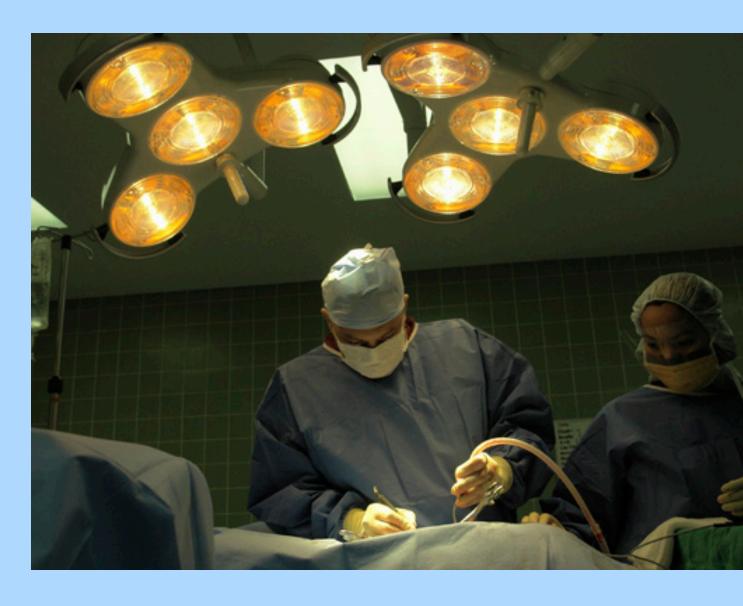The Fraser Institute

Hospital Report Card

British Columbia 2008

by Nadeem Esmail and Maureen Hazel

12 FAQs about the Hospital Report Card

The Fraser Institute
Hospital Report Card: British Columbia 2008

by Nadeem Esmail and Maureen Hazel

12 FAQs about the Hospital Report Card

Data Tables and Appendices of the Hospital Report Card

Question 1   How are some measures (e.g., deaths associated with hip replacement surgery) that do not apply to all hospitals (because they do not perform this type of procedure) handled in calculating an overall mortality score? Did you try to pick things for the Hospital Mortality Index that many hospitals did? This is particularly relevant for smaller hospitals (which may not offer a full range of services), specialty hospitals, and individual sites within a hospital corporation or city (where for quality or efficiency reasons some types of care may be concentrated in some site or another).

The Hospital Mortality Index (HMI) was developed as a result of an interest in a summary measure of patient care outcomes from our study. We started with 39 indicators and initially hoped to include all of them in an overall index that represented a composite measure of quality and patient safety. This proved impossible for a number of reasons, including the matter of coverage, where not all of the procedures and conditions are found in every hospital. To give examples from 2005/06, we have only four hospitals with data for the CABG Mortality Rate indicator and only five for Percutanerous Transluminal Coronary Angioplasty volume. Through a process of elimination (described in Appendix F), we have ended up with the HMI and its nine measures of mortality.

The HMI has a reasonably large hospital count of 25 in terms of coverage in the latest year. In terms of adequate patient record sample size, an indicator was not used in calculating the HMI if it did not represent at least 75% of patient records for that year. For example, in 2005 an indicator had to contain at least 291,785 records in order to be included in the HMI (please see Appendix F for further details on calculating the HMI, ranks and scores).

With regard to small numbers of cases at a hospital, we have used the AHRQ recommendations and do not show information where there are five or fewer cases. This is done for reasons of confidentiality and comparability. CIHI provided our database and has a standard policy of censoring any data cells that are three or fewer.

Question 2   How are the measures combined to calculate a composite score in the Hospital Mortality Index rankings? Do they receive equal weighting? This may mean that outcomes for an area that very few patients experience (e.g. a highly specialized type of surgery) are given the same weight as those for another type of care that thousands of patients experience each year. On the other hand, if indicators are not equally weighted, the score values some outcomes more than others. Previous research on composite measures in many fields has shown that changing the weights of components often has a large impact on final scores.

The measures in the Hospital Mortality Index (HMI) are equally weighted (for further information on calculating the scores, ranks and HMI, please see Appendix F). This is a standard approach of The Fraser Institute and is used in much of our research when indexing components with unknown weights. One alternative would be to weight according to the populations at risk, the denominator of our indicators. In that case, we would have the largest weight for Death in Low Mortality DRGs, as that is the broadest measure. To take the example of Anonymous Hospital 25, the

hospital with the largest number of records, this indicator has 10,807 cases in the denominator, while the other components of the HMI have between 538 (Gastrointestinal Hemorrhage Mortality) and 272 (Stroke Mortality) cases in their denominators.

This then brings up a relevant question: how important are these indicators when compared to each other? Is it just a matter of how many patients are treated? There is no obvious answer and so we really want to emphasize that the HMI is a summary measure but people should always look to the individual components and the other indicators of quality and patient safety to understand the circumstances at any given hospital. This is explicitly stated in the Introduction, the Overview and Observations, and the text that is on our website.

