Taking the Pulse:

Hospital performance indicators from the patient's perspective

By Julia Witt, Ph.D.

August 2007

Atlantic Institute for Market Studies

The Atlantic Institute for Market Studies (AIMS) is an independent, non-partisan, social and economic policy think tank based in Halifax. The Institute was founded by a group of Atlantic Canadians to broaden the debate about the realistic options available to build our economy.
AIMS was incorporated as a non-profit corporation under Part II of the *Canada Corporations Act* and was granted charitable registration by Revenue Canada as of October 3, 1994; it recently received US charitable recognition under 501(c)(3) effective the same date.

The Institute's chief objectives include:

a) initiating and conducting research identifying current and emerging economic and public policy issues facing Atlantic Canadians and Canadians more generally, including research into the economic and social characteristics and potentials of Atlantic Canada and its four constituent provinces;
b) investigating and analyzing the full range of options for public and private sector responses to the issues identified and acting as a catalyst for informed debate on those options, with a particular focus on strategies for overcoming Atlantic Canada's economic challenges in terms of regional disparities;
c) communicating the conclusions of its research to a regional and national audience in a clear, non-partisan way; and
d) sponsoring or organizing conferences, meetings, seminars, lectures. training programs, and publications, using all media of communication (including, without restriction, the electronic media) for the purpose of achieving these objectives.

2000 Barrington Street, Suite 1302, Halifax, Nova Scotia B3J 3K1
Telephone: (902) 429-1143; fax: (902) 425-1393
E-mail: aims@aims.ca; Web site: www.aims.ca

Taking the Pulse:

Hospital performance indicators from the patient's perspective

Julia Witt, Ph.D.

Atlantic Institute for Market Studies
Halifax, Nova Scotia

August 2007

Published by the Atlantic Institute for Market Studies
2000 Barrington Street, Suite 1302
Halifax, Nova Scotia B3J 3K1

Telephone: (902) 429-1143
Fax: (902) 425-1393
E-mail: aims@aims.ca
Web site: www.aims.ca

Design by Gwen North.

AIMS would like to thank The Max Bell Foundation, the Marigold Foundation and the Lotte & John Hecht Memorial Foundation for their financial support for this project.

CONTENTS

About the Author

Julia Witt is a Research Fellow at the University of Melbourne's Institute of Applied Economic and Social Research in the area of Health Economics. Her current research includes the recruitment and retention of doctors and nurses, measuring the quality of care, occupational choices of nursing graduates, discrete choice experiments, and the effects of domestic violence on households.

EXECUTIVE SUMMARY

This paper, *Taking the Pulse*, examines indicators used as tools to measure, compare and improve the quality of care in hospitals. It is a growing practice in the health care field, not just examining accreditation of a facility, but also collecting data to determine the success of a hospital in providing care to its patients. For example, knowing which hospitals have high rates of surgical site infections is information a patient going into surgery is likely to want to know to make a more informed choice about the hospital in which he or she will be treated. At the same time, such information provides added incentive to a facility to determine the cause of the problem and correct it, thereby improving the quality of health care.

Indicators are collected and published in different countries for a number of reasons. In the United Kingdom, they often are used to rank hospitals in relation to each other, so patients can make informed choices and decide where they will likely get the "best quality" care. In the United States, there are a number of agencies at both the state and national level that publish data on indicators to evaluate and compare the performance of hospitals. In Canada, report cards published by, for example, the Ontario Hospital Association, the Fraser Institute and the Canadian Institute for Health Information, are aimed at helping hospitals improve their performance and allowing patients to make more informed decisions.

This paper is the second in a two-part background series in preparation for a hospital report card to be issued by the Atlantic Institute for Market Studies (AIMS). The first, *Finger on the Pulse*, focused on the accreditation mechanisms and quality assessment systems currently in use in the United States, Europe, and Canada. The purpose of this paper is to familiarize the reader with some of the indicators that are often used to compare providers and to draw attention to areas where improvements are necessary in the way that the indicators are being applied and used. Additionally, some indicators that *should be* collected are suggested, which are ones that would

be beneficial to patients to properly assess which hospital provides the best quality of care for particular procedures.

INTRODUCTION

Amidst mounting efforts to measure, compare and improve the quality of care in hospitals, the use of indicators as a tool to meet these goals is a rising trend, and there are many reasons for this. From a practical point of view, indicators can often be calculated from data that hospitals routinely collect. For example, surgical complications are recorded in patient charts, and the number of hours that a nurse works in a specific ward is standard administrative data. Indicators can also affect the relative attractiveness of hospitals to those who intend to use them by quantifying outcomes that are desirable to know. Knowing which hospital has a high rate of surgical site infections, for example, is beneficial to patients who need to undergo surgery, because they probably want to avoid hospitals that have a high rate.

Indicators are published in different countries for a variety of reasons. In the U.K., indicators are used in league tables to rank hospitals in relation to each other so that patients can make informed choices about where to get the "best quality" care, and so that the National Health Service (NHS) can use these as incentives for hospitals to improve their care.[1] In the U.S., a number of agencies, some at state level, others at the national level, publish data on indicators to evaluate and compare the performance of hospitals. There are a large number of such reports available, probably making the U.S. the most sophisticated at reporting these statistics. The most prolific of the reporting agencies is probably the Agency for Healthcare Research and Quality (AHRQ). They have defined and validated a number of indicators that have been adapted by other agencies and in other countries, including Canada. In Canada, the Ontario Hospital Association was the first to widely publish comparative hospital data in the late 1990s. These reports publish information on whether hospitals perform above or below average, along with actual rates for each indicator. Other reports are publicly available that compare hospitals based

[1] For discussions about NHS performance measures and league tables, see, for example, Witt (2005); McKee and Sheldon (1998); Adab et al. (2002); Lakhani et al. (2005).

on a set of indicators, such as the Fraser Institute's *Hospital Report Card* for Ontario hospitals, and Ontario's Institute for Clinical Evaluative Sciences' *Quality of Cardiac Care in Ontario*. At the national level, the Canadian Institute for Health Information and the Hay Group have annually been releasing their *Benchmarking Comparison of Canadian Hospitals* since mid 1990. This report is released to participating hospitals so that they can compare themselves to their peers and identify areas where their performance could be improved relative to their peers. Performance benchmarks are set by the best performing hospital in each category (surgical/medical intervention and case mix group), and hospitals can view where they stand in relation to the "best" hospital.

This paper is the second in a two-piece background series, and focuses on indicators. One purpose of this paper is to familiarize the reader with indicators, specifically the ones that are used in the *Benchmarking Comparison of Canadian Hospitals*. Some of these indicators are discussed in more detail, mostly those that are most directly related to the quality of care. Another purpose of this paper is to draw attention to areas where improvements are necessary in the way that the indicators are being used and reported. The paper also includes a section on what indicators *should be* collected. These are ones that would be beneficial to patients in being able to properly assess which hospital provides the greatest quality of care for particular procedures.

