not; in addition, there is no clear-cut point on the preventability scale at which a death should be categorized as preventable. Unlike nearly all-previous measures of preventable death, our study is explicit in the meaning of "preventable death" and acknowledges the probabilistic nature of preventability. Studies that use the HMPS methodology do not recognize or account for the uncertainty in preventability ratings. The studies that use HMPS methodology define AEs as cases in which 2 out of 3 reviewers assigned a rating of greater than 4 out of 6; importantly, this may have occurred by chance because of low inter-rater reliability. In the HMPS methodology, the uncertainty of the preventability ratings is not formally acknowledged and therefore leads to over estimation of preventable deaths. The bias caused by the use of a suboptimal test to measure a rare condition is made transparent by our analysis.

Our study has numerous other strengths. We used 480 consecutive deaths instead of a purposive sample thereby eliminating error introduced when correcting for sampling frame. Our study had 4 reviewers per case, which is more than any other study that aimed to measure preventable death rate. Reviewers functioned independently and thereby avoided the decreased reliability that occurs when using a consensus processes (43). Our case creation process started with a broad screening process involving 2 physicians and 1 nurse who independently searched for any way care could have been improved prior to the physician reviewers assigning preventability ratings. This process likely increased the sensitivity to detect adverse events. Finally our study took place at a multi-site academic teaching hospital in a major Canadian city that is unlikely to be an outlier for preventable death rate. Judging by the Hospital Standardized Mortality Ratio (the ratio of hospital mortality rate divided by the expected mortality rate) our hospital is doing better than many in the country (66).

### *4.7 Limitations*

A major limitation is our failure to generate an estimate of the proportion of deaths that are completely preventable by better medical care. At the outset of the study we anticipated that the latent class model would specify a class that had a high probability of receiving a high preventability rating and could then be labeled "preventable death". No model contained such a class. A related limitation is different definitions of the "possibly preventable death" class between models. Each time an analysis is conducted the meaning of "possibly preventable death" will be slightly different, preventing comparison of possibly-preventable-death rates between measurements at different centers or over time. Despite this limitation the meaning of the "possibly preventable" class was very similar in each of our latent class models. This issue could be remedied if a standard definition of "possibly preventable death" could be specified in the model and is an opportunity for future research. Specifying a latent class definition would likely decrease model fit.

The author (DK) who created the summaries may have introduced bias in the case creation process. Considerable efforts were made to guard against this by having summaries written by an independent nurse and physician first. DK then read these summaries, reviewed the chart again and created each case summary. Any gaps in care picked up by the nurse or physician were included in the case summary and we used a standard format to record the summaries (Appendix 4). Each case included only facts about the case including quotes from the chart as appropriate. Never the less there is still risk of bias in the process.

Reviewers were predominately internists, which may have biased the results. All cases had a physician from the specialty of the admitting service summarize the death to create the case summary. This physician also gave expert opinion in select cases when the reviewer was unlikely to have clinical expertise to make the judgment. Even with these precautions there could be bias introduced by the reviewer population. Another concern is that DK reviewed all deaths in the sample so my ratings had more weight than any other reviewer. This is not necessarily a weakness because I was not an outlier in the proportion of high or low ratings assigned and my calculated sensitivity and specificity were good, although that could be partially due to reverse causation.

Bias is almost certainly present in the reviewer's judgments. It has been shown in prospective studies that physicians are overly optimistic about their patient's survival (67, 68). If this is the case in retrospective chart review also then reviewers may be over calling the preventability of death, blaming the death on the adverse event because they expected the person not to die. Hindsight bias, also called the knew-it-all-along effect, is another well-described phenomenon where an outcome is judged to be more likely when the outcome is known prior to making the judgment (69, 70). This effect may be at play when a reviewer reads about sub-optimal care, already knowing that the patient dies. It is more likely the reviewer will judge the outcome of the AE to be death when in fact the death was caused by other factors. Both hindsight bias and optimism bias would likely inflate the estimate of preventable deaths.

### *4.8 Implications*

Although preventable deaths are not common, they are the most serious of adverse events and warrant efforts to reduce and eliminate them. Significant resources have gone to reducing AEs and preventable deaths since the IOM report "To Err is Human" 15 years ago. Despite this effort, the possibly preventable death rate per admission from this thesis is similar to the original measurements performed over 30 years ago by *Brennan et al.* While the methodology used in this thesis is not directly comparable the similarity of the numbers is interesting. Researchers involved in the first measurements of adverse events have also raised concern that patient safety has not improved in the last two decades (13, 71). A 2010 study of patient safety in 10 hospitals in North Carolina USA found that there was no change in patient safety outcomes over a 6 year period despite significant efforts (72). Another recent study had more encouraging results, showing a decrease in adverse events from 2005 - 2011 in patients admitted for heart failure, myocardial infarction and pneumonia (73).

A first step in reducing preventable deaths is having reliable measures so that progress can be tracked. The hospital standardized mortality ratio is a system wide measure that is supposed to monitor preventable deaths. As our study has shown, the proportion of truly preventable deaths is less than 6%, therefore the signal to noise ratio for the hospital standardized mortality ratio (HSMR) will be low. *Girling et al.* quantified the sensitivity of the HSMR to variations in preventable death with a modeling study and found that if the preventable death rate is 6% then the predictive value of the HSMR is only 9% (14). There are several studies that correlate standardized mortality rates with other measures of quality care, all showing little or no correlation (15, 74, 75). With no accurate and reliable system wide measure of preventable deaths, we are left with the peer review methods used in this thesis and in previous work. Chart review methods are resource intensive, have low reliability and the methods are not standardized. Most peer review studies do not even acknowledge that preventability is a continuous latent variable. In its current state peer review is too labor intensive and unreliable to use for system wide monitoring of preventable deaths.

Before devoting more effort to measuring preventable deaths we need to consider how these measurements will be used to make a safer health system. There are many types of AEs, each with a longer list of potential root causes, which may be specific to the local environment where the AE occurred. Because the root causes of preventable adverse events are diverse there is no clear action to take if preventable deaths are high in a particular hospital or region. A strategy to reduce preventable deaths needs to focus on improving specific processes and outcomes that have good evidence for best practice (for example hand hygiene to prevent hospital acquired infections) (76). Outcomes related to a specific process are easier to measure and the changes needed to improve the outcome are easier to address. Measuring preventable deaths on the other hand does not directly suggest how to reduce preventable deaths. It makes sense for hospitals to perform peer review measures of preventable death because it draws attention to patient safety issues and highlights common and critical types of AE. At the system level peer-review methods of measuring preventable death are too resource intensive and unreliable to track progress. Instead patient safety indicators need to be more granular, focusing on specific evidence based problems that can be measured and improved upon. This is the approach that has been taken by health quality organizations like Health Quality Ontario, the World Health Organization and the Agency for Healthcare Research and Quality in the United States.

### *4.9 Conclusion and future research*

We found that 6% of deaths occurring in hospital are possibly preventable. This is consistent with previous findings but unique in that we are explicit about the meaning of the term "possibly". The number of truly preventable deaths is certainly much lower than 6%, but even 6% "possibly preventable" is unacceptably high. The path to reducing preventable deaths is focusing on upstream processes and outcomes that can be measured more easily and more reliably than preventable deaths. Focusing only on preventable deaths gives a small sample of the diversity of system issues that are responsible for preventable morbidity and mortality. Never the less, periodic audits of preventable death motivate continued patient safety efforts and draw attention to some of the most critical flaws in care processes.

Future measurements of preventable death need to use a standardized definition of preventable death that treats preventability as a continuous latent variable. We are working on this currently. We identified preoperative risk assessment and recognition of sepsis as two care processes that have caused preventable deaths at our hospital. Root cause analysis and standardization of these processes needs to occurs so that the best evidence is applied to every patient.

