THE ELUSIVE QUALITY OF HEALTH CARE

Av

Boo Svartbo

AKADEMISK AVHANDLING


Avhandlingen baseras på följande arbeten:


ABSTRACT

Background
Wide variations in hospital inpatient death rates have been observed over many years, and concerns have been expressed that such variations could reflect important differences in the quality of medical care. The aim of the study was to develop and apply a method to monitor the quality of medical care and to study the accuracy by which the cause of death was determined.

Methods
From 31 hospitals, data on every discharge were linked with other routinely collected data. In addition, inpatient hospital death rates were analysed as well as data three-month after admission. A five-year study included almost all hospitalisations in one larger city and eight counties. The total population covered was 2.87 million. The census provided socio-economic stratum, type of housing accommodation, and education. The causes-of-death registers provided the date and of death. The observed outcome percentage was recorded and compared with percentage of deaths expected in a hospital.

In Sweden during a one-year period, all deceased patients from the age of 15 to the age of 74 were classified according to age and according to four types diagnosis. Simple random samples were drawn from each class. For each case, the hospital records, including the autopsy report were obtained from the institution where the death certificate was issued. The records included the text and reports of laboratory tests and clinical treatments. An expert panel with representatives from internal medicine, social medicine, and forensic medicine examined the chain of events leading to death and issued a new death certificate, blinded to the original death certificate and coded as, and compared with the original.

Results
We found significantly higher observed than expected mortality more often in small hospitals versus middle sized and large hospitals with respect to the following causes of death: asthma, cardiovascular diseases, infectious diseases, malignant tumours, gastrointestinal diseases, injuries and poisoning.

The analyses of the cause of death certification showed that the agreement between the expert panel and the issuing doctors decreased with the age of the patient. In addition, issuing doctors aged 45 to 54 received the best grading. Institutes of forensic medicine were the ones most accurate in identifying the underlying cause of death. The most rural districts had the lowest degree of agreement and larger hospitals had significant greater degree of agreement than small hospitals and were much better than nursing homes and general practitioners. The largest differences were found as to pneumonia while breast cancer deaths was very accurately certified. Autopsy cases were not coherent with larger accuracy than other cases.

Conclusion
Current administrative data offer substantial practical advantages for widespread quality assessment.
THE ELUSIVE QUALTY OF HEALTH CARE

Boo Svartbo

Umeå 2000
ABSTRACT

Background
Wide variations in hospital inpatient death rates have been observed over many years, and concerns have been expressed that such variations could reflect important differences in the quality of medical care. The aim of the study was to develop and apply a method to monitor the quality of medical care and to study the accuracy by which the cause of death was determined.

Methods
From 31 hospitals, data on every discharge were linked with other routinely collected data. In addition, inpatient hospital death rates were analysed as well as data three-month after admission. A five-year study included almost all hospitalisations in one larger city and eight counties. The total population covered was 2.87 million. The census provided socio-economic stratum, type of housing accommodation, and education. The causes-of-death registers provided the date and of death. The observed outcome percentage was recorded and compared with percentage of deaths expected in a hospital.

In Sweden during a one-year period, all deceased patients from the age of 15 to the age of 74 were classified according to age and according to four types diagnosis. Simple random samples were drawn from each class. For each case, the hospital records, including the autopsy report were obtained from the institution where the death certificate was issued. The records included the text and reports of laboratory tests and clinical treatments. An expert panel with representatives from internal medicine, social medicine, and forensic medicine examined the chain of events leading to death and issued a new death certificate, blinded to the original death certificate and coded as, and compared with the original.

Results
We found significantly higher observed than expected mortality more often in small hospitals versus middle sized and large hospitals with respect to the following causes of death: asthma, cardiovascular diseases, infectious diseases, malignant tumours, gastro intestinal diseases, injuries and poisoning.

The analyses of the cause of death certification showed that the agreement between the expert panel and the issuing doctors decreased with the age of
the patient. In addition, issuing doctors aged 45 to 54 received the best grading. Institutes of forensic medicine were the ones most accurate in identifying the underlying cause of death. The most rural districts had the lowest degree of agreement and larger hospitals had significant greater degree of agreement than small hospitals and were much better than nursing homes and general practitioners. The largest differences were found as to pneumonia while breast cancer deaths was very accurately certified. Autopsy cases were not coherent with larger accuracy than other cases.

**Conclusion**

Current administrative data offer substantial practical advantages for widespread quality assessment.
THE ELUSIVE QUALITY OF HEALTH CARE

Av

Boo Svartbo

AKADEMISK AVHANDLING
som med vederbörligt tillstånd av Rektorsämbetet vid Umeå Universitet för avläggande av Medicine doktorsexamen kommer att offentligen försvaras i Sal D, Tandläkarhögskolan, by 1D, 9tr, Norrlands universitetssjukhus, Fredagen den 3 november 2000 kl 13.00.
Fakultetsopponent: Professor Gudjón Magnússon, Nordiska Hälsovårds­högskolan, Göteborg.

Avhandlingen baseras på följande arbeten:


INTRODUCTION

Wide variations in hospital inpatient death rates have been observed over many years, (1-4), and concerns have been expressed that such variations could reflect important differences in the quality of medical care available in different hospitals (5,6). Until now, research has provided contradictory evidence about the relationship between hospital mortality and quality of care (6-9). While differences in patients' age and severity of illness may explain some of the variation in hospital death rates, adjustment for age, sex, and severity leaves a large amount of unexplained variation (10-16).

National death certification was first introduced in England and Wales in 1837. The Registrar General was created, and death registration became compulsory as a way legally to prove death and as a way to improve mortality data. The complex legislation concerning certification, registration and disposal of the dead was reviewed in the Brodrick report in 1971 (17). Originally, anybody knowing the deceased could declare the cause of death, but by 1845 the Registrar General provided forms to registered doctors. By 1874, only registered medical practitioners could officially certify death and penalties were imposed for not registering a person’s death. ‘Violent’ or ‘suspicious’ deaths – and from 1885, deaths from ‘unknown cause’ and ‘sudden and uncertified’ deaths were to be referred to the coroner by the registrar of Births and Deaths. If no inquest was held and an official certificate was unavailable, death registration used the best available information. Certification by unregistered medical practitioners, unqualified midwives, and chemists caused continuing public concern about uncertain deaths from ‘poison, violence or criminal neglect’. It was not, however, until administrative changes in 1914 and legislative changes in 1926, that the coroner referral system finally dealt with deaths not certified by an attending registered medical practitioner. In Sweden doctors certified the cause of death from the beginning of the 20th century.