THE LITERATURE ON INDICATORS

The evidence regarding usefulness of indicators is mixed. Indicators enable, or at least facilitate, comparisons of outcomes that are to some degree a function of the quality of care that was received to attain that outcome. This should allow patients to make more informed choices about where to go for certain treatments. A hospital that routinely has fewer surgical complications per 1,000 surgeries, for example, would be preferred by patients who need to undergo surgery.

However, whether existing indicators really quantify what they are supposed to is debatable. Many studies have shown otherwise for a variety of reasons, ranging from inconsistencies in reporting standards among hospitals (and hence lack of comparability) to the question of whether differences in outcomes really embody quality, or just a whole slew of other factors that affect the outcome.

Health outcome indicators are supposed to reflect the quality of care that a patient received during a specific treatment. For example, one can compare mortality within 28 days following admission to the hospital for acute myocardial infarction (AMI, meaning a heart attack). Some of these deaths occur because patient characteristics make them less likely to survive an AMI episode: older patients who are quite sick already are less likely to survive than young, otherwise healthy patients. However, part of the probability of survival depends on how quickly patients are treated in the hospital. Delays in administering thrombolysis and primary angioplasty (surgical repair or recanalization of a blood vessel) have been shown to be closely correlated with mortality among patients with AMI (see, for example, De Luca et al, 2004). Delays in getting life-saving treatments are clear indications of poor quality of care, since the benefits of timely treatment are well-documented and widely known. The hospital that administers life-saving therapy more quickly would be deemed better quality than the one that has longer delays because patients can expect to have better outcomes when treated in the "faster" hospital. There

are a variety of problems with attributing a good outcome to good quality of care, chief among them being that patients are not homogenous, and that survival is highly correlated with patient characteristics. When indicators are calculated, hospital-level data are typically used. This means that adjustments are made for the overall case mix and patient severity that the hospital treats, but individual cases are not evaluated one-by-one. There are, of course, practical reasons for doing so: a hospital treats many patients, and it would not be feasible to calculate indicators taking every individual patient characteristic into consideration. However, the cost is that the resulting indicator is unlikely to accurately reflect a hospital's quality of care. Statistically speaking, hospitals differ in their patient mix, and many outcomes occur simply as a result of this. Depending on the degree of heterogeneity and the adjustments that are made for it, indicators can be quite misleading in their assessment of the quality of care.

Likely the biggest hurdle that indicators are faced with is separating the extent to which health outcomes are the result of factors not related to the quality of care that the hospital provided from true quality differences. This is not a small task, since health outcomes are the intricate result of numerous factors. In fact, Arah and Westert (2005) studied the correlations between health and health care performance of Canadian provinces and territories. They found that health is relatively better associated with community and health system characteristics than health care performance is, in other words, their results suggest that health is better related to things like health behaviours, living and working conditions, personal resources and environmental factors than to health care performance (such as safety, effectiveness and appropriateness). With respect to indicators, this implies major difficulties in measuring quality especially if quality of care is not even the "deciding factor" for health outcomes.

There are numerous papers that deal more specifically with this key concern that indicators may not be able to separate quality from all other factors, usually with reference to a particular indicator or specific group of patients. Following is a small sample of the findings from the literature on this subject.

The worst outcome that a patient can have is death, and so mortality rates are popular indicators because patients would want to avoid hospitals that have high mortality rates, and, from a technical perspective, death rates are easy to measure. However, it is difficult to attribute a patient's death to poor quality care. Hofer and Hayward (1996) address this problem of attributing excess mortality to the quality of care. They used a Monte Carlo simulation model to examine whether mortality rates could be used to distinguish average quality from poor quality hospitals. Their results showed that, for individual diagnosis-related groups (DRG), mortality rates were a poor measure of quality, even with the optimistic assumption that case mix was perfectly adjusted. In their simulations of 191 hospitals across Michigan, they test whether a 20% difference in preventable death rates is detectable using mortality rates. The "average-quality" hospitals have a death rate of 5%, while "poor-quality" hospitals have a death rate of 25% in their analysis. The results showed that DRG-specific mortality rates are not good indicators of poor-quality hospitals. They conclude that "[a]n indicator based on mortality rates would generate more false-positive findings than true positives and would identify very few of the poor-quality hospitals." (pg. 747). In other words, what their study showed was that, in order to distinguish good quality from poor quality using mortality indicators, the difference in death rates between hospitals would have to be very large. In fact, the difference they use to demonstrate their point is 20%, which is so high that it is very unlikely to exist among any type of comparable provider.

Thomas and Hofer (1999) studied whether risk-adjusted mortality rates are valid indicators of the quality of hospital performance using random measurement error, sensitivity and predictive error of mortality rate in a hypothetical group of hospitals with perfect risk adjustment and equal patient volume. Poor quality of care hospitals were those with an observed mortality rate much greater than their expected mortality rate, i.e., those in which a high mortality rate could not be attributed to random variation alone. They found that less than 12 per cent of poor-quality hospitals were identified as high mortality rate outliers, and that more than 60 per cent of outliers were associated with good-quality hospitals. They calculated sensitivity to be less than 20 per cent, and predictive error greater than 50 per cent, thus concluding that measuring quality with

risk-adjusted mortality rates is highly inaccurate. The false positive/false negative rates improved the larger the number of patients treated in each hospital, but even for a hospital with 900 patients (which would be relatively large), the problem remained significant.

Similarly, Mant and Hicks (1995) found that disease-specific mortality is not sensitive enough to compare quality differences among hospitals, even with the assumption of perfect severity adjustment and identical case ascertainment and definition. In their analysis of two hypothetical hospitals, treating the same number of identical patients each year, a mortality difference of 9 per cent (mortality rate in poor quality hospital 30 per cent; mortality rate in good quality hospital 21 per cent) would require a sample size of 389, assuming that the poor quality hospital uses no effective interventions (such as aspirin, beta blockers, etc.), and that the good quality hospital uses these 100 per cent of the time. Detecting a quality difference of 3 per cent (good quality hospital mortality rate 27 per cent) requires a sample size of 3,619. The assumption, however, of perfect and zero intervention use is, of course, unrealistic, as are the very large differences in mortality rates and identical patients. Hence, sample sizes are underestimates of what would actually be needed, taking these heterogeneities into account. They conclude that "…disease specific mortality is an insensitive tool with which to compare the quality of care among hospitals" (pg. 795).