## References


1. Canada H. About Health Canada: About Mission, Values, Activities 2011 [cited 2015]. Available from: http://www.hc-sc.gc.ca/ahc-asc/activit/about-apropos/index-eng.php - share.
2. T KL, M CJ, MS D. To err is human: building a safer health system. Washington, DC: Institute of Medicine, 2000.
3. Baker GR, Norton PG, Flintoft V, Blais R, Brown A, Cox J, et al. The Canadian Adverse Events Study: the incidence of adverse events among hospital patients in Canada. CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne. 2004;170(11):1678-86.
4. Stelfox HT, Palmisani S, Scurlock C, Orav EJ, Bates DW. The "To Err is Human" report and the patient safety literature. Quality & safety in health care. 2006;15(3):174-8.
5. McLaws ML, Pantle AC, Fitzpatrick KR, Hughes CF. Improvements in hand hygiene across New South Wales public hospitals: clean hands save lives, part III. Med J Aust. 2009;191(8 Suppl):S18-24.
6. Sax H, Allegranzi B, Uckay I, Larson E, Boyce J, Pittet D. 'My five moments for hand hygiene': a user-centred design approach to understand, train, monitor and report hand hygiene. J Hosp Infect. 2007;67(1):9-21.
7. Mehta JA, Sable SA, Nagral S. Updated recommendations for control of surgical site infections. Ann Surg. 2015;261(3):e65.
8. Padula WV, Makic MB, Mishra MK, Campbell JD, Nair KV, Wald HL, et al. Comparative effectiveness of quality improvement interventions for pressure ulcer prevention in academic medical centers in the United States. Jt Comm J Qual Patient Saf. 2015;41(6):246-5.
9. Information CIfH. HSMR: A New Approach for Measuring Hospital Mortality Trends in Canada Ottawa Canada: CIHI, 2007.
10. Information CIfH. Learning From the Best:
Benchmarking Canada’s
Health System Ottawa ON: CIHI, 2011.
11. James JT. A new, evidence-based estimate of patient harms associated with hospital care. Journal of patient safety. 2013;9(3):122-8.
12. Kennerly DA, Kudyakov R, da Graca B, Saldana M, Compton J, Nicewander D, et al. Characterization of adverse events detected in a large health care delivery system using an enhanced global trigger tool over a five-year interval. Health Serv Res. 2014;49(5):1407-25.
13. Leape LL, Berwick DM. Five years after To Err Is Human: what have we learned? Jama. 2005;293(19):2384-90.
14. Girling AJ, Hofer TP, Wu J, Chilton PJ, Nicholl JP, Mohammed MA, et al. Case-mix adjusted hospital mortality is a poor proxy for preventable mortality: a modelling study. BMJ quality & safety. 2012;21(12):1052-6.
15. Goodacre S, Campbell M, Carter A. What do hospital mortality rates tell us about quality of care? Emerg Med J. 2015;32(3):244-7.

16. van den Bosch WF, Kelder JC, Wagner C. Predicting hospital mortality among frequently readmitted patients: HSMR biased by readmission. BMC health services research. 2011;11:57.
17. van Gestel YR, Lemmens VE, Lingsma HF, de Hingh IH, Rutten HJ, Coebergh JW. The hospital standardized mortality ratio fallacy: a narrative review. Medical care. 2012;50(8):662-7.
18. Shojania KG, Forster AJ. Hospital mortality: when failure is not a good measure of success. CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne. 2008;179(2):153-7.
19. Leape LL. Institute of Medicine medical error figures are not exaggerated. Jama. 2000;284(1):95-7.
20. McDonald CJ, Weiner M, Hui SL. Deaths due to medical errors are exaggerated in Institute of Medicine report. Jama. 2000;284(1):93-5.
21. Gaieski DF, Mikkelsen ME, Band RA, Pines JM, Massone R, Furia FF, et al. Impact of time to antibiotics on survival in patients with severe sepsis or septic shock in whom early goal-directed therapy was initiated in the emergency department. Crit Care Med. 2010;38(4):1045-53.
22. Brennan TA, Leape LL, Laird NM, Hebert L, Localio AR, Lawthers AG, et al. Incidence of adverse events and negligence in hospitalized patients. Results of the Harvard Medical Practice Study I. N Engl J Med. 1991;324(6):370-6.
23. Hiatt HH, Barnes BA, Brennan TA, Laird NM, Lawthers AG, Leape LL, et al. A study of medical injury and medical malpractice. N Engl J Med. 1989;321(7):480-4.
24. Wilson RM, Runciman WB, Gibberd RW, Harrison BT, Newby L, Hamilton JD. The Quality in Australian Health Care Study. Med J Aust. 1995;163(9):458-71.
25. Vincent C, Neale G, Woloshynowych M. Adverse events in British hospitals: preliminary retrospective record review. Bmj. 2001;322(7285):517-9.
26. Thomas EJ, Studdert DM, Burstin HR, Orav EJ, Zeena T, Williams EJ, et al. Incidence and types of adverse events and negligent care in Utah and Colorado. Medical care. 2000;38(3):261-71.
27. Schioler T, Lipczak H, Pedersen BL, Mogensen TS, Bech KB, Stockmarr A, et al. [Incidence of adverse events in hospitals. A retrospective study of medical records]. Ugeskr Laeger. 2001;163(39):5370-8.
28. 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.
29. 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):U1909.
30. Soop M, Fryksmark U, Koster M, Haglund B. The incidence of adverse events in Swedish hospitals: a retrospective medical record review study. International journal for quality in health care : journal of the International Society for Quality in Health Care / ISQua. 2009;21(4):285-91.
31. Sox Jr HC, Woloshin S. How many deaths are due to medical error? Getting the number right. Eff Clin Pract. 2000;3(6):277-83.
32. Dubois RW, Brook RH. Preventable deaths: who, how often, and why? Ann Intern Med. 1988;109(7):582-9.

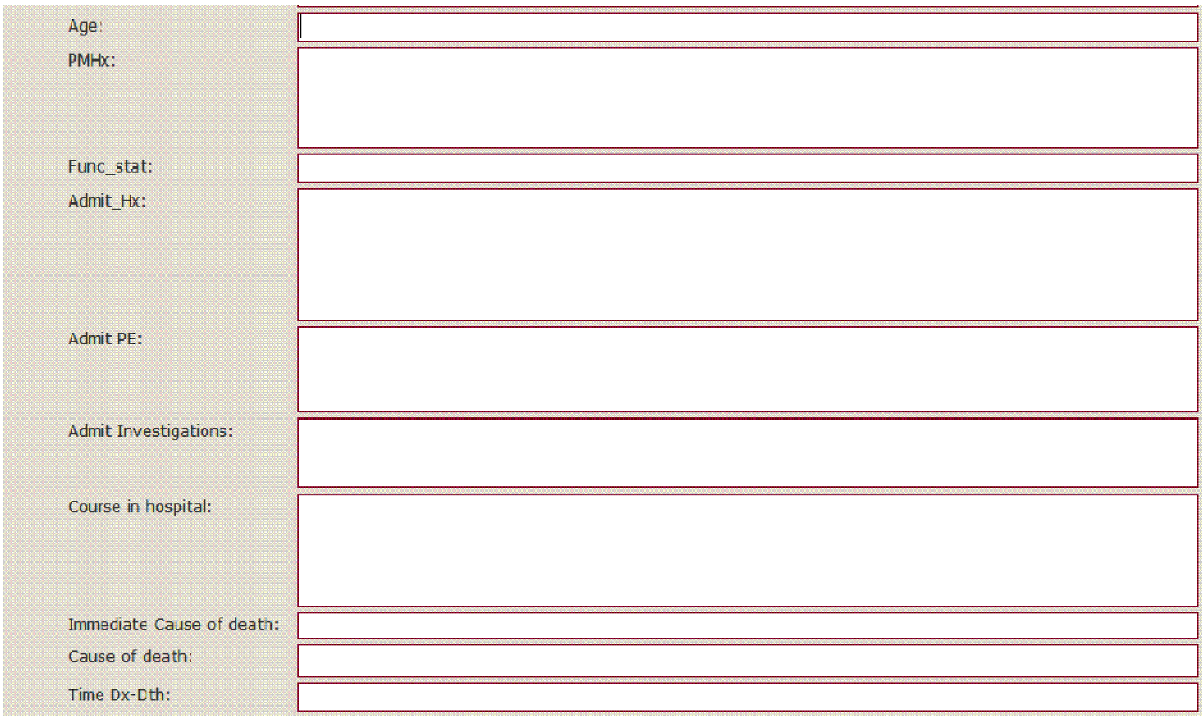
33. Hayward RA, Hofer TP. Estimating hospital deaths due to medical errors: preventability is in the eye of the reviewer. Jama. 2001;286(4):415-20.
34. Zegers M, de Bruijne MC, Wagner C, Hoonhout LH, Waaijman R, Smits M, et al. Adverse events and potentially preventable deaths in Dutch hospitals: results of a retrospective patient record review study. Quality & safety in health care. 2009;18(4):297-302.
35. 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 quality & safety. 2012;21(9):737-45.
36. Thomas EJ, Lipsitz SR, Studdert DM, Brennan TA. The reliability of medical record review for estimating adverse event rates. Ann Intern Med. 2002;136(11):812-6.
37. Localio AR, Weaver SL, Landis JR, Lawthers AG, Brenhan TA, Hebert L, et al. Identifying adverse events caused by medical care: degree of physician agreement in a retrospective chart review. Ann Intern Med. 1996;125(6):457-64.
38. Goldman RL. The reliability of peer assessments of quality of care. Jama. 1992;267(7):958-60.
39. Hayward RA, McMahon LF, Jr., Bernard AM. Evaluating the care of general medicine inpatients: how good is implicit review? Ann Intern Med. 1993;118(7):550-6.
40. Rubenstein LV, Kahn KL, Reinisch EJ, Sherwood MJ, Rogers WH, Kamberg C, et al. Changes in quality of care for five diseases measured by implicit review, 1981 to 1986. Jama. 1990;264(15):1974-9.
41. 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):2349-54.
42. Zegers M, de Bruijne MC, Wagner C, Groenewegen PP, van der Wal G, de Vet HC. The inter-rater agreement of retrospective assessments of adverse events does not improve with two reviewers per patient record. J Clin Epidemiol. 2010;63(1):94-102.
43. Hofer TP, Bernstein SJ, DeMonner S, Hayward RA. Discussion between reviewers does not improve reliability of peer review of hospital quality. Medical care. 2000;38(2):152-61.
44. Hayward RA, Heisler M, Adams J, Dudley RA, Hofer TP. Overestimating outcome rates: statistical estimation when reliability is suboptimal. Health Serv Res. 2007;42(4):1718-38.
45. Coory M, Gibberd R. New measures for reporting the magnitude of small-area variation in rates. Statistics in medicine. 1998;17(22):2625-34.
46. Gatsonis CA, Epstein AM, Newhouse JP, Normand SL, McNeil BJ. Variations in the utilization of coronary angiography for elderly patients with an acute myocardial infarction. An analysis using hierarchical logistic regression. Medical care. 1995;33(6):625-42.
47. Hofer TP, Hayward RA, Greenfield S, Wagner EH, Kaplan SH, Manning WG. The unreliability of individual physician "report cards" for assessing the costs and quality of care of a chronic disease. Jama. 1999;281(22):2098-105.