Comparisons of hospital inpatient death rates, published annually in the United States as league tables, have resulted in lively discussion and debate about their compilation and usefulness (13, 18-20). Meaningful comparisons of hospital death rates require adjustments for severity of illness, often indirectly measured by length of hospital stay, age, diagnosis, and type of admission. In the United States suitable standardised hospital death rates are used both as indicators of quality of care and as a way to set standards.

Medical care is the interaction between physician and client. In theory,
this interaction is divisible into two domains. The technical performance, i.e. medical knowledge and technology to maximise benefit and minimise risks taking account of the preferences of each patient. The other domain is the management of personal relationship with the patient in a manner that conforms to the ethical requirements, social conventions and the legitimate expectations and needs of the patient. The two are, however, highly correlated and in practice not possible to separate.

One way to evaluate medical care is to use implicit criteria i.e. a senior skilled physician knowing all particulars in the case is reconstructing the case in his or her mind and conduct the care that he or she would have recommended under the circumstances. Implicit criteria are time consuming, costly and the qualification of any assessor may be challenged. Another way is to use explicit criteria, i.e. guidelines from leading experts or derived from the average practice of physicians in a community. As opinion about diagnostic decisions and treatment differs the standard should be subclassified into reasonably homogeneous “categorical” and “contingent” classes. The criteria are lists of procedures that should be or may be performed in some cases but not in others. A further refinement is to specify for each procedure the frequency with which it is expected to be performed (21).

Most systems for monitoring the quality of care employ a three-stage approach. A common way is to start with the structure, i.e. the number, mix and qualification of the staff, how the staff is organised and governed, but also space, equipment and other physical facilities. Next part is the direct observations, a process describing how the cases are found, diagnostic work, the treatment and the rehabilitation. This “physiology” or “pathophysiology” of a service is seen as the core of medicine (21). The most simple way of measuring the process is to count number of visits and in-hospital days. The third stage is a study of the outcomes of the services and measured as influence on sickness, death rates, handicap, secondary effects, progress of diseases, specific diagnoses results, post-surgery death rates, sickness. Other results are influences on knowledge, attitudes and health behaviour. It is also possible to measure changes in quality of life.

Monitoring by following the structure, the process or by measuring the results are three ways possible to use. All are based on the condition that there is a connection between the medical care and the results. If the correlation is strong, measuring the structure or the process is as good as measuring the result (21).
Two methods of assessing the quality of care can be put in a separate category. The first may be called the “trajectory” method and that is to follow the patients from the time they come for out- or inpatient care to some time after their care presumably ends. The other is to select a number of diagnosis or conditions as indicators of the quality of care in each subpart of the healthcare system. Each diagnosis functions as a “tracer” and the set of tracers can be considered to provide what is analogous to set of carefully selected soundings of an unexplored terrain.

Studies in the US have shown that an important factor for quality of care is the type of hospital. The best care is provided in university hospitals and less good in private hospitals in cities and in small hospitals wherever they are. It is of importance if the doctor is a specialist as long as he or she is working in his or her specialist area. Doctors in large groups give better care, mainly because they use more of specialist caregivers. But most of the variations of results used to be unexplained. Analysing the quality of care may be done by combining the structure with results/outcome and, to give a closer look at the process, by adding tracers or “trajectories”. Another way is to combine process with results/outcome.

The objective of this study is to monitor the Swedish health care system by applying and developing a method by means of tracers with small regional variations to find out whether the patient survives three months after admission. Another aim is to study relevance and variations in the cause-of-death statistics. We found some tracers that often differed among the hospitals mostly the small hospitals were worse off. We also found differences in relevance regarding cause-of-death for the different hospitals.

**METHODS**

We examined the quality of health care by focusing on the structure of hospitals. We also examined the quality of health care by focusing on the process of treatment and its outcomes.

**Hospital structure and outcome**

We invited eight counties and one of the larger cities in Sweden to participate in the project. The study based on structure included practically all hospitalisation in those areas from 31 hospitals. The total population covered was 2.87 million. The counties and the city of Gothenburg provided Statistic Sweden with their computerised records of in-patient stays for the
period from 1987 through 1993, except for one province which provided
data for the years from 1987 through 1992 and another province which
provided data for the years from 1987 through 1991. Each record identified
the hospital and the department, the patient’s personal identity number,
dates of admission and discharge and the diagnoses. For the purpose of this
study, hospitalisation was defined as patients with at least one hospital stay
during a calendar year. If patients had more than one stay per year, one of
these was drawn at random.

The data for the study were obtained from the hospitalisation register of
each county, the national causes-of-death registers, and the national census
registers. From the 1985 and 1990 censuses the following information was
collected for each case: marital status, cohabitation, socio-economic stra­
tum and type of housing accommodation. The causes-of-death registers
provided the date and time of death. The registers were linked by the patients’
personal identity numbers, which subsequently were erased. Thus, for each
sampled hospitalisation, data on outcome variables and control variables
became available for the analysis. The outcome variable, for this study, was
whether the patient survived for three months after admission to the hospi­
tal.

We selected 27 diagnoses as tracers each constituting a dependant variable
in the outcome analysis. The diagnoses were selected because they were
common in all types of hospitals. The specific groups were delineated after
inspecting the rates of admission in the nine provinces in order to delete
those diagnostic groups with considerable geographical variation in hospi­
tal admission that could not be attributed to geographical variation of disease
incidence.