There are a large variety of indicators that are related to morbidity (the rate of sickness), rather than mortality, though their interpretation faces similar challenges. Asthon, et al. (1995) examined data of male veterans using Veteran Affairs hospitals, and found that only under certain circumstances were early (within 14 days), unplanned readmissions associated with the quality of care during the previous hospital stay. Their study population included men that were recently discharged for diabetes, chronic obstructive lung disease or heart failure. Of unplanned early admissions for each of these diseases, one in seven, one in five, and one in twelve for diabetes, heart failure and chronic obstructive lung disease, respectively, was attributable to substandard quality of care. Although there are some caveats to their findings, an indicator, based on their data, for the quality of care given to diabetes patients, for example, would have to

be able to distinguish one unplanned early readmission that was the result of poor quality from six unplanned early readmissions that were the result of other contributing factors.

Scott et al. (2004) tried to find out if actual differences in quality of care exist between hospitals, with an analysis based on twelve diagnosis-outcome indicators (for example, readmissions for acute myocardial infarction), using different hospital groupings [(comparing only peer groups (i.e., only tertiary, community or district hospitals) and comparing all hospitals]. They mention some problems with the use of indicators that have been identified in the previous literature. Among these was inadequate adjustment for case mix and illness severity between hospitals, which makes comparisons difficult. As well, generalising hospital performance based on just a few indicators was found to be problematic at best. While additionally, smaller hospitals with smaller caseloads would probably fare worse because there are too few cases to statistically assess them properly. Using data relating to all acute care episodes for the financial year 1999/2000, and adjusting for identified risk factors (co-morbidities), they found significant systematic variation for the indicator "readmission for AMI" when all hospitals were compared, but not when hospital peer groups were compared. This suggests limited scope for using indicators: only in certain settings are comparisons actually meaningful. The authors state that indicators should be chosen on the basis that they *do* identify poor performers, not on the premise that they are *likely* to reflect quality differences. They conclude that indicators should be chosen based on the following criteria: that they use only discrete, high volume diagnoses (to avoid detecting random errors rather than systematic variation); that data collection is standardised; and that data are validated as markers of systematic variation.

Weissman et al (1999) used a hospital-level simulation to find that "Related Adverse Readmissions" (RAR) did not discriminate well between average and low quality hospitals. Based on a statistical analysis of Medicare beneficiaries in four U.S. states between 1991 and 1992, they found only small differences in initial quality of care between patients readmitted for RAR and patients who were not readmitted. Patients readmitted for reasons other than RAR also tended to have lower quality of care, but these differences were also very small. They conclude that RAR are not likely to be useful in identifying hospitals that provide low quality of care.

Jha et al. (2005) found that the quality of care in American hospitals varied widely depending on the type of indicator and on the diagnosis. Hospitals that scored very highly for indicators relating to AMI often also provided high quality of care for congestive heart failure (CHF), but not necessarily for pneumonia. Thus, high quality care for AMI was a good predictor of high quality care for CHF, but not for pneumonia. Their results suggest that "high quality" hospitals are not consistent in their performance across conditions, and thus not easily identified, and that evaluating hospital performance would probably need to be based on a large number of different conditions.

Statistically valid indicators are clearly important, but there are other aspects that need to be considered when deciding what indicators to use, and for what and for whom their use is intended. Hibbard and Jewett (1997) show that, if indicators are not well understood by the consumers, then regardless of how salient they are, consumers dismiss them as unimportant. Schneider and Epstein (1998) surveyed 474 patients who had undergone Coronary Artery Bypass Graft (CABG) surgery on their knowledge of an existing consumer guide to CABG. Only 12 per cent of patients knew of the report before their surgery, and less than 1 per cent of patients knew the correct rating of their surgeon or hospital, and had, at least moderately, used this information in their provider selection. Knutson et al. (1998) compared the selection of health plans of two sets of employees, one with access to a health plan report card, the other without, and found that there was very little difference between the choices these two groups made about their health plans. These findings suggest that the use of report cards, at least for those making the decision, is very limited. Possible explanations are that the report cards are too difficult to interpret, or that they are too long and time-consuming to read. In fact, it has been shown that how quality data are presented has a major impact on whether or not it gets used by consumers and so those who disseminate the information have a responsibility for doing this properly since information that is not correctly presented could even be used to manipulate consumers (Hibbard et al., 2002). For hospital report cards, this is an important consideration, especially since the presentation of medical data to non-medical consumers is not straightforward.

Notwithstanding these criticisms, indicators constitute an important part of accountability and transparency in health care. They are indispensable because they allow patients to make more informed choices, while alerting hospitals to areas where they can improve their services. However, relying heavily on routine calculations of indicators that may be misleading can be more detrimental than helpful. Moreover, there are indicators, other than the ones that are currently being used, that more appropriately reflect quality related to what is important for patients.

There has been some literature on what constitutes a good indicator. The Agency for Healthcare Research and Quality (AHRQ) lists five key questions that help identify an indicator's validity.

(1) "How strong is the scientific evidence supporting the validity of this measure as a quality measure?" This includes, for example, whether or not studies in peer-reviewed journals have been published regarding use of the indicator.

(2) "Are all individuals in the denominator equally eligible for inclusion in the numerator?" In the calculation of the indicator, only relevant cases, including those at risk, should be included. For example, the indicator "In-hospital hip fractures of the elderly" measures the proportion of elderly patients who fractured their hip in the hospital, and so in the calculation should only include elderly patients at risk for hip fractures, and not, for example, non-elderly patients at risk for hip fracture.

(3) "Is the measure result under control of those whom the measure evaluates?" For example, when assessing hospital quality, it is important that the "quality" indicators can really be influenced by the hospital. Otherwise, there is little effect in reporting them, and maybe even detrimental in attributing a poor outcome to a hospital that has no influence over the outcome.

(4) "How well do the measure specifications capture the event that is the subject of the measure?" In some cases, outcomes may not always be clearly attributable to the measure in question.

(5) "Does the measure provide for fair comparisons of the performance of providers, facilities, health plans, or geographic areas?" In some cases, hospitals may not be comparable because they are different in nature, or because they serve different types of patients.

Hofer et al (1997) propose asking four questions before using an indicator.

(1) When cases relating to a specific indicator are examined, can a set of definable and preventable processes of care that are known to lead to a bad outcome be found?

(2) Can a review instrument be created by which problems within the process can be identified by the providers?

(3) Does the indicator identify substantially more process problems, and can the sensitivity and specificity of the indicator be defined?

(4) Who is the indicator useful for – the provider for quality improvements or an external measure to compare quality across providers?