48. Schulzer M, Anderson DR, Drance SM. Sensitivity and specificity of a diagnostic test determined by repeated observations in the absence of an external standard. J Clin Epidemiol. 1991;44(11):1167-79.
49. Escobar GJ, Greene JD, Scheirer P, Gardner MN, Draper D, Kipnis P. Risk-adjusting hospital inpatient mortality using automated inpatient, outpatient, and laboratory databases. Medical care. 2008;46(3):232-9.
50. van Walraven C, Escobar GJ, Greene JD, Forster AJ. The Kaiser Permanente inpatient risk adjustment methodology was valid in an external patient population. Journal of clinical epidemiology. 2010;63(7):798-803.
51. Hofer TP, Asch SM, Hayward RA, Rubenstein LV, Hogan MM, Adams J, et al. Profiling quality of care: Is there a role for peer review? BMC health services research. 2004;4(1):9.
52. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull. 1979;86(2):420-8.
53. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet. 1986;1(8476):307-10.
54. Hofer TP. Commentary on Sharek: adverse events and errors-important to differentiate and difficult to measure. Health Serv Res. 2011;46(2):679-84.
55. Colins LM, Lanza ST. Latent Class and Latent Transition Analysis: With Applications in the Social, Behavioral and Health Sciences: Wiley; 2010.
56. Koehler KJ. Goodness-of-fit tests for log-linear models in sparse contingency tables. Journal of the American Statistical Association. 1986;81(394):483-93.
57. Koehler KJ, Larntz K. An empirical investigation of goodness-of-fit statistics for sparse multinomials. Journal of the American Statistical Association. 1980;75(370):336-44.
58. G. M, D. P. Finite Mixture Models. New York: Wiley; 2000.
59. Gelman A, Meng XL, Stern HS. Posterior predictive assessment of model fitness via realized discrepancies. Statistica Sinica. 1996;6:733-807.
60. Fewell Z. Statistical Evaluation of Measurement Errors: Design and Analysis of Reliability Studies. G Dunn. London: Arnold, 2004, pp. 224, £50.00 (HB), ISBN: 0-340-76070-2. International Journal of Epidemiology.34(2):499-.
61. Sousa P, Uva AS, Serranheira F, Nunes C, Leite ES. Estimating the incidence of adverse events in Portuguese hospitals: a contribution to improving quality and patient safety. BMC health services research. 2014;14:311.
62. Jenkins TC, Price CS, Sabel AL, Mehler PS, Burman WJ. Impact of routine infectious diseases service consultation on the evaluation, management, and outcomes of Staphylococcus aureus bacteremia. Clin Infect Dis. 2008;46(7):1000-8.
63. Lopez-Cortes LE, Del Toro MD, Galvez-Acebal J, Bereciartua-Bastarrica E, Farinas MC, Sanz-Franco M, et al. Impact of an evidence-based bundle intervention in the quality-of-care management and outcome of Staphylococcus aureus bacteremia. Clin Infect Dis. 2013;57(9):1225-33.
64. Robinson JO, Pozzi-Langhi S, Phillips M, Pearson JC, Christiansen KJ, Coombs GW, et al. Formal infectious diseases consultation is associated with decreased mortality in Staphylococcus aureus bacteraemia. Eur J Clin Microbiol Infect Dis. 2012;31(9):2421-8.

65. Fleisher LA, Fleischmann KE, Auerbach AD, Barnason SA, Beckman JA, Bozkurt B, et al. 2014 ACC/AHA guideline on perioperative cardiovascular evaluation and management of patients undergoing noncardiac surgery: a report of the American College of Cardiology/American Heart Association Task Force on practice guidelines. J Am Coll Cardiol. 2014;64(22):e77-137.
66. CIHI. Your Health System: Candian Institue for Health Information; 2015 [cited 2015 July 3rd ]. Available from: http://yourhealthsystem.cihi.ca/hsp/inbrief?lang=en - !/indicators/005/hospital-deaths-hsmr/;mapC1;mapLevel2;trend(C1,C5001);/.
67. Christakis NA, Lamont EB. Extent and determinants of error in physicians' prognoses in terminally ill patients: prospective cohort study. West J Med. 2000;172(5):310-3.
68. Glare P, Virik K, Jones M, Hudson M, Eychmuller S, Simes J, et al. A systematic review of physicians' survival predictions in terminally ill cancer patients. Bmj. 2003;327(7408):195-8.
69. Fischhoff B. Hindsight not equal to foresight: the effect of outcome knowledge on judgment under uncertainty. 1975. Quality & safety in health care. 2003;12(4):304-11; discussion 11-2.
70. Fischhoff B, Beyth R. "I knew it would happen"-Remembered probabilities of once-future things. (1974) "I knew it would happen"-Remembered probabilities of once-future things 22 pp Oxford, England: Oregon Research Inst, Research Bull; England. 1974.
71. Vincent C, Aylin P, Franklin BD, Holmes A, Iskander S, Jacklin A, et al. Is health care getting safer? Bmj. 2008;337:a2426.
72. Landrigan CP, Parry GJ, Bones CB, Hackbarth AD, Goldmann DA, Sharek PJ. Temporal trends in rates of patient harm resulting from medical care. N Engl J Med. 2010;363(22):2124-34.
73. Wang Y, Eldridge N, Metersky ML, Verzier NR, Meehan TP, Pandolfi MM, et al. National trends in patient safety for four common conditions, 2005-2011. N Engl J Med. 2014;370(4):341-51.
74. Silva LK. Validity of the risk adjustment approach to compare outcomes. Cad Saude Publica. 2003;19(1):287-95.
75. Thomas JW, Holloway JJ, Guire KE. Validating risk-adjusted mortality as an indicator for quality of care. Inquiry. 1993;30(1):6-22.
76. Allegranzi B, Bagheri Nejad S, Combescure C, Graafmans W, Attar H, Donaldson L, et al. Burden of endemic health-care-associated infection in developing countries: systematic review and meta-analysis. Lancet. 2011;377(9761):228-41.