Question 3   How precisely are the scores being ranked? How meaningful are the differences based on the scores? Is it fair to say that indicator results tend to be more precise for larger hospitals or municipalities than smaller ones? In producing rankings, it is important to take into account the extent to which differences in indicator results may be explained by chance alone, as opposed to real differences in care. Statistical tools such as confidence intervals are often used to evaluate how likely it is that observed differences are simply the result of random variation. Likewise, to what extent does a small difference in overall score (which may make a big difference in ranking) represent a true difference in the quality of care and patient safety?

The scores and rankings are a direct result of the underlying indicator rates. We produced both in order to help people understand the relative position of the hospitals for any given indicator (for further information on calculating the scores and ranks, please see Appendix F). In addition, we have compared each institution's and each municipality's risk-adjusted rate (per indicator) to the upper and lower bounds of a 95% confidence interval (CI). This additional analysis was performed to measure the statistical significance of each result. Those below the lower CI are statistically "better than average" and those that are above the upper CI are "worse than average" (with the exception of IQIs 22, 23 and 34, where those below the lower CI are "worse than average" and those above the upper CI are "better than average").

Question 4   Whose results are reflected? Are results for municipalities based on patients treated in hospitals in that area or patients from that area regardless of where they were treated? To what extent were results adjusted for the fact that people who live in some communities (e.g. rural or remote regions) may be more likely to be transferred to specialized centres for care? Depending on how indicators were calculated, this may affect mortality and other indicator results.

The municipality results are based on the location of the patient and this is determined from the first three digits of their postal code (the Forward Sortation Area). There is no exact match of municipality to hospital, as every municipality has patients at more than one hospital. On the other side, every hospital in our study has

patients who are from different municipalities. We have made no adjustment to the municipality measures for the degree to which patients receive care at different hospitals. They are simply measures of results for patients from a given municipality, no matter where the hospital is located.

Question 5 Some types of adverse events are relatively common; others are very rare. In selecting indicators appropriate for a particular level of reporting (e.g. in this case the hospital or municipality level), to what extent has this been taken into account? For example, measures based on rare events (such as foreign objects left in a patient's body after a procedure) may not be valid for small populations, such as individual hospitals or communities.

It is true that adverse events tend to be rare and smaller places will not always see these consequences of patient care. This was a major reason why only two out of 15 of the patient safety measures were used in the overall Hospital Mortality Index summary measure for the study. It cannot be imputed that a high score on these types of indicators is due to fewer adverse events for those places with relatively low numbers of cases (this is further discussed in Appendix E). Their volume of activity may simply be inadequate to produce the inevitable adverse event. AHRQ can be referenced for work in this regard.

Question 6 How were the AHRQ indicators adapted for use in Canada? The ways that Canadian hospitals capture information about the types of health problems and procedures that patients have, differ from the methods used in the United States and have changed over time. For example, the AHRQ indicators used in this study were designed for a classification system that was historically used in some Canadian hospitals. Hospitals in British Columbia historically used a different, though similar, classification system but all switched to a new system in 2001. Comparing results based on these classification systems is challenging (e.g. because clinical understanding of conditions has changed over time and the level of detail available differs). Also, have the APR-DRGs been adapted for use with the current classification systems in use in Canada?

Appendix J outlines our entire coding methodology. Both the AHRQ indicators and 3M risk adjustment software are measured in the American 9th version of the International Classification of Diseases (ICD-9-CM), whereas in British Coumbia, the Canadian International Classification of Disease, Version 10 (ICD-10-CA/CCI) has been in use since 2001. We are dealing with over 10,000 classification codes in the 9th version and over 30,000 codes in the 10th version. In order to compensate for differences between ICD-9-CM and ICD-10-CA/CCI, conversion tables were purchased from CIHI and applied to the codes in the DAD. Each code that did not directly translate between the two classifications was individually analyzed with respect to each indicator and other codes that contained the same information. A concentrated effort was applied to this process (which took months to complete) in order to ensure the most accurate translations. All of this is discussed in the Appendices.

Question 7   Has the validity of the data used in calculating specific indicators been assessed? The quality of much hospital data is high but the extent of reporting and consistency of some data varies between institutions and over time. For example, there are known historical issues that may affect the comparability of some of the indicators cited. How likely do you think that there were data processing or coding mistakes in the data you bought from CIHI? Or, did you do the coding yourself?

CIHI's Discharge Abstract Database (DAD) contains information on hospital stays in Canada. Various CIHI publications note that the DAD is used extensively by a variety of stakeholder groups to monitor the use of acute-care health services, conduct analyses of health conditions and injuries, and increasingly to track patient outcomes. The DAD is a major data source used to produce various CIHI reports, including annual reports on the performance of hospitals and the health-care system and for seven of the health indicators adopted by the federal, provincial, and territorial governments. These data have been used extensively in previous reports on health care performance, and form the basis for many journal articles.