This literature suggests that indicators need to be selected on the basis of very stringent criteria; moreover, it is probably useful to individually assess the validity of each indicator for the specific setting that it will be used for. For example, an indicator may be valid for a teaching hospital in an urban centre but may not prove to be valid for a non-urban hospital. Then, not only will such careful consideration validate an indicator for its specific use, having to answer these types of questions will also force agencies that publish indicators to provide and think about very specific issues regarding the intent of information that gets conveyed when each indicator is published. The AHRQ has done substantial work in validating and improving indicators, and addressing many of these issues. However, gaps remain to be filled, and the universal validity and adaptability of indicators is not obvious. And while it is difficult to imagine that a "perfect" indicator will ever be feasible, the widespread use of, and growing reliance on indicators without clear warnings is worrying. In addition, the use of indicators is going beyond simple comparisons. CIHI uses indicators to set performance benchmarks, and while this is not much different from simple comparisons, the idea of formalizing their comparative function reinforces concerns that they will be misinterpreted or used to set benchmarks that are not very meaningful.

BENCHMARKING

In the 1980s, Xerox made 'benchmarking' popular, and since the process has become an important aspect in the success of organizations (Heylar et al., 1998). Benchmarking promotes performance improvement by using information that is obtained through ongoing measurement and comparison of work processes across organizations.

The unique feature of benchmarking is that it sets targets based on leaders in the field. That is, the benchmark-setting hospital is the one that has the best outcomes. An advantage of benchmarking is that it compares actual processes that are already in place. As such, it avoids setting unrealistic targets for processes that may only be possible under "ideal" conditions. However, this may also be a disadvantage of benchmarking: if the leader in the field has considerably more resources available to outperform others in a particular process, then the target set by benchmarking may not be achievable for other providers, making them look relatively worse than they actually are. The leader may also have chosen to allocate a relatively large amount of resources into one specific process, trading off quality elsewhere. Thus, being a leader in one particular process may be at the expense of other processes. By the same token, if the leader of a specific group of providers is just average compared with providers outside the group, then the benchmark for that group will be mediocrity.

The Canadian Institute for Health Information (CIHI) annually publishes *Benchmarking Comparisons of Canadian Hospitals*, which allows Canadian hospitals to compare their clinical efficiency, operational efficiency and quality of care to peer hospitals (for example, to all other teaching hospitals, which are based on the hospitals' descriptive indicators and characteristics) across Canada (CIHI, 2003). The benchmarks are set by the top performers for each specific category, making the performance levels a comparative measure that can only identify hospitals that are significantly different from the average, though this may not necessarily mean that they are poor performers. The categories by which benchmarks are set are finely stratified according

to case mix, age group and complexity of the cases, by the intervention type, and by the diagnosis. There are potentially thousands of categories in which benchmarks can be set, though not all of these actually exist. In order to set a benchmark, there needs to be a sufficient number of cases that have been treated, otherwise the benchmark would not be meaningful. It is important to note that a benchmark is not criterion-based; it is simply a comparison of peer hospitals treating roughly the same type of patient.

In CIHI's comparison, benchmarks are set for certain indicators, which allow hospitals to see how they rank relative to their peers. In addition to the benchmark comparisons, there are a number of indicators that are published for each hospital, which allows hospitals to identify processes that could be improved.

INDICATORS

The focus of this paper is to discuss the indicators that are used in the *Benchmarking Comparisons of Canadian Hospitals* report. These indicators are not exclusive to CIHI's report, and most of them are widely used by other agencies. The reason for focusing on these indicators, and not others, is that our hospital report card will use the same data that was used to construct the indicators in the benchmarking report. Hence our report card will use the same indicators, and this paper will serve as a background paper and a cautionary note for these. Many of the issues and possible difficulties associated with using these indicators are "transferable" to other indicators.

There are two broad areas that will be discussed in this paper. The first are the "clinical efficiency" indicators, as they are used by CIHI in their benchmarking report. Discussion of these will explain how performance levels are set and used, and what some potential problems might be in using them. The other group of indicators are the "standard" type, which includes mortality-based indicators and readmissions following certain treatments. These are probably the ones that people first think of when they hear "indicator". They identify areas where hospitals can improve themselves, and they are also meant to help patients identify hospitals with "high-quality" care.

CLINICAL EFFICIENCY

Within clinical efficiency, there are two types of measures, one to provide an overview of the type of hospital, which includes different measures of hospital characteristics; the other is a comparison of performance levels of peer hospitals.

There are eleven hospital descriptive indicators, providing an overview of each hospital, and used in the benchmarking comparisons. The first hospital descriptive indicator are the hospital patient volumes, stratified into total number of inpatient cases, and total number of qualifying day procedures. Qualifying day procedures are those that do not require patients to stay in the hospital overnight, and the percentage of day procedures that a hospital performs is considered an important indicator because acquisition of new technologies allows procedures that were traditionally performed on an inpatient basis to be done on an ambulatory basis. Thus, the more "qualifying" day procedures the hospital performs, the "better" it is deemed to be at the acquisition of new technologies, which are likely to improve patient satisfaction and outcomes since recovery is usually easier for patients when they are allowed to go home. Moreover, it is cost-saving for the hospital, allowing resources to be allocated elsewhere.

There are two descriptive indicators that assign resource intensity weights (RIW). Since cases differ by the amount of resources that are needed for treatment, this indicator estimates the resource utilization of each specific case. One of these indicators is the hospital average inpatient weight per case, which assigns RIW to each case. The other is hospital average inpatient weight per day, which estimates the average daily resource intensity. Another set of indicators stratifies inpatient cases by primary/secondary and tertiary/quaternary. This measures resource intensity to some degree (quaternary services require much more specialized resources than primary care services), and thus also indicates the type of hospital. Primary care refers to basic hospital care; secondary care to that provided by a specialist. Tertiary care requires specialized skills, technology and support services, such as heart surgery. Quaternary care requires highly

specialized skills, technology and support services, such as multiple organ transplants. Not all hospitals provide tertiary care, and quaternary care is usually concentrated in hospitals that specialize in that type of care. Hospitals with many quaternary inpatient cases are not very comparable to mostly primary care hospitals, because they would provide very different services to their patients, who would have very different needs.

Another descriptive indicator splits the hospital's patients into age groups (0 – 17, 18 – 69 and 70 and up). Since younger patients are more likely to have a speedy and more complete recovery, and also to suffer from fewer co-morbid conditions, this is an important indicator that is used to adjust some of the health outcome related indicators that are also used.

The remaining indicators are ones that specify "exclusions", or are issues of caution. Typical and outlier cases and inpatient days stratify patients into those who have received a standard program of care in a single institution ("typical"), and those who were discharged after a longer length of stay than would usually be necessary for that type of care ("outliers"). Outliers are excluded from the clinical efficiency analyses, because it is possible that some "atypical" characteristic of the patient has led to the longer than usual length of stay. "Transfer from" cases are also excluded from clinical efficiency comparisons; these are patients that were transferred from another acute care institution. It is not clear which of the multiple acute care institutions the outcome of the treatment should then be attributed to, which is why this category is for exclusion. Finally, alternate level of care (ALC) days, although included in the clinical efficiency comparisons, are recorded separately because they are not directly part of the episode of care. These include, for example, days in the acute care facility following the completion of the acute care treatment that are spent waiting for placement to another facility. They are included (rather than excluded) in the calculation of total patient days because there are variations between hospitals and provinces/territories in reporting ALC days. In other words, since there are differences in when hospitals consider "ALC" days to start, they cannot be excluded from calculations.