Two factors contributing greatly to variations in registered post-hospital
mortality are varying diagnostic customs and diagnostic criteria. Most often
these influence the average severity of a disease in hospitalised patients. The
proportion of patients suffering from the disease in the population, who
are hospitalised, largely determines the severity. We did not take into account
the possible variation in incidence or prevalence of disease in the catchments
areas of the hospitals when selecting highly specific diagnoses except for
myocardial infarction. When selecting tracers, we consequently used
hospitalisations per population as a measure of diagnostic criteria and criteria
for the need of in-hospital care. The relative range of these measures over
provinces, i.e. maximum minus minimum as a percentage of the average, constituted our range measure. The coefficient of variation between counties was divided by the coefficient of variation owing to random variation. The resulting ratios were tested regarding deviation from 1. The criteria for selection of diagnoses were that the absolute number of hospitalisations in the material should preferably exceed 1,000, and the ranges of incidence and prevalence should converge and their coefficient of variance over the random variation ratios should preferably not exceed 5 and not deviate highly significantly from 1. For each group of diagnoses at each hospital, the number of hospitalisations and the observed outcome percentage were recorded. This was the proportion of hospitalised patients who actually died during the follow-up period.

These data were contrasted against the corresponding expected percentage of deaths. For the diagnosis group in question, the expected figure was computed on the basis of all the hospitals participating in the study. The expected proportion of deaths was defined as the hypothetical percentage of deaths among the patients at the hospital in question, given that their mortality was the same as that among all patients with similar characteristics in the whole population at all the hospitals in the study. This means that patient characteristics, such as age, previous hospitalisation, living conditions, etc., were taken into account when computing the expected percentage.

The 31 hospitals included in the analysis were grouped into three classes according to the number of specialties in the hospital and the number of patients treated per year. Three independent variables emerged: university hospitals and large provincial hospitals, small provincial hospitals and large local hospitals, and small local hospitals. A proportional hazards model was applied (22), to verify the influence of hospital class on mortality, which took into account confounders like a patient’s socio-economic or demographical background. If there were several hospital stays for one patient during one calendar year, a random stay was chosen in order to prevent bias that might occur when a patient is admitted to a hospital on more than one occasion.

For each year and each tracer of diagnosis, a logit model for the outcome in terms of the control variables was estimated for the population of all hospitalisations in the participating counties. The estimation of the model was made by PROC LOGISTIC of the SAS System (SAS Institute Inc., NY, USA, 1990).
The control variables included (dichotomised variables, where not otherwise stated):

- age, expressed in years exceeding age 45 (continuous variable);
- age, ten-year intervals below 45 (continuous variable);
- gender;
- living in small house/apartment;
- living in nursing homes and similar establishments;
- married or cohabiting;
- born in non-Nordic country;
- worker, unskilled;
- emergency/planned admission;
- admitted within three days after discharge from other hospital;
- certain lethal conditions, N codes: 141-160, 162-172, 174-208; 412-414, 426-429; 571-572, 582-583, 585-587 (ICD-9);
- certain chronic conditions, N codes: 249-298, 300E, 303, 304, 714, 715, 723, 724, 726, 728 (ICD-9);
- days in hospital previous year (continuous, transformed by square root);
- time since last hospitalisation (continuous, transformed by inversion);
- not in census.

For each patient the model yields an expected outcome probability \( P(x) \) where \( x \) is the vector of control variable values for the patient. Finally, the observed and the expected outcome proportions for hospitals and groups of hospitals are obtained by aggregation, as averages of the corresponding outcome for the patients. This step is also implemented in the SAS System.

The aggregation step also includes a computation of the estimated variations in the difference between the observed and the expected proportions, needed to express the random uncertainty. We used a “smaller” estimated variance that merely expresses that variability which is due to the limited number of hospitals. (Not use a “larger” variance that also includes a further variance term expressing a natural variability between hospitals.)

Notable approximations involved here are that, we consider only the variability in the expected term and disregard that in the observed and we consider the hospitalisations as independent, even though one patient may have more than one hospitalisation in different calendar years. We also consider previous hospitalisation as a control variable.
Process and outcome

In order to see behind the figures in our findings based on the structure of hospital care, we investigated the process in one medium sized hospital for 1,036 patients for three years of hospital records with stroke as diagnosis. The records were checked against a few explicit criteria of good care as CT-scanning, medication, ADL-status, and housing specified. The hospital in question was according to our formula average compared to other hospitals of similar kind.

Another aspect of the process of care was the accuracy by which the cause of death was determined. The diagnostics means that a sequence of events is coded and an underlying cause of death is determined. It is a relatively easy process to evaluate. A population-based study was performed which included all deceased in Sweden during one year in the interval 15 - 74 years of age, in all 36,642 cases. The population was stratified according to age (15-44, 45-64, 65-74 years) and into four diagnostic classes: pneumonia N codes: 480-486 (ICD-9), cardiovascular disease N codes: 410-438 (ICD-9), breast cancer, and other causes. Simple random samples were drawn from each stratum. For each case, the records of the stage when death occurred and the autopsy report, when relevant, were obtained from the institution where the death certificate was issued. In the youngest age class 50 cases were drawn (all 36 classified as pneumonia) and 100 cases in the two other age groups in the four diagnostic classes were drawn.

As for pneumonia, we had 236 cases recorded as underlying causes of death. These categories denoted viral pneumonia, pneumococcal pneumonia, other forms of bacterial pneumonia, pneumonia due to other specified organisms, acute interstitial pneumonia, unspecified bronchopneumonia and unspecified pneumonia. The non-pneumonia stratum consisted of a self-weighted sample of all the non-pneumonia diagnoses. All cases in the youngest age-class were sampled and 100 cases from each of the two older age classes (sampling rates 0.44 and 0.17) were sampled. Another 1,140 non-pneumonia cases comprised the self-weighted sample from the three other categories; breast cancer, cardiovascular disease and other diagnoses. In 10 pneumonia and 61 non-pneumonia cases, no record could be obtained and these cases were therefore removed from the sample, leaving 226 pneumonia deaths and 1,079 non-pneumonia cases. Most records came from hospitals, 83 came from health centres or nursing homes, and 272 from Institutes of forensic medicine. Autopsies of persons who
died at home or otherwise outside institutions were often performed at the latter institutes.