Once a patient is discharged, the data for the patient's stay is subject to a detailed abstraction process conducted by a health records professional and then results are submitted to CIHI. CIHI applies a comprehensive edit and correction system and inaccuracies or incorrect information are followed up on at the hospital level when the DAD is sent back to the hospitals for data validation.

The data are collected under consistent guidelines, by trained abstractors, in all acute-care hospitals in British Columbia. The data undergo extensive edit checks to improve accuracy but all errors cannot be eliminated. However, in order to produce good information about data quality, CIHI established a comprehensive and systematic data quality program, whose framework involves 24 characteristics relating to five data quality dimensions of accuracy, timeliness, relevance, comparability, and usability.

There are a number of publications that have addressed data quality issues that are discussed in our report. Of note are CIHI's reabstraction studies that go back to the original patient charts and recode the information using a different set of expert coders. [1]

The reabstraction studies, performed in the province of Ontario, note the following rates of agreement between what was initially coded compared to what was coded on reabstraction:

a) non-medical data: 96%–100%

b) selection of intervention codes (procedure codes): 90%–95%

c) selection of diagnosis codes: 83%–94%

d) selection of most responsible diagnosis: 89%–92%

e) typing of co-morbidities: pre-admit: 47%–69%; post-admit: 51%–69%

f) diagnosis typing (which indicates the relationship of the diagnosis to the patient's stay in hospital) continues to present a problem; discrepancy rates have not diminished with adoption of ICD-10-CA.

[1] Reabstractors participating in the study were required to have several years of coding experience, experience coding in ICD-10-CA and CCI in particular, experience coding at a tertiary care centre, and attendance at specific CIHI educational workshops. They were also required to attend a one-week training session and to receive a passing score on the inter-rater test.

The coding issues in points **(e)** and **(f)** do not affect our results since the most responsible diagnosis is coded with a high degree of agreement and the AHRQ indicators do not discriminate between diagnosis types. Overall, when the rates of agreement in the third year of this reabstraction study (performed on data coded in ICD-10-CA) were compared to the rates of agreement of the previous years' data (coded in ICD-9-CCP), the rates were as well as or better than the rates previously.

However, with regard to the coding of pneumonia, a potential data quality issue exists because some reabstraction coders selected pneumonia instead of chronic obstructive pulmonary disease (COPD) as the most responsible diagnosis. **[2]** This could potentially create false positive results for Pneumonia Mortality rate (IQI 20) since this indicator counts deaths due to pneumonia in situations where the primary diagnosis is a pneumonia diagnosis code. We have noted this proviso in our report.

[2] Canadian Coding Standards for ICD-10-CA and CCI 2004.

With respect to specific conditions related to the health indicators examined, those that are procedure driven (i.e. cesarean section (C section), CABG, and total knee replacement) were coded well with low discrepancy rates. The following had less than a 5% rate of discrepancy: C section, coronary artery bypass graft (CABG), hysterectomy, total knee replacement, vaginal birth after cesarean (VBAC), and total hip replacement. The following had greater than 5% discrepancy: acute myocardial infarction (AMI), 8.9%; hip fracture, 6.0%; hospitalization due to pneumonia and influenza, 6.9%; and injury hospitalization, 5.3%. **[3]**

[3] DAD Data Quality Reabstraction study. Combined findings for FY 1999/2000 and 2000/2001. Dec 2002.

Discrepancy rates were noted in conditions that are diagnosis driven: AMI, **[4]** stroke, pneumonia, and COPD **[5]** (as described above). Only the pneumonia codes are potentially affected in our report.

[4] DAD Data Quality, Reabstraction study. Combined findings for FY 1999/2000 and 2000/2001. CIHI 2002, pg 8.

Overall, according to CIHI, findings from their three-year DAD re-abstraction studies have confirmed the strengths of the database, while identifying limitations in certain areas resulting from inconsistencies in the coding of some data elements. In addition, the findings from the inter-rater data (that is, comparison between reabstractors) were generally similar to the findings from the main study data (that is, comparison between original coder and reabstractor). This suggests that the database is coded as well as can be expected using existing approaches in the hospital system.

[5] Data Quality of the DAD flowing the First year implementation of ICD-10-CA/CCI.

Question 8   How was palliative care handled? Some studies suggest that Canadians receiving end-of-life care in hospital (rather than in a hospice or at home) are more likely to die than similar patients in many other countries. Within Canada, the extent to which end-of-life care occurs in hospital varies from community to community. Deaths among these patients are not unexpected and do not necessarily indicate any issues with quality of care. Identifying these patients is complex but important, particularly when calculating results for indicators such as deaths among patients with pneumonia. For example, about 15% of in-hospital deaths were palliative-care cases in acute-care hospitals. Furthermore, a substantial number of patients who were hospitalized mainly for other conditions also received palliative care services during their stay.