Clinical efficiency comparisons use two types of indicators: day procedure (DP) performance levels, and average length of stay (ALOS) performance levels. The former measures the number

of eligible procedures that are done on an ambulatory basis. The latter assesses how long patients stay in the hospital.

The day procedure performance level (DPPL) is used to "…identify potential opportunities to increase the proportion of care provided on an ambulatory basis" (CIHI, 2003, pg. 17). For each hospital and CPA (Case Mix/Complexity/Age Group) combination, the percentage of day procedures is calculated (= qualifying day procedure cases / inpatient and qualifying day procedure cases). To be able to set the benchmark for the DPPL for a given CPA combination, hospitals must have at least 30 eligible cases and an ALOS ≤ 5 days. Performance levels for each CPA combination are calculated by identifying the hospital with the highest percentage of day procedures for the given combination. For each hospital, the performance levels are applied to cases within each CPA combination, resulting in the number of additional day procedure cases that would have been required for the hospital to achieve the performance level. The difference between the performance level and the actual number of day procedure cases for a given CPA combination represents the cases to be shifted in order to achieve the CPA performance level. For example, a hospital has 85 qualifying day procedure cases, and 105 inpatient cases. Therefore, qualifying day procedures = 85 / (85 + 105) = 44.7 per cent. If the DPPL is 50 per cent, then the hospital is "short" of the target, because only 44.7 per cent of eligible cases are performed on an ambulatory basis. There are: 85 + 105 = 190 total cases in this CPA group eligible for a day procedure. With a 50 per cent PL, the hospital should be treating 95 patients on an ambulatory basis. If all 105 inpatient cases are one-night stays, then applying the performance levels should save 10 inpatient cases, or 9.52 per cent of inpatient cases. If only 6 are one-night stays, and the rest two-night stays, then 8 inpatient cases should be shifted to ambulatory care (i.e., 6 x one-night + 2 x two-nights = 10 one-night stays, but only 8 different patients).

There are some possible sources of problems with setting such performance benchmarks. The most obvious is whether treating on an ambulatory basis really is better or just cheaper? CIHI uses a screen, similar to that of the 1996-1997 Day Procedure Exclusion List, to identify "qualifying day procedures". The cases are then stratified by CPA (of which there could be more

than 5,000 possible combinations: there are 477 case mix groups (CMG), that could be stratified into four complexity (Plx) levels, and then further stratified into three age groups, 0-17, 18-69, 70 and up). Benchmarks are set for each of these, and there can be many categories, depending on how many CPAs there are. The performance level is set by the hospital that has the highest percent of day procedures for the given combination (given that the day procedure "qualifies"). However, "qualifying" a day procedure does not take CPA into consideration, while performance levels do. The problem might be that, even though a procedure qualifies as a day procedure, it may not be the best type of care for that particular CPA group. For example, a simple procedure might not be so simple for patients with several chronic diseases, such as for patients with diabetes and asthma. If the procedure is considered an ambulatory one, a hospital simply treating patients based on this may be ignoring the fact that "high level" CPA patients might have additional requirements.

In other words, for a certain CPA combination, ambulatory care may not be optimal at all. However, the hospital that has the "best" performance level (i.e., the highest number of ambulatory cases per total cases) sets the performance level for all other hospitals, even though within that CPA a "lower" performance level might be better. Should it be considered "better quality" when a hospital routinely treats all patients (regardless of level of complexity) on an ambulatory basis for qualifying day procedures, as compared to a hospital that keeps patients who have more complex cases (for example) overnight? Rather than setting good performance standards, this may actually identify poorer quality, because those hospitals with the highest percent of qualifying day procedures may not be applying them correctly (to groups of patients who may benefit from an overnight stay).

Setting targets by the shortest Average Length of Stay (ALOS) can be as problematic as choosing performance levels based on the percentage of qualifying day procedures. The ALOS is calculated as (inpatient days – days shifted to ambulatory care) / (inpatient cases – cases shifted to ambulatory care), where cases were shifted from inpatient to ambulatory care to achieve the DPPL. The remaining cases are the ones with longer lengths of stay. Performance levels are calculated by identifying the hospital with the shortest ALOS for a CPA combination. Do longer

lengths of stay really reflect lower efficiency or is it something idiosyncratic, related to the patients' needs? Shorter lengths of stay could be the result of insufficient funds or space to keep patients for a more optimal length of time in the hospital. If there is pressure on freeing up beds, patients may be released earlier in order to make beds available. Additionally, there is the proximity of the patient: when patients live near to the hospital so that they are able to return quickly if side effects were to occur, is that a reason for releasing them earlier? When patients live far away, are they kept longer because if they left it would be too far for them to return in time to treat side effects? Some of those additional days may also be ALC days: patients who need to be transferred to another facility because they do not have support at home may end up staying longer simply because they have no where else to go.

It can be seen then that while clinical efficiency measures are an effective means of assessing hospital resource allocation they are not as effective a measure of quality of care. They do not so much measure desirable health outcomes as they measure desirable resource outcomes. From a patient's perspective, while resource efficiencies will likely result in a broader suite of services, of more immediate interest is an appropriate resource allocation that achieves a positive health outcome in the current instance.

HEALTH OUTCOME INDICATORS

This section defines and discusses the indicators that are related to health outcomes that matter to patients. Some are discussed in more detail than others, which partly reflects the degree to which the indicators have been studied in the literature. Many of the concerns that are specifically addressed for only some of the indicators are applicable to all of them.

Mortality and Complications after Surgery

Post-operative mortality is a risk that all patients undergoing surgery face. Some of these deaths are preventable, and it is these that reveal the quality of care received. Using mortality after surgery as an indicator assumes that some of the excess mortality is attributable to the quality of care received. Hospitals with higher death rates following surgery are deemed lower quality providers because some of this excess mortality must have been avoidable. The key to being able to use this indicator is that there must be a sufficient number of cases to be able to differentiate deaths attributable to quality from everything else. Dimick et al. (2004) found that only mortality following coronary artery bypass graft (CABG) surgery was a useful measure of hospital quality when the definition of poor performance was twice the national average death rate for each procedure. All other surgeries that they examined were not performed often enough to warrant their use as an indicator of quality, since only a small proportion of hospitals exceeded the minimum number of cases needed to reliably detect poor performance. They identify the problem, "…small samples and low event rates combine to limit the statistical power of a comparison between an individual hospital and a population-based benchmark" (pg. 847 – 848), as a type II error. Michaels (2003) found that using mortality data in performance comparisons is only meaningful when high-quality data with clinical information are available, so that differences in case mix and patience selection can be incorporated. The two main concerns then with using mortality as an indicator are that there are many factors other than quality of care that affect mortality after surgery, and that there may not be enough cases, particularly in smaller

hospitals, to be able to make meaningful comparisons (due to limited statistical power based on small samples).