Using after-stratification, the underlying cause of death from cerebrovascular diseases (CVD) was studied. This sample contained 75 cases with CVD as the underlying cause of death, while 1,301 were non-CVD cases. In 71 cases, no record could be retrieved and the cases were removed from the sample, leaving 72 deaths from CVD and 1,233 non-CVD deaths. The majority of records and autopsy reports came from hospitals, 83 from general practitioners, and 272 from Institutes of Forensic Medicine. The large number of medico-legal autopsy reports is explained by the fact that most people who died outside institutions or from unnatural causes were subject to a medico-legal autopsy at the time of the investigation.

Of the 36,642 cases, a simple random sample was drawn, stratified as mentioned above for the class ischemic heart disease (IHD). There were 250 cases with IHD as the underlying cause of death and 1,126 non-IHD cases. The 250 in the IHD group, 162 cases were classified as acute myocardial infarction (AMI) and accordingly 1,214 cases were classified as non-AMI. In 71 cases, no record could be retrieved and these cases were excluded, leaving 239 deaths from IHD and 1,066 non-IHD deaths. When studying AMI deaths, nine cases had to be excluded because clinical records could not be found, leaving 153 AMI and 1,152 non-AMI cases.

The number of originally registered breast cancer cases in the sample was 250. In 6 cases no record could be retrieved, leaving 244 breast cancer cases. For non-breast cancer cases, the sample contained 1,126 cases, and since 65 cases could not be retrieved, 1,061 non-breast cancer cases were actually sampled.

An expert panel with specialists in internal medicine, social medicine, and forensic medicine was convened to determine the chain of events leading to death. This panel used implicit criteria to determine the underlying cause of death. The panel was ‘blinded’ to the original death certificates. The records of each case were scrutinized and assessed and the results were reported to the panel. Each case where the reported diagnosis was not identical for all specialists, was discussed in the panel until consensus was reached. A new death certificate was then issued by the panel and coded by Statistics Sweden according to the 9th revision of the International Classification of Diseases, WHO (1977).
Only when the clinician had reported the course of symptoms and physical signs of pneumonia in a patient free from underlying serious pulmonary disease, or when the patient did not have generally impaired health, the panel noted ‘pneumonia’ as the underlying cause of death. If the physician reported that when the patient, prior to the onset of the symptoms of pneumonia, had a worsening disease or injury impairing his or her resistance to the infection, or if the patient was bedridden, due to a condition other than pneumonia, then the progressing disease (or the disease necessitating bed-care) or injury was discussed as the possible underlying cause of death.

Consensus within the panel was achieved in all cases. A small minority of cases proved difficult to assess, e.g. cases with a generally worsening disease that had remained stable for a long time. Obvious progression or exacerbation qualified a disease as constituting the underlying cause of death. If the disease had been stable for long time, the disease was noted as a contributory cause rather than the underlying cause. In instances of alcoholism, the autopsy often provided guidance, especially in cases of lobar pneumonia. The underlying cause of death, according to the original death certificate, was compared with the panel’s underlying cause of death the cause selected according to the ICD instructions. The effect of ICD coding rules was demonstrated by comparing the panels underlying cause of death before and after the coding rules had been applied. The age of the issuing doctor, the speciality in which he or she worked and the type of hospital, were all coded from the registers of doctors and hospitals, respectively. A retest was made more than one year after the first assessment.

The degree of agreement between the “panel’s underlying cause of death” and the official statistics was described with Cohen’s kappa (23). Kappa < 0.00 indicated “poor” agreement, kappa values of 0.00 – 0.19 and 0.20 – 0.39 and the following quintiles indicated “slight, fair, moderate, substantial, and almost perfect” agreement (24, 25). Statistics were weighted when appropriate. The SYSTAT program pack (Systat Inc, Evanston, Illinois) was used.

RESULTS
Hospital structure and outcome

From the in-patient records collected from 9 provinces out of 25 in Sweden for the years from 1987 through 1993, 3.6 million stays were analysed. The question being whether survival 3 months after admission was dependent
on where the care was given. We defined some diagnostic classes as more specific and with the same diagnostic criterion across provinces.

There are altogether 27 inpatient diagnostic classes. The “specific” diagnoses were classified into chapter groups, e.g., “specific neo-plastic disease”. We found significant higher observed mortality than expected more often among small hospitals than among medium and large size hospitals. Medium sized hospitals seldom demonstrated such excess mortality. The hospital types differ little regarding outcome when it comes to the total mortality, but there were six diagnostic classes of higher interest (Table 1).

<table>
<thead>
<tr>
<th>Disease</th>
<th>Hospital size</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Large N = 8</td>
</tr>
<tr>
<td>Asthma</td>
<td>3</td>
</tr>
<tr>
<td>Cardiovascular diseases</td>
<td>2</td>
</tr>
<tr>
<td>Infectious diseases</td>
<td>0</td>
</tr>
<tr>
<td>Malignant tumours</td>
<td>4</td>
</tr>
<tr>
<td>Gastro intestinal diseases</td>
<td>2</td>
</tr>
<tr>
<td>Injuries and poisoning</td>
<td>3</td>
</tr>
</tbody>
</table>

The effects of transferred patients may influence the results. To eliminate this effect we have computed the outcome not for the last visit during a calendar year but for a randomly drawn visit during the calendar year.

**Process and outcome**

**Stroke as a tracer**

The results from our quality control of the structure analyse were mirrored by data on stroke from a medium sized hospital in the hospitals files. About half of the case books were hard to read, a quality indicator in itself. Each case is unique and there are many reasons for individual treatment. The general quality level was therefore difficult to judge. It looks though as if diagnostic procedure has the best quality, better than the medical treatment,
while plans for rehabilitation of those who were sent home were rare (see paper one).

**Pneumonia as underlying cause of death**

Pneumonia as the underlying cause of death (UCD) was present in 2% of the total number of deaths and in 6% as contributing cause of death. Among the 226 patients, whose official UCD was pneumonia, only 53 actually had pneumonia as the UCD according to the panel. On the other hand, of the 1,079 patients originally classified as non-pneumonia cases, only 5 had pneumonia as the UCD. The corresponding weighted proportion originally recorded as pneumonia was 2.3% of deaths (true or false pneumonia) but according to the panel 1.9% of deaths were falsely reported (false pneumonia). On the other hand, 0.2% of deaths were in fact caused by pneumonia even though they were recorded in the register as non-pneumonia cases (false non-pneumonia).