# Appendix 1

## Form used to create case summary's


Age:

PMHx:

Func_stat:

Admit_Hx:

Admit PE:

Admit Investigations:

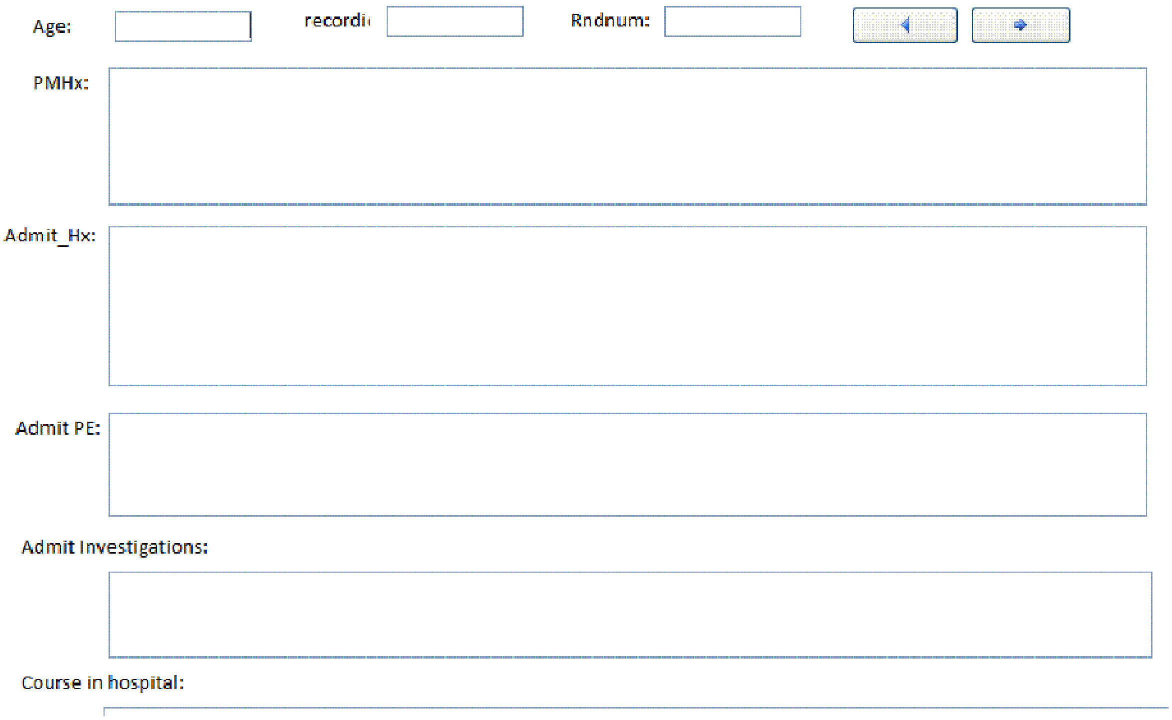
Course in hospital:

Immediate Cause of death:

Cause of death:

Time Dx-Dth:

## Form used by reviewers to assign preventability ratings


Age:

recordi

Rndnum:

PMHx:

Admit_Hx:

Admit PE:

Admit Investigations:

Course in hospital:

**What could have been done differently, that would be considered standard of care and could have decreased the risk of this patient dying in hospital?**

**What is the probability the changes you described would have prevented the death?**

0

**Was the death avoidable by actions taken before the hospitalization, during the hospitalizaiton or both?**

- ○ Before the hospitalization
- ○ During the hospitalization
- ○ Both before and during the hospitalization

**If this patient recieved high quality of care would they have lived for 3 more months?**

- ○ Almost certainly would be alive
- ○ Uncertain but probably would be alive
- ○ Uncertain but probably would be dead
- ◉ Almost certainly would be dead

# Appendix 2

**Instructions to reviewers**

You have been asked to review deaths from the Ottawa Hospital and judge whether high quality care with no errors would have prevented each death. Previous studies of this question have shown that between 1-6% of in hospital deaths have some element of preventability. All of these deaths occurred between September and December 2013.

Definitions:

*What is high quality care*

High quality of care is the best care you can expect in the average Canadian hospital. High quality care minimizes the risk of complications, maximizes the likelihood of a good outcome and is respectful of patient's wishes. When this definition is not met there are gaps in the quality of care.

*Standard of care*

Do not hold other physicians to standards that are idiosyncratic to you or controversial with in the profession. Instead the standard should be care processes that would be agreed upon by a group of your peers.

*What is an adverse event?*

An adverse event is an unintended physical or mental injury that occurs as the result of medical management rather than by the patient's underlying disease process. An adverse event could result from acts of commission (incorrect diagnosis, treatment or poor performance) or acts of omission (a failure to diagnose or treat).

*How to know if a death is preventable:*

A death should be judged as preventable when there are gaps in quality of care resulting in an adverse event that cause death.

<u>For each case you will be asked the following questions:</u>

1) What could have been done differently, that would be considered with in the standard of care, to reduce the risk of this patient dying during the current hospitalization? (Free text)

Each case has been reviewed by 2 physicians and a nurse who spoke with the care team. Any questionable circumstances are described. If you are told in the description that the patient was diagnosed with pneumonia and treated then you should assume the diagnosis is correct and it was done properly.

Do not describe actions that could have led to faster discharge so the patient could die somewhere other than hospital.  Please only describe acts of commission or omission by the healthcare system that could have reduced the chance of the patient dying during the current hospitalization at TOH.

2) What is the probability the changes you described would have prevented the death? (0-100)

Judge the probability that the changes you described would allow the patient to be discharge from hospital, either to home or to another facility.

There are two levels of uncertainty with this question.  There is information missing because data is retrospective there will therefore be uncertainty that the changes in care you described would have actually been appropriate and considered “standard of care” .  There will also be uncertainty that the changes you describe would have prevented the death in hospital.  Please give the combined probability taking into account both levels of uncertainty.  If you are finding this difficult it may be helpful to first decide on each of these probabilities and then multiply them to get the final probability.

3) Was the death avoidable by actions taken before hospitalization, during hospitalization or both? (Options)

The term hospitalization refers only to the patient's time admitted to The Ottawa Hospital. If they were previously at a peripheral hospital where there were errors in care then the "before hospitalization" option should be chosen.

4) If this patient received high quality care would they have lived for 3 more months? (Options)

- Almost certainly would be alive
- Uncertain but probably would be alive
- Uncertain but probably would be dead
- Almost certainly would be dead

If there are no changes in the patient's care that could have reduced their chances of dying in hospital then enter a probability of 0 and pick the "almost certainly would be dead" option for the last question. Leave the other questions blank.

Examples:

Here two example cases.   One is considered highly preventable while the other has no element of preventability.

Case #1:  89 yo M presents to ER with dyspnea and chest pain worsening for 18 hours.

PMHx:  moderate – severe dementia, CAD, Severe AS with AVA 0.79 cm – has previously been seen by cardiology regarding AVR and after discussion with the patient and family it was decided he was not a surgical candidate or trans-aortic valve implantation candidate because of frailty and limited benefit.

O/E: decrease LOC – makes incoherent noises when stimulated. HR 101 BP 73/40 RR 26 Sat 100% on 5L  Resp: Crackles through out lung fields, increased WOB  CVS: JVP 6 cm ASA, typical AS murmur heard.

Investigations:

TnI 23.08, CK 201 ECG: ST depression in pre-cordial leads.  CXR pulmonary edema.

Course in Hospital: He was initially started on BiPAP, given antiplatlet agents and IV heparin for ACS. Cardiology assessed the patient and discussed treatment of ASC, hypotension and severe AS. He had ongoing increased work of breathing despite BiPAP and became more hypotensive. In discussion with the family it was clear that more aggressive treatment would be of limited benefit and not what the patient would want. The goals of his care were changed to comfort only and he died 6 hours later.