Complications from surgery include cardiac arrest after major surgery, acute myocardial infarction after major surgery, surgical site infections and readmission via ER of surgical cases with wound infection. Issues with using these are similar to the ones associated with using mortality after major surgery: quality cannot always be disentangled from all the other factors that contribute to these outcomes. Additionally, there may not be enough observations to be able to make meaningful comparisons.

The Well-Being and Safety of Elderly Patients

This includes falls of elderly inpatients, and the development of pressure ulcers. Rubenstein et al (2001) found that impaired gait and balance are the most significant underlying causes of falls, and these are also common consequences of falls. A study by Frels et al (2002) found that a previous fall, benzodiazepine (used for night sedation) intake and the need for maximum assistance were significant predictors of falling in the hospital among elderly patients. Vassallo et al (2003) found that characteristics, such as confusion, history of previous falls and unsafe gait are the main determinants of falls among elderly patients. In light of these studies, it may be difficult to attribute falls to the quality of care.

Rate of Reported Misadventures for Surgical Patients

According to Pennsylvania Health Care Cost Containment Council (PHC4) data, 3.4 out of every 1,000 admissions to Pennsylvania hospitals had a "misadventure" coded in 2003. This includes things like accidental cuts and punctures, foreign objects left in the body, and contaminated or infected blood. These cost nearly US$64 million, and accounted for roughly 8,000 hospital days (PHC4, 2004).

Caesarean Section Deliveries

There are two indicators that measure the rate of caesarean deliveries: one as the rate of all deliveries, the other as the rate of first-time mothers only. Caesarean deliveries can either be

elective, or emergency. The indicator does not distinguish between these. In addition, outcomes are more important: when a vaginal delivery was performed, were the outcomes optimal, or could there have been less trauma had a caesarean been done?

Use of Laparoscopic Cholecystectomy for Uncomplicated Cases

Cholecystectomy (surgical removal of the gallbladder) can be done with a laparoscope (a small video camera) or via open surgery (major abdominal surgery). Laparoscopic cholecystectomy has been identified as an underused procedure in the U.S., and it is associated with less morbidity in less severe cases (AHRQ). This indicator identifies the proportion of cholecystectomy cases that were done laparoscopically out of all cholecystectomies, a higher number implying better performance since that means more cholecystectomies were performed using a laparoscope. However, there are cases where laparoscopic surgery is not recommended. Additionally, the success of laparoscopic surgery depends highly on the skill of the surgeon performing it. Using this as an indicator raises questions of validity and concerns about setting targets for the number of laparoscopic cholecystectomies that should be performed. If a hospital has "sicker" patients and a less skilled surgeon in this area, open cholecystectomies may actually lead to better outcomes in the patient.

Paediatric Admissions Treated for Asthma

Asthma that is well-controlled rarely requires hospitalization. The purpose of this indicator, then, is to identify problems within primary care in the community where children with asthma are not taught very well how to control it. Since asthma is linked to environmental factors, such as living in an urban centre, this indicator may not only reflect differences in primary care, but also in asthma rates that are the result of the hospital's location. This indicator is not directly related to the quality of care in hospitals.

Use of Breast-Conserving Surgery for Breast Malignancy

This is the proportion of women who received breast-conserving surgery among those with a breast malignancy who received surgery. Using less-invasive surgery (breast-conserving) leads to similar outcomes as more invasive (e.g., mastectomy) surgeries. There has been a long-

standing debate in the literature whether mastectomy is better than breast-conserving surgery. In a 2002 issue of the *New England Journal of Medicine*, two studies (Fisher et al (2002); Veronesi et al (2002)) showed that survival is the same, regardless of which of the two treatments the woman received. Better indicators would be the long-term survival of women with certain interventions, i.e., was the appropriate surgery chosen by the hospital, or the utility of the patient, i.e., is the patient happy with the procedure and the outcome?

Percentage of Inpatient Days Reported as Alternate Level of Care

The number of ALC days measures the timeliness of access to post-acute services, these being mostly attributable to elderly patients with several chronic illnesses (CIHI, 2004). There are differences in how these days are reported. This indicator measures the availability of services outside the hospital and is not a direct measure of the quality of care in hospitals.

Percentage of Admissions Classified as May Not Require Hospitalization

This indicator measures the percentage of patients hospitalized with conditions (as derived from the Case Mix Group) that can be treated on an ambulatory basis. Since some patients require hospitalization for what are usually ambulatory interventions, this indicator measures the percentage of cases that are eligible for ambulatory treatment. Use of this indicator is faced with similar challenges as those of any other measuring the proportion of "day procedures".

Short Stay Admissions

This measures the percentage of medical and mental health patients who are admitted via the ER and discharged within three days. The rationale is that high short stay admission rates might indicate possibilities for shifting some of these to ambulatory care. The arguments for and against this indicator are similar to those as for "qualifying day procedures".

Admissions via the Emergency Room

Hospitals are frequently used as the first point of care for patients who require or want medical attention right away, so this indicator may measure a variety of things. There are two possible

sides to the outcomes: a low value may reflect the availability of other resources in the community; it may also be a sign that the hospital does not admit patients who maybe should be admitted. The latter would be an indicator of quality; the former outside the scope of the hospital.

Average Acute Care and ALC Days for Discharge to Rehabilitative or Continuing Care

This indicator is meant to help recognize areas where improvement in getting patients out of the hospital and into other non-acute care facilities could be made. Again, this is often not under the control of the hospital, and having longer ALC days might be a sign of good quality, because the hospital may be willing to keep patients longer to ensure that they are released to appropriate non-acute care facilities, rather than the "first available spot".

Percentage of Mental Health Inpatients Readmitted within One Month

This indicator measures the availability of other resources in the community, as well as the appropriate use of ambulatory care. The indicator does not distinguish between planned and unplanned readmissions.