According to the panel, 0.7% (true pneumonia and false non-pneumonia) of deaths were due to pneumonia and 0.4% were included in the official statistics. This indicates a high degree of sensitivity, a low specificity, and a very low prediction value of pneumonia in the official register. Cohen's kappa was 0.32, indicating fair agreement. The underlying causes of death among the 173 false pneumonia cases were, according to the panel, in many cases heart disease, malignant neoplasm, cerebrovascular disease, senile dementia, alcoholism, injuries, and poisoning. In fact, a whole range of diseases and injuries were represented.

A reliability study of the judgements made by the expert panel was performed 3 years after the first assessment by the panel. A simple random sample of 100 cases were selected for re-test. Four of these 100 cases had to be excluded from the study as the records were not available; therefore, 96 cases were analysed. The same procedure as in the first study was applied to reach consensus in the team for individual cases, and on both occasions the coding was carried out according to the ICD rules (ICD-9). The agreement at the three-digit category levels of the code for underlying cause of death between the two studies was 77%.

Within the age range of the sample, 15 – 74 years, agreement between panel and the issuing doctor decreased with the increase in the patient's age. The degree of agreement for those below age 45 was substantial (kappa
0.65) while the degree of agreement for those 45 – 74 years old at death, agreement was fair (kappa 0.25 – 0.30).

To assess the importance of the experience of the individual certifying doctor we checked the agreement between the two sets of certificates in relation to the age of the first certifier.

Doctors aged 45 – 54 received the best grading, viz. ‘moderate’ agreement. The importance of the certifier’s professional environment was demonstrated by the influence of hospital size and the clinical speciality of the ward from which the certificate was issued. The largest hospitals showed better agreement. The doctors at the Institutes of forensic medicine were the one most accurate in determining pneumonia as an underlying cause of death. The surgical specialities and thoracic medical wards were often found to inaccurately record pneumonia as being the underlying cause of death. There were also differences between districts, with the most rural districts having the lowest degree of agreement between the registered cause of death and the “true” one according to the panel.

Among the 366 patients autopsied in hospitals, the degree of agreement between the panel and the issuing doctor was slight (kappa 0.19), while among 287 patients autopsied at Institutes of forensic medicine, agreement was substantial (kappa 0.63). For non-autopsied cases, agreement was fair except in surgical specialities where it was worse. Certificates issued at departments of clinical pathology showed fair agreement with the panel’s opinion concerning pneumonia as the UCD (kappa 0.23). Those issued by specialists in internal medicine showed a greater degree of agreement in cases where an autopsy was performed. Other specialities exhibited a contrary tendency. Autopsied cases showed poorer agreement between the official cause of death and that of the expert panel.

**Stroke as underlying cause of death**

The overlap between the official register’s underlying cause of death (COD) and that of the panel was 72% at the 3-digit level. The official underlying cause of death from cerebrovascular diseases (CVD) was 72 cases (out of 1,376) in our sample, while 93 were deemed to have CVD by the panel. On the other hand, of the 1,233 cases originally reported as non-CVD, the panel deemed 57 as CVD. The corresponding weighted proportion originally stated as CVD (true and false CVD) was 8.1 % of deaths, but according to the panel 0.9 % (false CVD) were stated otherwise. On the
other hand 1.7% of the deaths were in fact caused by CVD (false non-CVD) even though they were recorded in the register as being non-CVD cases. The CVD deaths constituted, according to the panel, 8.8% (true CVD and false non-CVD) and the majority of them were present in the official statistics, 7.1% (true CVD).

This leaves us with a sensitivity of 81% and an 88% predictive value of CVD in the official death register. Kappa was 0.83, indicating an ‘almost perfect’ agreement. The age range of the sample excluded the very young and the very old. However, sensitivity and predictive value were worse in the oldest age group (65–74 years of age) than in the two younger groups, even though Kappa was for all groups either ‘substantial or almost perfect’ (0.78–0.94). The type of hospital was rather significant, the large hospitals again having a significantly greater degree of agreement than the intermediate and smaller hospitals, and far better than nursing homes and general practitioners.

Regional differences were found for CVD, where the most urban areas had the best agreement between the registered cause and the “true” one, according to the panel. The agreement between the registered cause and the true one for CVD was ‘almost perfect’, irrespective if an autopsy was performed or not. However, certificates issued at the Institutes of forensic medicine had the greatest degree of agreement, followed by clinically autopsied cases and cases where no autopsy was performed. The earliest disease or injury in the sequence notified by the panel was compared with the true underlying cause when the code rules of the ICD were applied. Four cases of hypertension before coding became cardiovascular disease after coding. Sixteen cases of general arteriosclerosis became 6 cases of chronic ischemic heart disease, one case of acute myocardial infarction, 6 cases of cardiovascular disease, one aortic aneurysm, one gangrene and one case of heart failure when coding rules had been applied.

Ischemic heart disease and acute myocardial infarction as underlying cause of death (UCD)

Patients whose official underlying cause of death in register was ischemic heart disease (IHD) were 239, of whom 215 were deemed by the panel to have IHD. Within the IHD group, 153 deceased had acute myocardial infarction (AMI) as the official underlying cause of death and out of these, 113 were deemed by the panel to have AMI. On the other hand, in the 1,066 cases originally reported as non-IHD, IHD was determined the true
underlying cause in a number of cases which, after applying the sample weights, corresponded to 34 cases, while in the 1,152 originally reported as non-AMI, a corresponding number of 38 were considered by the panel to be true AMI cases.

After the appropriate weighting this leaves us with a high degree of sensitivity and specificity and a good predictive value of IHD in the official death register. However, regarding AMI, these measures were not that good. When evaluated with kappa, the agreement was substantial.

The death certificate includes all causes: direct, intermediate, underlying, and contributory. The condition which should have been stated as the underlying cause, was sometimes designated as direct, intermediate or contributory. Almost all cases of IHD were present in the register, if all causes on the certificate were taken into account.