Case #2 69 yo F with fever and decreased LOC x 6 hours

PMHx: HTN

O/E HR 110 BP 100/60 Temp 38.4

GCS 10 no focal neuro deficits

Investigation: WBC 18.2 Bands 1.2

Course in Hospital: She was started on the sepsis protocol and orders were written to recieve pip-tazo + vanco. Full cultures including an LP were performed. CSF WBC 543 gram stain negative. She was ordered to receive ceftriaxone, ampicillin and vancomycin. Admitted to ICU because of decreased LOC and possible need for intubation. 24 hours into the admission she was more hypotensive requiring vasopressors and was intubated. It is now noted in the chart that the orders for antibiotics previously ordered were not signed off and she appears not to have received them. She was now started on the previously ordered antibiotics for meningitis at > 24 hours since presentation. PAD #2 oxygenation became worse. Requiring oscillator with NO – ongoing hypoxemia. Sats max at 88% on 100% FiO2, PEA arrest. Unsuccessful CPR.

Further Information:

Reviewing 170 cases should take you 4-6 hours. You can see which number you are on at the bottom of the screen.

A death is not preventable when the patient and/or medical team decide to NOT offer a treatment because of lack of benefit in their judgment or a shared decision making process.

Complications of procedures or nosacomial infections are only preventable if there are measures known to reduce their incidence. For example, if a Clostridium difficile infection contributes to a patient's death then there is a element of preventability because some C diff infections can be prevented with proper infection control measures.

If you have any questions or technical issues please call me or email me.

Civic Hospital: 613-737-8899  Ext 15935

dkobewka@toh.on.ca

# Appendix 3

Ten Fictitious cases given to every reviewer created to be highly preventable or non-preventable.

| Case | Preventable death | Summary of case |
|---|---|---|
| 1 | Yes | 55 year old with history of alcohol abuse and hypertension. Presented with classic symptoms of meningitis and septic shock. Lumbar puncture was performed and ceftriaxone and vancomycin were given. After 2 days cultures came back positive for listeria. Ampicillin was then added to antibiotics. Patient had ongoing septic shock with multi-organ failure and died. |
| 2 | Yes | 65 year old patient with recent diagnosis of renal cell carcinoma presents with tachycardia and hypotension. Treated as septic shock with broad-spectrum antibiotics no other causes considered according to chart documentation. On day 2 of admission has worsening hypotension and ongoing tachycardia. Note in chart this day documents unilateral leg swelling, erythema and tenderness to palpation. He is started on full dose anticoagulation for possible PE but has cardiac arrest before diagnosis can be confirmed. |
| 3 | Yes | 62 year old male presents with sepsis. He was given piperacillin-tazobactam despite documented penicillin allergy. The patient had an anaphylactic reaction and developed hypoxemia and refractory shock. Cardiac arrest ensued despite appropriate management. |
| 4 | Yes | 76 year old female with a history of type 2 diabetes, obesity and hypertension presents with nausea, vomiting and dyspnea. An ECG and TnI are not performed until day 3 in hospital which show a missed STEMI. She goes on to die from cardiogenic shock despite aggressive management. |
| 5 | Yes | 51 year old male with cirrhosis secondary to HCV presents to hospital with confusion. Diagnosed with hepatic encephalopathy secondary to constipation. He does not improve over 4 days despite lactulose. At this point a paracentesis is performed that shows bacterial peritonitis. He is treated with ceftriaxone but refractory shock ensues |

| | | |
|---|---|---|
| | | leading to death. |
| 6 | No | 48 year-old female with metastatic breast cancer. She had mastectomy and radiation 5 years ago but cancer recurred and is now on experimental chemotherapy. Mets to lungs, brain and bone. Presents with generalized weakness and drowsiness. CT Head, abdo, pelvis show progression of metastatic disease. Care plan decided on by family and oncologist is to proceed with comfort care only. The patient died 2 days later. |
| 7 | No | 96 year-old male with advanced dementia developed pneumonia and was treated with ceftriaxone and azithromycin. He improved but was aspirating with all food consistencies. The patient's substitute decision maker decided against a feeding tube. The patient had very poor PO intake and died after 3 weeks in hospital. |
| 8 | No | 64 year-old female with history of alcohol abuse presented to hospital with an upper GI bleed. An upper endoscopy showed large varicies. They were banded but 6 hours later bleeding recurred. Bleeding could not be controlled endoscopicly. A Blakemore tube was inserted but bleeding with refractory hypotension was ongoing. The patient died despite massive transfusion. |
| 9 | No | 58 year-old male presented with STEMI. He went directly to the cath lab for primary PCI. During the PCI the patient had a VF arrest and was successfully returned to Sinus rhythm with electrical defibrillation and stents were placed in the LAD. Three hours later he had recurrent VF. Multiple attempts at defibrillation were unsuccessful. The patient received CPR, lidocaine, amiodarone and epinephrine over 45 minutes with no success. |
| 10 | No | 66 year-old female with advanced ILD on 3 L of home oxygen presents with worsening hyoxemia. CT chest is consistent with progressive ILD. She was treated with steroids and antibiotics but continued to be increasingly hypoxemic. After 2 days the patient asked to be treated for symptoms only and she died shortly after. |

## Appendix 4

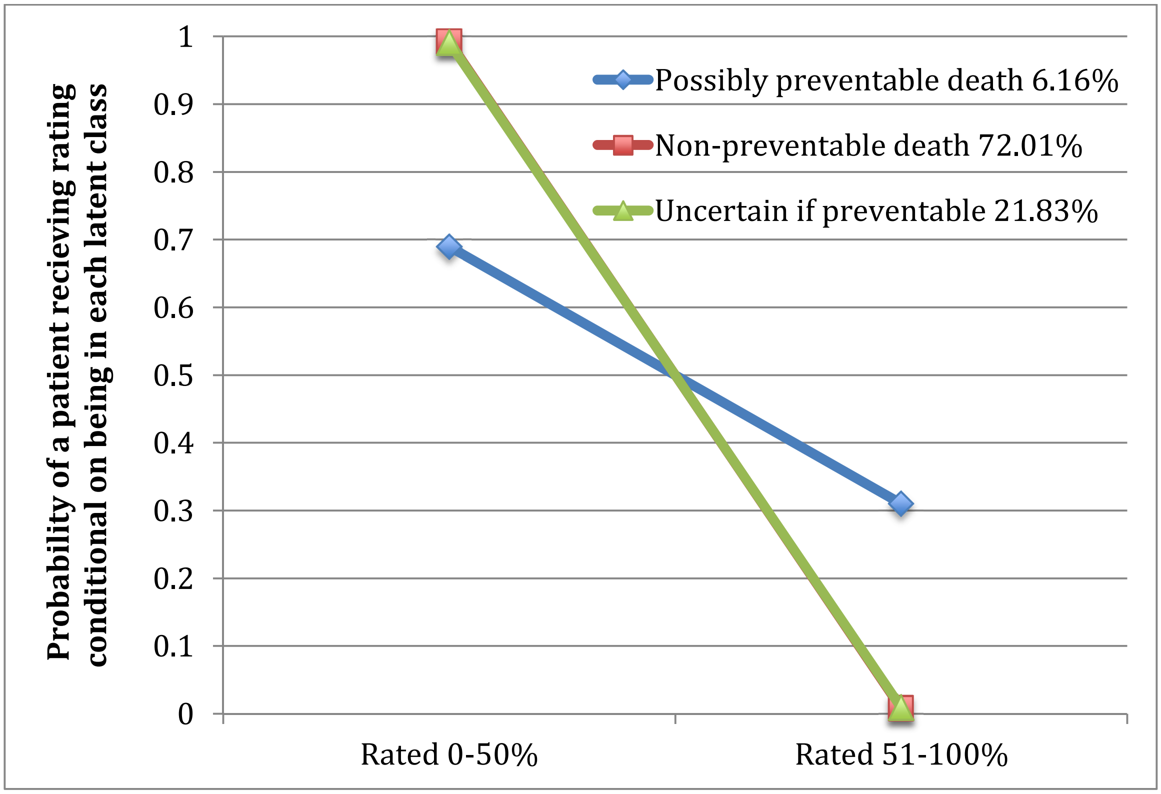



Figure 6: Probability of receiving rating conditional on being in each latent class.

Two-class-model with probabilities in terciles.