Acute Myocardial Infarction Length of Stay

This indicator measures the median of the length of stay difference, i.e., the difference between the observed length of stay and the expected length of stay, taking into account age, gender, cardiac re-vascularisation procedures and co-morbidities. A shorter LOS would reflect better technology uptake, and availability of non-acute resources in the community. However, a shorter LOS may not actually be better for the patient, or even possible. Among other things, LOS is influenced by the treatment that is administered: Berger et al (2006) found that LOS was significantly decreased when patients received primary angioplasty with stent implantation than when they received balloon angioplasty in the treatment of AMI. To routinely administer treatments because they result in shorter lengths of stay is probably not in the interest of the patient. Some patients may not be eligible for treatments with shorter LOS. Additionally, hospitals with long waiting lists for beds may be under pressure to release their patients earlier, resulting in shorter LOS. This would, clearly, not be a better quality hospital.

In addition, median calculations are not always a good indicator. The median is the midpoint, above which exactly half of all cases have a longer LOS and below which the other half of cases have a shorter LOS. If the median is zero (i.e., expected LOS = observed LOS), that means half of all cases were longer than zero, and half were shorter than zero. Medians do not reflect the range of values. That is, a hospital could look very good because exactly half of all cases are *slightly* lower than a median of zero (assuming a median of zero is desirable), and the other half much higher than the zero median. As long as at least half of all cases are below the median, it does not matter how much above the median the remaining cases are. 95 per cent confidence intervals are reported with the median. This means that the median falls 19 times out of 20 (i.e., 95 per cent of the time) within the upper and lower limits of the confidence interval. As such, the confidence interval is a range; however, it is the top 5 per cent of cases that are potentially of interest as far as quality of care goes. LOS outside the confidence interval probably occurs more often in complex cases, or cases where quality has been compromised, so it is these that are useful in assessing quality of care. Thus, it may be better to evaluate a hospital based on the more complex cases, so a better measure might be to measure the outcomes of patients with long LOS, potentially looking at those above the median separately. Looking at the distribution then, not just at confidence intervals, means and medians, could potentially shed a different light on the interpretation of the outcomes.

Acute Myocardial Infarction Readmission Rate

Readmission for AMI is supposed to reflect the quality of care that was received during the initial AMI episode. The drugs that are prescribed to the patient, the availability of resources and help outside the hospital and the care that the patient received during the initial episode are all related to readmission. However, there are other important factors that influence the rate of readmission, which are not captured by this indicator, such as lifestyles (smoking and sedative lifestyle). The usefulness of this indicator is muddled by these additional important contributing factors, as well as by the support that the person receives outside the hospital (which are community related). It may not be very clear what this indicator measures.

Hysterectomy Length of Stay

This is similar to AMI LOS, the main difference (as far as indicator interpretation is concerned) being that hysterectomy is a planned admission, while AMI is an unplanned admission. There may be selection bias: patients may choose hospitals that are (a) better at performing hysterectomies, and (b) have a shorter waiting list for hysterectomies.

Hysterectomy Readmission Rate

Readmission rates are calculated as a relative risk compared to the rate at all other hospitals. In Canada, there were 441 hysterectomies performed per 100,000 women in 2001 (CIHI). The concerns are similar to those for readmissions for AMI.

Percent of Day Surgery Cases for Sentinel Procedure Groups

This measures the percentage of surgery cases that are day surgeries (discharged within 24 hours following surgery) for certain Day Procedure Groups (DPG). Sentinel procedure day surgeries are appropriate to be performed in a day, for example cataract extraction. Day surgeries are inherent to the same challenges as day procedures: releasing patients quickly may not always be in the best interest for the patient, even if the surgery is normally a day one.

Wait Time in the ER before Admission to the Inpatient Unit

This is calculated as the difference between the time a patient was assigned a bed in the inpatient unit and the time the patient was designated as requiring admission as an inpatient. It would probably be useful to stratify this indicator according to condition: some cases require faster admission than others, and so a longer wait time for a bed does not affect all patients the same.

Utilization of Medical Beds for Pneumonia and Influenza by Seniors

This is the number of medical beds (measured by length of stay) that are used by patients 65 and over with a most responsible diagnosis of influenza or pneumonia, measured as a proportion of the length of stay of all patients 65 and over. This indicator captures the rate at which seniors are hospitalized for these conditions. High rates may point to problems in the community, such as lack of immunization. These are both communicable diseases and high rates in an area may

trigger higher rates of hospitalization. The indicator may be more useful if the incidence or prevalence of pneumonia or influenza in the community were taken into account as well.

Long Term Complications of Diabetes

This is the proportion of hospitalizations of diabetes patients with complication from their diabetes of all diabetes inpatients. This indicator measures problems outside the scope of the hospital. Complications often arise from poor disease management, and so high rates of complication-associated hospitalizations are a sign of poorly controlled diabetes.

Percent of Stroke Patients Discharged to Inpatient Rehabilitation

This is the proportion of stroke inpatients released to rehabilitation facilities. Stroke victims benefit substantially from rehabilitation, and so it is important to ensure that they are provided with the care they need. However, this indicator only measures patients transferred to an inpatient rehabilitation centre. There are other options available, such as home care, or rehabilitation in the hospital. Home care is captured in the next indicator.

Percent of Stroke Patients Discharged Home, Referred to Home Care

This is the proportion of stroke inpatients discharged home and referred to home care for rehabilitation.

Percent of Knee Replacement Patients Discharged to Inpatient Rehabilitation

Again, rehabilitation is highly beneficial for patients who have had knee replacement. This indicator measures the proportion that is discharged to inpatient rehabilitation.

Percent of Knee Replacement Patients Discharged Home, Referred to Home Care

This is the proportion of knee replacement patients that are discharged home and are referred to home care for their rehabilitation.

Post-Admission Pulmonary Embolism or Deep Vein Thrombosis[2]

Cook et al (2000) found evidence that prophylaxis for venous thromboembolism (VTE; defined as deep vein thrombosis with or without pulmonary embolism) in medical-surgical intensive-care unit patients is under-prescribed, and that VTE diagnostic tests are infrequent. Pulmonary embolism is among the most common preventable causes of inpatient deaths. This indicator measures important preventive measures that hospitals take. It is calculated as the ratio of observed to expected embolisms for every CPA.

Accidental Puncture or Laceration

This indicator captures technical difficulties that lead to complications in medical care. It is calculated as the ratio of observed to expected lacerations for each CPA. Gallagher et al found that not all accidental lacerations or punctures are recorded: for example, in their study of New York data, of inpatients undergoing hysterectomy as the principal procedure that had bladder injuries, 27.3 per cent were not recorded. Bowel injuries in cholecystectomy patients were not captured by the indicator 20.8 per cent of the time.

Death in Low-Mortality Case Mix Groups[2]

There are two indicators, one of medical case mix groups, the other of surgical case mix groups. Both measure the excess mortality rate in groups with less than 1 per cent mortality. It is the ratio of observed deaths to expected deaths. Separating deaths related to poor quality from other causes of death may be difficult with such low rates.