The experience of the doctor who issued the original death certificate was investigated by dividing the doctors into different age groups. No significant differences between the old and the young doctors could be found, either for IHD or for the more circumscribed diagnoses AMI. The Institutes for forensic medicine were the ones most accurate in setting both IHD and AMI as the underlying cause of death (kappa 0.95 and 0.84 respectively). Regional differences were found for AMI, the most rural areas having a lower degree of agreement between the registered cause and the “true” one, according to the panel. For patients with the diagnosis IHD, no significant differences were found.

The agreement between the registered cause and the “true” one for both IHD and AMI was “substantial” or “almost perfect”, irrespective of whether an autopsy was performed. Certificates issued at the department of clinical pathology showed “substantial” agreement with the panel’s assessment as the underlying cause of death for both IHD and AMI (kappa 0.79 and 0.78 respectively).

**Breast cancer as underlying cause of death (UCD)**

Most cases where the registered underlying cause of death was breast cancer were deemed to be correctly recorded by the panel, most cases registered as non-breast cancer cases were also correctly reported. False negative cases were 3 out of 1,061, and false positive cases 5 out of 244. The three false
negative cases of breast cancer had been recorded as pneumonia (2 cases) and pulmonary embolism. All three had breast cancer registered as a contributory cause of death. One deceased had been autopsied. The 5 false positive cases were, according to the panel, caused by other cancers (3 cases), ischemic heart disease and pneumonia.

**DISCUSSION**

The series of studies in this thesis takes into account various aspects of the quality of medical care. The study of all discharged patients in 9 provinces concentrated on structural aspects, used some discharge diagnoses as tracers, and used mortality as the outcome measure. In our study of structure and outcome, there may have been considerable random variability in the outcome, owing primarily to the limited number of hospitals.

The studies of all deceased in the country below 75 years of age were performed partly to indicate how valid the mortality measure was, but they were also used to investigate the diagnostic process and then to measure accuracy of death certificates as an outcome measure. The structural aspects regarding the diagnostic process, were in focus.

**Mutual adjustment of hospitals.**

Hospitals vary in their death rates, sometimes widely. The rationale for risk adjustment is to remove two sources of this variation, the patient and disease characteristics leaving residual differences to reflect quality. The underlying assumption is that outcomes result from a complex mix of factors: patient outcomes equal the theoretical possible potential effectiveness of treatments plus patient risk factors that affect response to treatment plus quality of medical and nursing care plus random chance. Controlling for patient risk allows us to begin isolating quality differences. But “risk adjustment” is a meaningless phrase without first answering the question: risk of what? Identical risk factors may have different relationships to different outcomes (i.e., an attribute suggesting high risk for one outcome may indicate low risk for another outcome). For example, when refining diagnosis related groups (DRG) to improve their sensitivity to severity for hospital payment, researchers found that medical patients dying within 2 days of admission had relatively low-cost hospitalisations (26).

Many diverse patient attributes affect risks, including age, sex, acute physiological stability, principal diagnosis (i.e., the main reasons for the
hospital care) and its severity, the extent and complexity of co-morbid illness, functional status, psychosocial and cultural factors, socio-economic characteristics, and preferences for specific outcomes (27). Measuring certain attributes is challenging and potentially expensive (requiring patient surveys or extensive medical record reviews). For example, physicians and patients may have divergent perceptions of the patient’s functioning (28). In addition, some factors affect outcomes not because of physiology but because of differences in the way people are treated. In the US it has been shown that African-American patients and uninsured patients (29) may receive lower-quality hospital care than others. Using risk-adjusted outcomes to evaluate quality, adjusting for race or socio-economic standing could mask these important differences.

Generally, an individual’s socio-economic standing, medical insurance, and network influence and individual’s access to quality medical care. Our 15 control variables might as well.

**Comparing hospital mortality rates**

Finding differences across severity measures at the level of individual patients led to this question: would judgements about whether hospitals look particularly good or bad differ using different severity measures for risk adjustment? The answer to this question is yes – sometimes. The immediate impression is that severity measures often flagged different rates, but different severity measures also frequently flagged different hospitals. No clear pattern emerged nor was agreement better for flagging the best 10% of hospitals than for the worst 10%. Again results differed by condition. Hospitals vary in their unadjusted death rates.

According to other findings (30-32) severity fail to explain fully these differences. As shown in California, differences across hospitals in accuracy of data can account partially for discrepancies in risk-adjusted death rates, using measures based on discharge abstracts (33). The central question remains unresolved: does severity adjustment isolate that residual quantity, namely quality-of-care differences, across hospitals? A definitive answer is unlikely any time soon. The research required is expensive and time-consuming, but methodologically not very complex. Only a handful of studies have addressed this question, and most have produced equivocal results.

Studies have failed to find any relationship between risk-adjusted mortality rates and hospital quality. Using rigorous severity and quality measures
Kahn et al (34) found that hospitals flagged as having unexpectedly high age-, sex-, race, or disease-specific death rates did not have worse quality than other hospitals. Another study found no association between quality of care and observed-to-expected mortality ratios calculated by the US Department of veterans affairs using discharge abstract-type data. (35). An important difference between our work and the US reports is that is has been possible for us to control for more variables than they have.

**Differences in data quality**

Variability in data quality across hospitals compromises the utility of administrative data for comparing hospital performance. Because hospitals code with different degrees of thoroughness and accuracy, one cannot tell whether coded differences reflect true differences among patients without further study. The re-abstraction study from California of 974 patients with heart problems found variations in coding accuracy across hospitals. Overall, at least one clinical risk factor was missing for 65% of the patients. In contrast, 31% of records contained at least one unsupported risk factor; this over-coding was more common at “low” mortality hospitals than at “high” mortality hospitals. Over-coding ranged from 10% at one “high” mortality hospital to 74% at one “low” mortality hospital. Variation in coding accuracy explained part of the differences between “high” and “low” mortality hospitals (33).

Knowing when events happen is crucial for assessing quality. However, the information in hospital discharge abstracts is retrospective: diagnoses are assigned after discharge. Discharge diagnoses reflect conditions that were diagnosed or treated at any time during the entire admission, regardless of when they occurred.