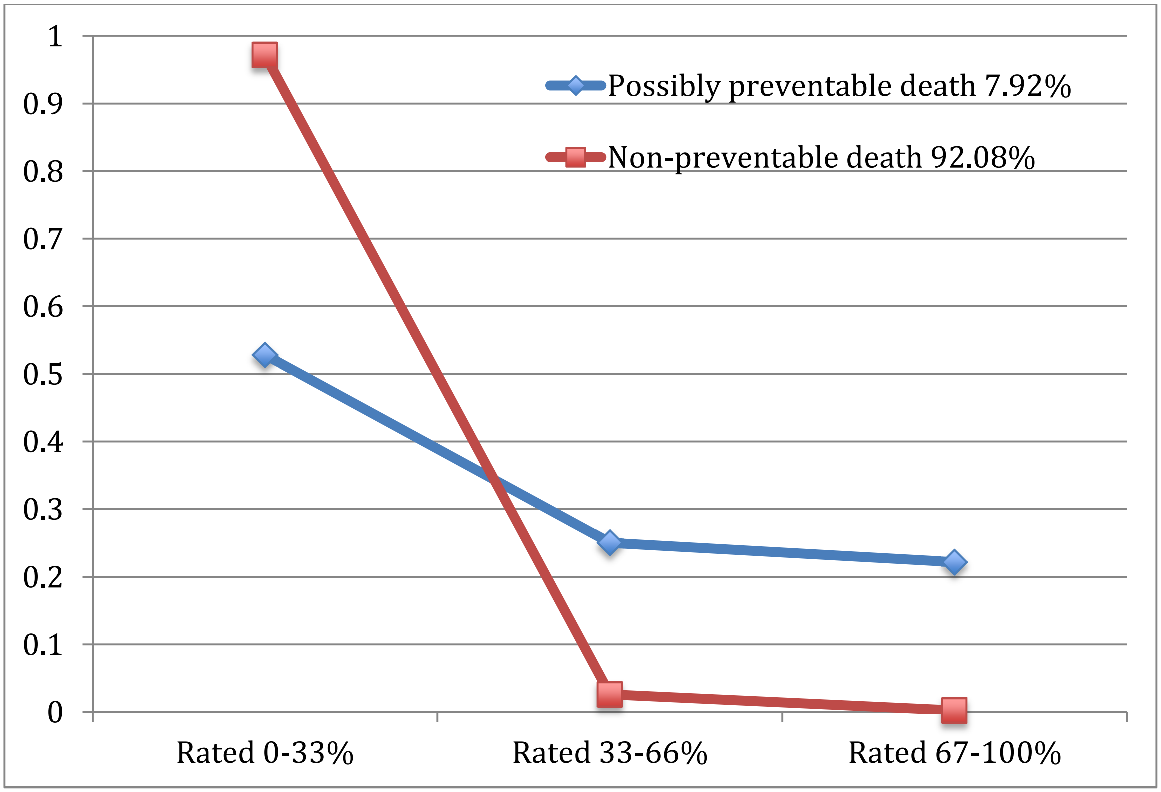



Figure 7: Probability of receiving rating conditional on being in each latent class.

Three-class-model with probabilities dichotomized.

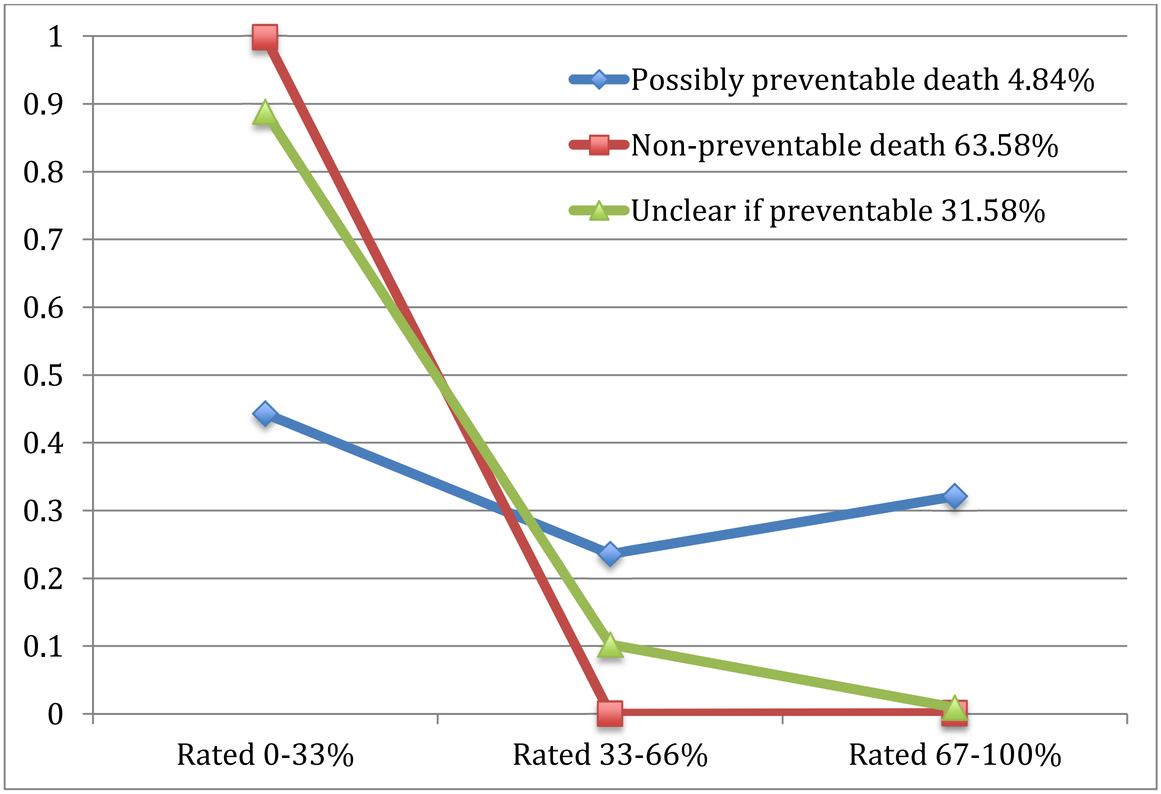



Figure 8: Probability of receiving rating conditional on being in each latent class.

Three-class-model with probabilities in terciles

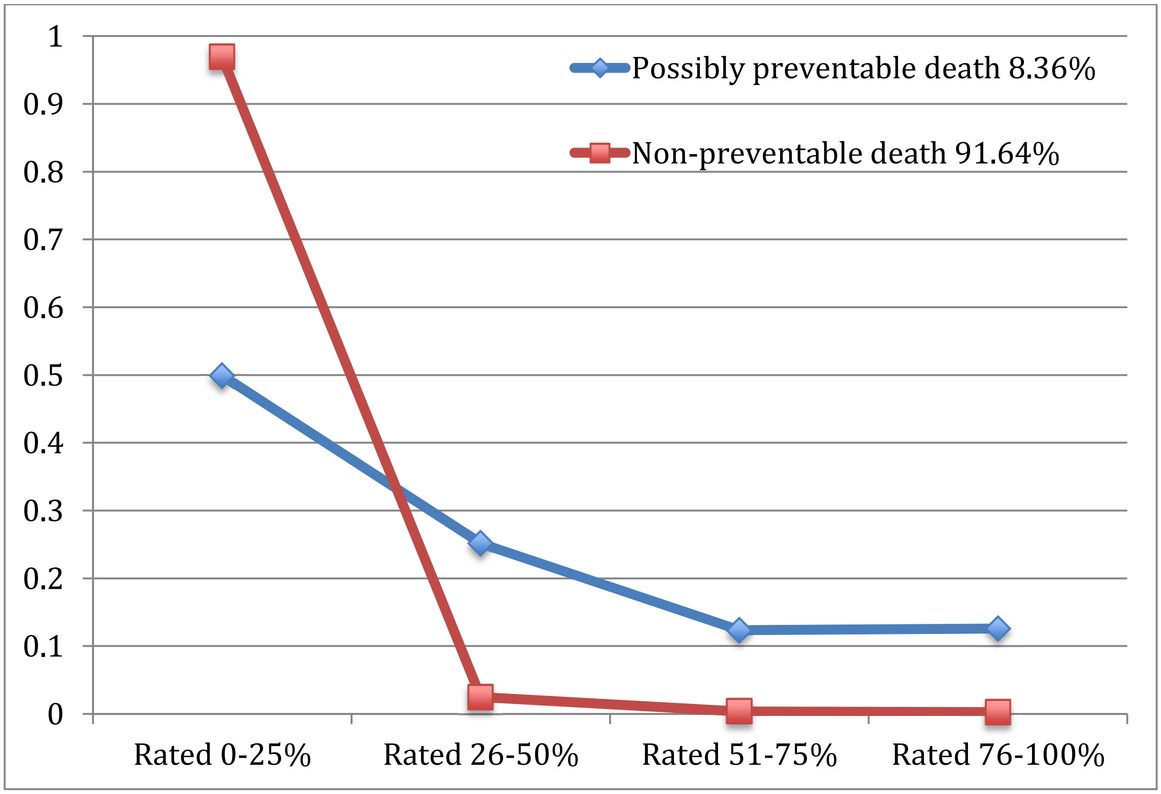



Figure 9: Probability of receiving rating conditional on being in each latent class.