The Discharge Abstract Database (DAD) is a national database for information on all acute-care hospital separations (discharges, deaths, sign-outs, transfers). Palliative patients are difficult to diagnose (and much palliative care is given outside the hospital setting) and are often identified as such only in hindsight. Only as recently as June 19, 2006 did CIHI begin instructing institutions on how to best indicate a palliative patient. Previously (and until FY2006/07 in their databases), there was no national coding standard to identify patients with terminal illness who are receiving palliative care in hospital. There is, however, an ICD-10-CA code for palliative care. In FY2005/06, the frequency of this code is 1.2% (or 4,608 of 397,769 patient records).

We hope to incorporate these improvements in the DAD in subsequent reports, as the information becomes available.

Question 9   Why is there so little in the report about cancer? Is it particularly difficult to report?

The treatment of cancer is not included in the AHRQ indicators. We chose the ARHQ methodology because it was objective, backed by a large body of research, in use in a number of jurisdictions, and based on administrative data. We have noted in the report that the indicators are for a very specific portion of hospital care: inpatient acute care. There is nothing directly related to cancer, ambulatory, clinical, ER, and so on, nor are there measures of things like patient satisfaction or the financial performance of hospitals. Comments on hospital performance should be conditioned with the fact that this is not a comprehensive survey of all hospital care. In fact, the main value is probably at the individual indicator level because that is most meaningful for a patient concerned with a certain condition or procedure. AHRQ has conducted extensive research on assessing performance on certain indicators that studies have shown are related to quality. AHRQ has identified four categories of quality indicators that appear to have relationships to the outcomes of care provided within hospitals: mortality for specific procedures, mortality for specific conditions, procedure utilization, and procedure volume. Research has confirmed that the rate of patient deaths for certain procedures and conditions may be associated with quality of care. While research can predict an expected range of patient deaths for a given procedure or condition, mortality rates above or below the expected range may have quality implications. For some procedures, research has shown that overuse, under use, and misuse (utilization) may affect patient outcomes. For certain procedures, the number of times (volume) the procedure is performed in a hospital has been linked to the patient's outcome.

Question 10   What do you see as the strengths of this report card?

The strengths of the report card are its transparency in terms of data and methodology, the detail provided at the hospital and indicator level, and the focus on patient-oriented information as well as the sample size of patient records, which over the five-year period was nearly two million.

Question 11   What about its weaknesses?

The weaknesses of the report card are its limited coverage (applying only to inpatient acute care), the unwillingness of hospitals and regions to participate by allowing themselves to be identified, and potential issues with data quality.

Question 12   What is the timeline on this project? What provinces will you add next year? When will you cover the whole country?

This is the first annual hospital report card for British Columbia. A hospital report card (2006, updated 2007) has already been produced for Ontario. We hope to prompt participation by hospitals and regions next year and to publish a report card on at least one province other than Ontario and British Columbia in 2008. We would hope to have full national coverage within five to seven years.

Question 13   Is this exactly the same methodology that New York and other states used in their hospital care surveys? Or were there some changes?

The AHRQ methodology is the same as that used in more than a dozen US states, including New York, Texas, Colorado, California, Florida, Kentucky, Massachusetts, Maryland, Minnesota, New Jersey, Oregon, Utah, Vermont, and parts of Wisconsin. There is also a recently released report by the Manitoba Center for Health Policy that used the AHRQ Patient Safety Indicators. **[6]**

  In order to use the CMS- and APR-DRG software, the DAD dataset received from CIHI required several standard modifications to account for differences in the Canadian and US coding methodologies. All standard modifications are explicitly detailed in Appendices B, C, and J.

[6] S. Bruce et al. (2006), *Application of Patient Safety Indicators in Manitoba: A First Look* (Manitoba Centre for Health Policy).

Question 14   To what extent did the risk adjustment improve the "fit" of the model used to describe the indicators? This is typically measured statistically by measures such as a t-statistic, which tells you how much better you were at predicting which patients would die when you used the risk-adjustment model compared to when you did not.

The AHRQ and 3M risk-adjustment processes are employed to control at least partially for variances in patient health status. The methodology employs three types of adjustments involving age, gender, and co-morbidities. They are not used to predict which patients would die. The risk-adjustment model has not been validated by us. It has been thoroughly validated in the course of developing the AHRQ program over the past decade. It also has additional value because the methodology is transparent, is in use in many other jurisdictions, and is done in an identical and therefore comparable way. The software required to run these programs is in the public domain, in contrast to similar reports, which have a proprietary risk-adjustment technique.