**Administrative data and quality measurement**

Despite the concerns discussed above, it is possible to patch together potential indicators of quality from administrative data. Even a modest administrative data-based quality standards may not, however, pass the first hurdle of validity. Gaps in clinical information, questions about coding procedures, and the billing context restrict the ability to derive valid appraisals of quality from administrative data. The assessment of validity, however, is not a yes-or-no proposition: Shades of grey are possible but administrative data may be very useful for descriptive purposes (36). For quality assessment, an intermediate position holds that administrative data are a useful screening
tool that highlights areas in which quality should be investigated in greater
dept by using detailed clinical information.

**Hospital structure**

We have traced a cohort of people admitted to hospitals, but the
delineation of the cohort is vague, because it partly depends on the health
resources in that area, the organizations of health services, differences in
the primary health care, out-patient care at hospitals, and traditions of
admittance to hospitals among the doctors outside the hospital. However,
practically all in-hospital service provided for the population in the 9
provinces in our study is covered by the registers. Furthermore, the outcome
is presented for narrow diagnostic classes where the variation between areas
in admittances per population was small. In that way the variations through
traditions of admittance, different diagnostic criteria, etc. were avoided to
some extent.

The meta-literature on post-hospital mortality is not encouraging. Authors
are concerned that it is difficult to assess the prognosis inherent in the
cases. Nevertheless, some important factors prolonging patients’ survival
have emerged, e.g. the size of the hospital’s teaching status, the clinical
experience of the staff and the degree and success of their
intercommunication. In these early studies, a higher level of technological
adequacy was correlated with low adjusted death rates. The score for
technological adequacy included characteristics such as the hospitals having
an intensive care unit, a pathology laboratory, rehabilitation services etc.,
or it was accredited by the appropriate medical authorities (37).

The most important single factor associated with the quality of hospital
care seems to be the nature of the hospital itself. Care is best from mortality
point of view in large, urban, university-affiliated hospitals and worse in
proprietary urban hospitals and other small hospitals, whether urban or
rural. Physician specialisation is also a factor, although its salutary influence
is less, and its influence is only when practice is confined to the area in
which the physician has specialized (21). Once specialists step outside their
domain, specialists may do worse than the generalist. The importance of
the hospital in safeguarding quality is most important for the generalist.
Besides the best hospitals, the specialization of the physicians is most
important safeguard. Physicians in the larger group practices provide hos-
pital care with lower mortality rates, but this appears to be mainly due to
the use and access of specialists by doctors in the consortium. Perhaps more
important than all these associations is the observation that a large part of the variation in performance remains unexplained.

The small local hospital has limited teaching responsibilities. The staff should benefit from the diversity of patients, but perhaps see rare cases too seldom. Communication should be easier in small hospitals, but with fewer specialists its outcome might be negligible. The crucial point is the severity of cases, which are rarely randomly allocated to the hospitals.

We have presented the outcome for diagnostic classes that seemed not to vary in diagnostic criteria. This was done to ensure that uniform degrees of severity were achieved.

That all in-patients in a province were covered and practically all inhabitants were treated in the province’s own hospitals should be a second assurance in comparing the larger hospitals.

A difficult problem in interpreting the results is the transferring of patients from small to larger hospitals during the diagnostic procedure and sometimes owing to complications, and back again after the diagnosis is established and treatment scheduled for the near future. This is most obvious, for example, after certain operations such as open-heart surgery and oncological treatment of lung cancer. When it comes to lung cancer cases, the larger hospitals probably refer patients back to smaller hospitals for the palliative treatment and this might be the cases for other diagnoses as well. We have therefore used a random visit of an individual during a calendar year to counteract some of this systematic error and one of the control variables was transmittance during last three days.

One important aspect of the differences between hospitals of different size is that diseases easily spread in hospitals. Communicable disease outbreaks are much more common in large nursing homes than in small ones (38). Because of the mortality outcome measures for these patients, such an association is concealed if the incidence of diseases during a stay is not taken into account. In our study the main disease might have occurred after the admittance.

**Hospital process**

It is not always easy to establish the chain of events leading to death for a single cause. Two or more quite different causes of death may be fully justified. The discussion by the panel scrutinizing the hospital process was sometimes lengthy lasting at the beginning of the study; for example as
regards to chronic alcoholism, since there is a wide range of other diseases and injuries that are exacerbated by alcoholism. Hospitalisation because of alcoholism, bronchial asthma, lung disease is always followed by an increased risk for death from other causes. However, for a disease or condition to be designated the underlying cause-of-death, according to the WHO-rules, it must have initiated the chain of events that led to death and not merely increased the susceptibility to other lethal conditions. Unless it is a distinct aggravation, the disease must be regarded as a contributing cause rather than the underlying cause.

The criteria for causes of death in the panel’s discussion were implicit. Where reliability could perhaps have been improved by explicit criteria, the procedure probably favoured high validity, while reliability was still acceptable. The panel’s certificate cannot always be regarded as a “gold standard” and we have therefore eschewed this term. The experts could however discuss the case amongst themselves, sometimes had more data, e.g. the autopsy record, and had more insight into the theoretical and practical problems of assessing the underlying cause of death, than had the average doctor. Nevertheless the panel accepted the attending doctor’s description of the clinical course of the disease and merely interpreted it in the form of an accurate death certificate, without seeing the attending doctors certificate. Our re-test demonstrated the accuracy of the panel’s work with very high agreement between the two events. The general outcome of studying ischemic heart disease, acute myocardial infarction and stroke, was that the results from the study based on in-hospital process are similar to the results we found starting from the structure of the healthcare system. Larger and medium sized hospitals had the best outcome.