Four-class-model with probabilities in quartiles.

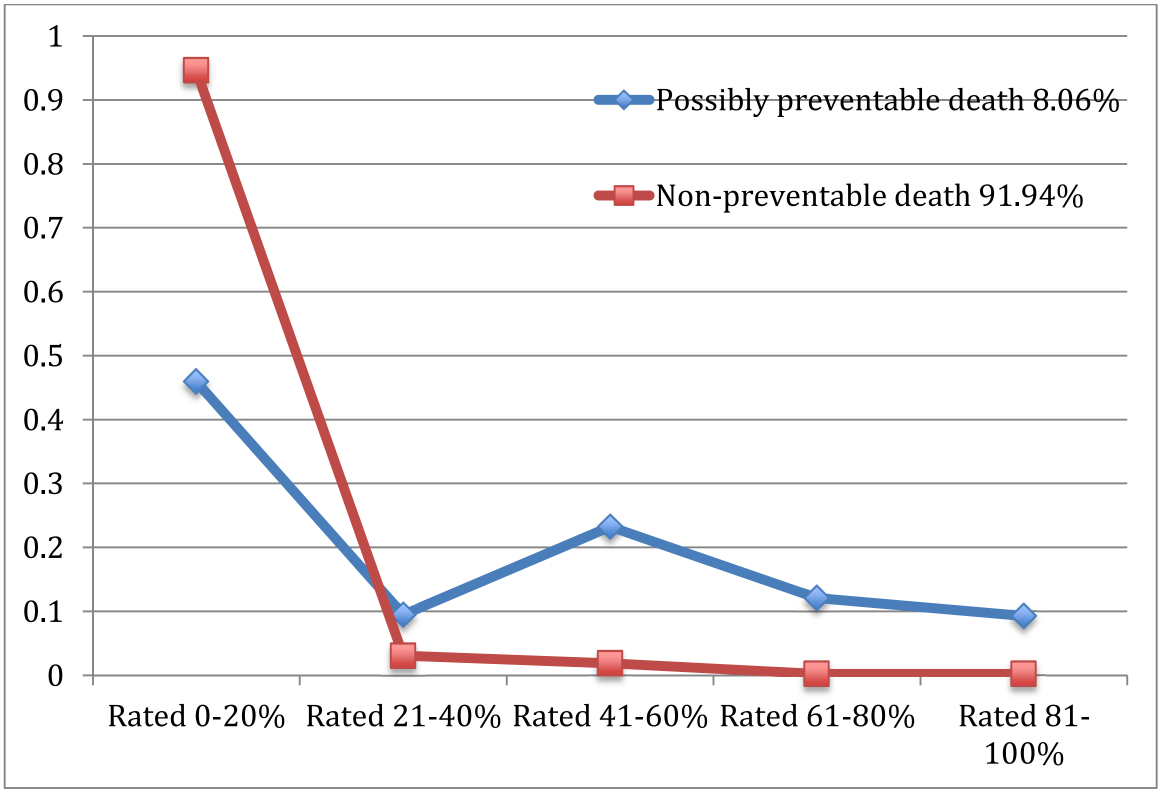



Figure 10: Probability of receiving preventability rating conditional on being in each latent class. Two-class-model with probabilities-of-preventability in quintiles.

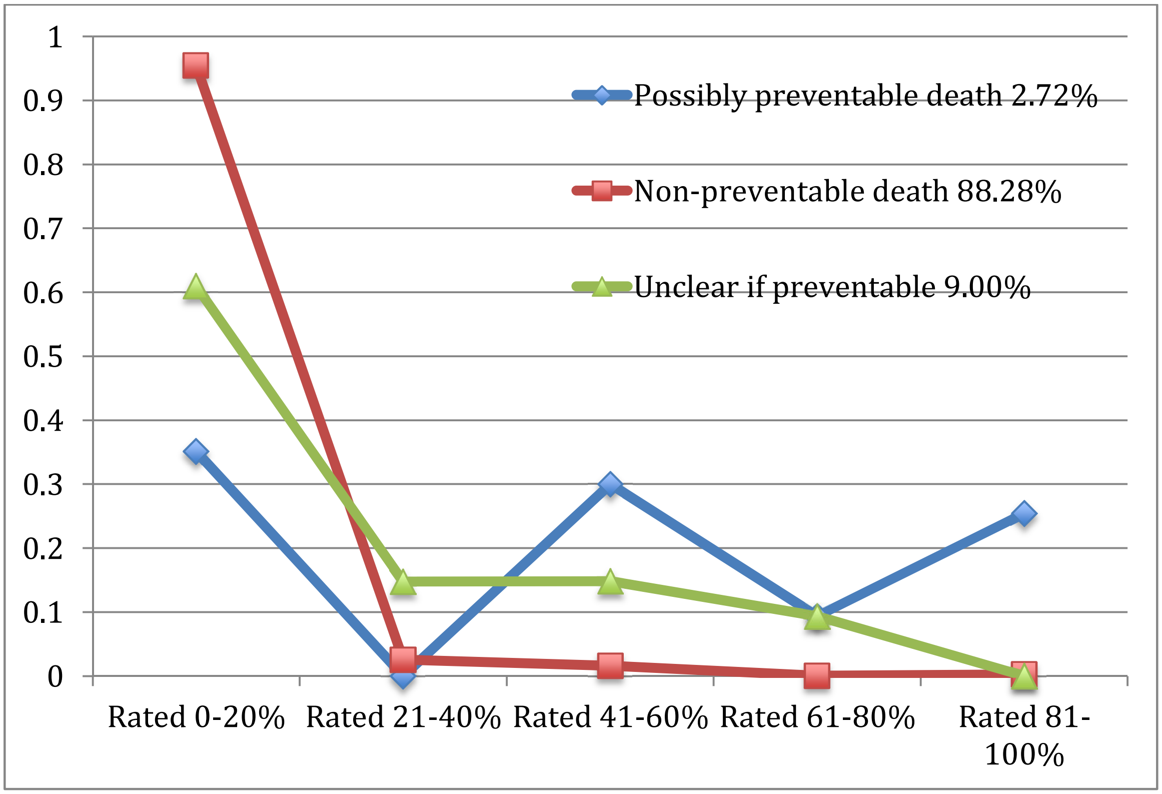



Figure 11: Probability of receiving preventability rating conditional on being in each latent class. Three-class-model with probabilities in quintiles.

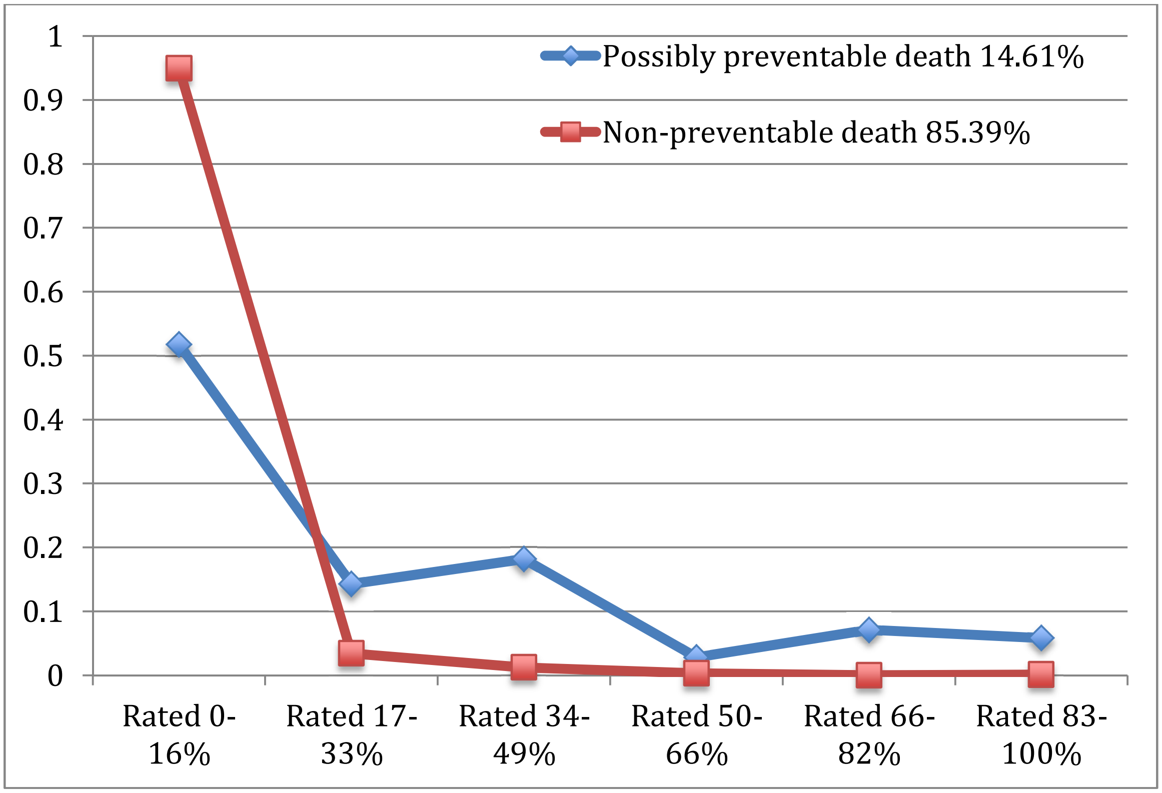



Figure 12: Probability of receiving preventability rating conditional on being in each latent class. Two-class-model with probabilities in sextiles.

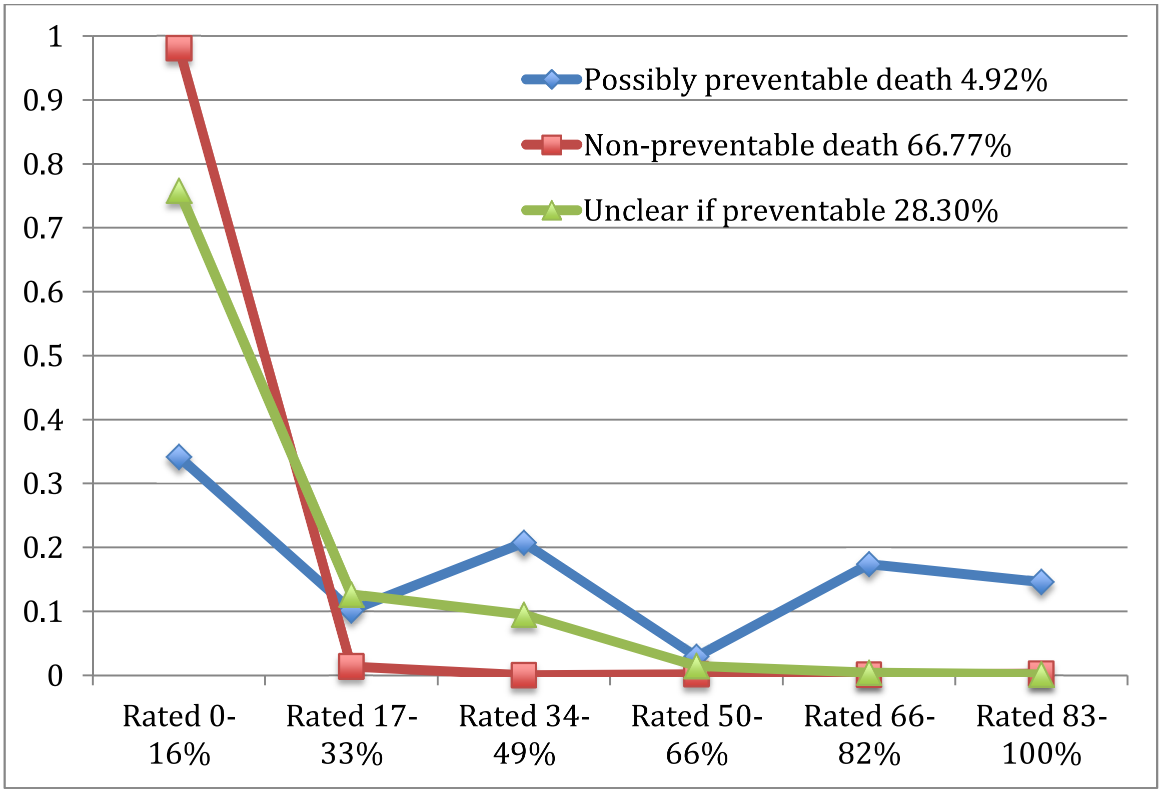



Figure 13: Probability of receiving preventability rating conditional on being in each latent class. Three-class-model with probabilities in sextiles.

**Table 12**: Fit statistics for 10 different models

| Model description | G-squared | AIC | BIC | CAIC | Adjusted BIC | DF | Entropy |
|---|---|---|---|---|---|---|---|
| 3 Classes data in sextiles | 265.2 | 299.2 | 370.16 | 387.16 | 316.2 | 1278 | 0.62 |
| 2 Classes data in sextiles | 283.57 | 305.57 | 351.48 | 362.48 | 316.57 | 1284 | 0.76 |
| 3 Classes data in quintiles | 194.34 | 222.34 | 280.77 | 294.77 | 236.34 | 610 | 0.83 |
| 2 Classes data in quintiles | 206.58 | 224.58 | 262.14 | 271.14 | 233.57 | 615 | 0.86 |
| 3 classes data in quartiles | 125.12 | 147.12 | 193.03 | 204.03 | 158.12 | 244 | 0.79 |
| 2 classes data in quartiles | 128.73 | 142.73 | 171.94 | 178.94 | 149.73 | 248 | 0.86 |
| 3 classes data in terciles | 76.38 | 92.38 | 125.77 | 133.77 | 100.38 | 72 | 0.52 |
| 2 classes data in terciles | 80.44 | 90.44 | 111.31 | 116.31 | 95.44 | 75 | 0.86 |
| 3 classes data dicotomized | 18.62 | 28.62 | 49.49 | 54.49 | 33.58 | 10 | 0.54 |
| 2 classes data dicotomized | 18.58 | 24.58 | 37.1 | 40.1 | 27.58 | 12 | 0.84 |

**Table 13**: Fit statistics for 10 models created with exclusion of the outlying reviewer.

| Model description | G-squared | AIC | BIC | CAIC | Adjusted BIC | DF | Entropy |
|---|---|---|---|---|---|---|---|
| 3 Classes data in sextiles | 211.56 | 245.56 | 316.51 | 333.51 | 262.55 | 1278 | 0.72 |
| 2 Classes data in sextiles | 228.58 | 250.58 | 296.49 | 307.49 | 261.57 | 1284 | 0.79 |
| 3 Classes data in quintiles | 148.47 | 176.47 | 234.9 | 248.9 | 190.47 | 610 | 0.93 |
| 2 Classes data in quintiles | 165.01 | 183.01 | 220.58 | 229.58 | 192.01 | 615 | 0.9 |
| 3 classes data in quartiles | 102.59 | 124.59 | 170.5 | 181.5 | 135.59 | 244 | 0.6 |
| 2 classes data in quartiles | 108.02 | 122.02 | 151.23 | 158.23 | 129.02 | 248 | 0.89 |
| 3 classes data in terciles | 62.16 | 78.16 | 111.55 | 119.55 | 86.16 | 72 | 0.58 |
| 2 classes data in terciles | 67.23 | 77.23 | 98.1 | 103.1 | 82.23 | 75 | 0.91 |
| 3 classes data dicotomized | 23.05 | 33.05 | 53.92 | 58.92 | 38.05 | 10 | 0.28 |
| 2 classes data dicotomized | 23.03 | 29.03 | 41.55 | 44.55 | 32.03 | 12 | 0.79 |