When ageing and/or a slowly progressing disease impairing the host defences sometimes reached a threshold where pneumonia developed; the first rule was to accept such a statement in the records and to code the mentioned disease as underlying the pneumonia. There were no cases of pneumonia on the certificate was totally irrelevant nor cases where it was totally missed. Most often the latest record contained the most accurate information. When trauma was the cause of death, the external cause was sometimes recorded only on the original death certificate, to which the panel had no access. In 31 instances, the doctor gave a more specific description of the cause of death on the certificate than in the text of the clinical record. The certifiers in these cases obviously knew more detail about the patient than were stated in the clinical record.
Our findings are that the pneumonia registration is a poor indicator of certified pneumonia deaths, which were in fact caused by other diseases. There were in all 280 registered cases with pneumonia as underlying, contributing or intermediate cause of death. Out of these, we did not have agreement on 143 cases and out of 226 cases that registered pneumonia as underlying cause of death the panel found that 173 were false while 53 actually had pneumonia.

**Underlying cause of death**

The coding sometimes meant that the panels’ decision as to underlying cause of death was changed. When alcoholism was seen as underlying the coding often stopped at an earlier step in the sequence. As regards this specification of alcoholism according to the ICD rules, we agree that the consequences of chronic alcoholism are most often more informative, but it is also important to follow the total impact of alcoholism in the official statistics.

Studies comparing the main diagnosis of in-patients with the underlying cause of death have revealed disturbing discrepancies. We have used more meticulous methods when evaluating our cases, where all relevant clinical information was taken into account, sequence of events reconstructed, and a new certificate prepared. We believe that there is a need to increase the awareness that the words “due to” on the medical certificate form, implies sequences where there is no direct causation but where an antecedent condition has led to the direct cause.

It is not surprising that the original certificate was more accurate when the deceased was younger and when the doctor worked in a large hospital. But autopsy is traditionally seen as a guarantee of quality and surprisingly this was true only of medico-legal autopsies. The difference between clinical and medico-legal autopsy indicates that it is of importance when the facts revealed are correctly used. Traditional pathologists emphasise morphological changes and often refrain from describing the chain of events, while clinicians tend to describe the sequence and sometimes interpret the morphological changes incorrectly.

According to our investigation, many of the clinicians may have misinterpreted autopsy results. Another explanation could be that a case was more complicated when an autopsy was performed and that Institu-
tions of forensic medicine have more accident cases. In view of this, the death certificate for autopsied patients should perhaps be issued jointly by the pathologist and the clinician. According to Swedish recommendations, it was primarily the attending doctor who was to issue the death certificate.

The recommendation is to obtain the latest clinical record in order to trace the clinical chain of events and to use experienced official coders to interpret the chain in order to establish criteria for the underlying cause agreed upon internationally. The most important is the diagnosis in the sequence that is specific and still early in the cause of events.

**Record linkage**

We have found the use of record linkage of official registers and the multivariate approach promising as tools for monitoring the elusive quality of medical care. We also found an association between hospital size and mortality rates but our findings need to be validated by further investigations. Quality of medical care is of greatest importance. The purpose of quality monitoring is to exercise constant surveillance so that departure from standards can be indicated early and corrected. A study of patients’ diagnoses and a study of the treatment of people admitted to hospitals with high and low standardised mortality ratios could help to elucidate these findings. In such an investigation, detailed data would have to be collected to allow for accurate adjustment of case mix and analyses made on the diagnosis level. When focusing on quality of care, financial and human resources are of interest. The number of doctors and nurses per hospital bed and the number of general practitioners per head of population from which hospital admission is drawn are important factors (17). Educational level and the number of doctors and the number specialists in charge is material for further investigation.

It would also be of interest to investigate further the differences between hospitals in different parts of Sweden when it comes to hospital care. In addition, study should be devoted to how hospitalisation affects work. It is possible to select data for the beginning and end of illness by using the health insurance register of sick-listings. The register of new invalid pension and date of pensioning can be used to illuminate the basic question of differences between hospitals and their catch-area.

The Cause of Death statement is in the internationally agreed format based on the philosophically enigmatic concept of Underlying Cause of
Death (39). Because UCD is defined pragmatically as the entity initiating the causal chain leading to death, questions arise about the multi-factorial nature of disease and how the causal sequence starts. The death certificate originally invited a single entry, and although multiple entries were commonplace, only one would be coded. An international standard Cause of death statement was then introduced that allowed multiple entries, but encouraged the certifier to identify the underlying cause of death in the sequence. Strict international rules for ‘single-cause coding’ were agreed upon in 1948 (17). The legal aspects of dying have long since required review, but organizational and technical ‘fixes’ are unlikely, to improve cause of death data by themselves. Some view the current system as more about ‘policing the dead’ than producing worthwhile data. It is known that inaccurate death certification is a problem, but it is not known how to reconcile the relative contributions of diagnostics and semantic errors and how to relate these to impact.

Beyond more robust evidence, a different educational perspective is required, advancing from merely urging educational input to evidence-based interventions. Certifiers are receptive to more education about death certification, but it is not yet known which interventions are best.

The flaws in the theoretical framework of cause of death and the routine nature of death certification are unavoidable, but necessary re-considerations are needed. Certifiers need practical feedback mechanisms to improve their understanding of the construction of mortality data. Autopsy by forensic medicine specialists is not the only answer to inadequate cause of death language and coding, but the potential contribution of pathologists cannot be underestimated in assisting the death certification of autopsied and un-autopsied deaths, and facilitating good practice.

CONCLUSION

Current administrative data offer substantial practical advantages for widespread quality assessment. We found significantly higher observed than expected mortality more often in small hospitals versus middle sized and large hospitals with respect to the following causes of death: asthma, cardiovascular diseases, infectious diseases, malignant tumours, gastrointestinal diseases, injuries and poisoning. These facts must be considered when deciding the structure of medical care systems.

The analyses of the cause of death certification showed that the agreement between the expert panel and the issuing doctors decreased with the age of
the patient. In addition, issuing doctors aged 45 to 54 received the best grading. Institutes of forensic medicine were the ones most accurate in identifying the underlying cause of death. The most rural districts had the lowest degree of agreement and larger hospitals had significant greater degree of agreement than small hospitals and were much better than nursing homes and general practitioners. The largest differences were found as to pneumonia while breast cancer deaths was very accurately certified. Autopsy cases were not coherent with larger accuracy than other cases.

Mortality data are essential for many aspects of everyday public health practice. More meaningful and accurate estimates are needed, as are evidence-based interventions, educational commitment, and continuing quality assurance at all levels.
REFERENCES