Birth Trauma – Injury to Neonate[2]

This is calculated as the proportion of injured live-born neonates (excluding pre-term and infants with osteogenic imperfecta) of all live births. Some issues with this indicator are that the definitions of birth trauma probably vary among physicians, the conditions that are included in the calculation of this indicator may not be entirely preventable, and there is no severity index (AHRQ).

[2] Available in 2004.

Obstetric Trauma – Vaginal Delivery with Instrument[2]

Vaginal, instrument-assisted delivery discharges with obstetric trauma as a proportion of all vaginal, instrument-assisted delivery discharges. Grobman et al (2006) found that obstetric trauma risk is significantly affected by patient and hospital characteristics: Obstetric trauma risk was higher for any vaginal or premature delivery, multiple gestation, excessive fetal growth and prolonged pregnancy. At the hospital level, obstetric trauma risk was higher in hospitals with high annual delivery volumes and high caesarean rates. Since separating hospital level factors from patient level factors is difficult, obstetric trauma is not a good quality indicator.

Obstetric Trauma – Vaginal Delivery without Instrument[2]

Vaginal delivery discharges with obstetric trauma as a proportion of all vaginal delivery discharges.

Obstetric Trauma – Caesarean Delivery

Caesarean delivery discharges with obstetric trauma as a proportion of all caesarean delivery discharges.

Proportion of Vaginal Deliveries Performed with Instrument-Assistance[2]

All vaginal deliveries that are performed with an instrument as a proportion of all vaginal delivery discharges. Studies have shown that risk of facial nerve injury and intracranial bleeding are increased with use of instruments during vaginal deliveries (see, for example, Gardella et al. 2001). However, there are situations when use of instruments is indicated (see AHRQ guidelines).

In-Hospital Hip Fracture in Elderly Patients[2]

In-hospital hip fractures in patients age 65 and up, as a proportion of all medical and surgical discharges of patients age 65 and up. Foss et al. (2005) analyzed 600 Danish patients that suffered hip fractures during their hospital stay, and found that 75 per cent of them were in acute

care wards. In their study, none of the in-hospital hip fracture patients had any preventive measures undertaken to prevent falling and fracturing. This suggests that there is room for substantial improvement in preventing in-hospital fractures.

Patients Satisfaction

A group of indicators that has not been mentioned so far, but one that is worth mentioning, is that of patient satisfaction. Arguably, "good" quality should, among other things, leave patients satisfied with their hospital stay. The Ontario Hospital Association, for example, publishes indicators related to "Patient Satisfaction" in their report card (OHA), and these include overall opinion of quality of care, satisfaction with the outcome of the hospital stay, opinions of nursing care and physician care, and so on.

There are, however, some rather major caveats associated with measuring consumer satisfaction. For one thing, outcomes are individual-specific: what "satisfies" one patient may not satisfy another. Avis et al. (1995) point out that satisfaction concepts assume that patients' expectations are compatible with each other, and this is clearly not a valid assumption. Moreover, being "satisfied" with one's hospital stay is not necessarily related to quality of care. In fact, satisfaction is more likely to be linked to other experiences, such as being looked after by a friendly nurse, than to the outcomes that are measures of quality.

Medically speaking, patients may not be able to judge how "good" their outcomes are, and may not even know that the quality of care they received would be considered "poor". Avis et al. (1995) go as far as saying that satisfaction conceptions are regressive because they promote a review of these health care processes rather than of health outcomes, which are central to quality of care. Having noted these caveats, though, does not imply that consumer satisfaction is unimportant. It is probable that satisfied patients experience less stress during their hospital stay, and thus may have better outcomes or that they may even be more likely to return to the hospital for minor complications thus improving their overall outcome. Consequently, patient satisfaction is an important dimension in the overall hospital experience, but may not necessarily be a valid indicator to compare quality of care in this context.

Quality of Care is Difficult to Separate from Other Factors

This summarizes the indicators that are used in CIHI's *Benchmarking Comparison of Canadian Hospitals*, and some of which will be used in AIMS' hospital report card. Insufficiencies and concerns have been noted along the way. The problem that is common to all the indicators is that quality cannot necessarily be separated from all other factors affecting the outcome. Compounding this problem is CIHI's approach to outliers: they are excluded from analysis in some cases, which perhaps leads to the elimination of those cases that really did receive poor quality. Since outliers contain information that is different than that of non-outlier cases, their exclusion should only be allowed for very good reasons, not simply because they are outliers.

Additionally, many of the indicators, such as the use of breast-conserving surgery for breast malignancy fail to recognize that there may be benefits to using the alternative (in this case, more invasive) intervention. Better measures of quality would be the long-term outcomes that measure whether the appropriate surgery was chosen. Publishing indicators that suggest that one type of surgery is better than another can encourage hospitals to use the "better" procedure more often, and this can be hazardous when individual cases would not warrant such action to be taken.

As with clinical efficiency indicators, for many of these outcome indicators (such as alternate level of care (ALC) days), it is not clear that they measure quality of care that matters to the patient, or whether they are simply areas where hospitals can "economize" their care. Indicators that measure short LOS, same day eligible procedures, ALC days and the like suggest that reducing these is desirable. However, reducing these would often be at the expense of the patients. A good indicator that measures quality from the perspective that patients care about might identify "good quality" as exactly the opposite. For example, staying overnight for a day procedure may be beneficial for some patients, even though they may be eligible to leave the same day, or a longer LOS may help the patient cope better in the long-term with their condition.

Patient safety indicators are important tools, both for patients and for hospitals. Miller et al. (2001), using 2.4 million discharge records from the 1997 New York State Inpatient Database, found that out of 10,000 discharges, events related to patient safety were identified in a range of 1.1 per 10,000 (foreign bodies left in patient) to 84.7 (birth traumas). Discharge records where patient safety events had been recorded had an up to threefold longer LOS, an up to 20-fold rate of in-hospital mortality, and up to eightfold higher charges (Miller et al., 2001). This suggests potential for improvement, and these are indicators that measure quality from a patient's perspective. However, some of the definitions may not be straightforward and coding practices may differ. Physicians are likely to have different thresholds to what they perceive as being an "error", and so what gets coded may differ substantially.

Finally, there are no severity adjustments for the indicators: DRG and complexity of cases do not distinguish between severity of individual cases, and it is usually severity that substantially affects the outcome. The Pennsylvania Health Care Cost Containment Council (PHC4) already insists that all organizations submit severity adjusted information about their outcomes of care, using a common methodology. This is very sensible and should be adapted for all measures that can be used in the development of indicators.

SUGGESTED NEW INDICATORS

Given the concerns highlighted in the preceding discussion, following is a list of indicators that would be more useful. Most importantly, these indicators are important to patients. They quantify desirable health *outcomes* rather than focusing on processes within the hospital.

(1) **Proportion of people who have appropriate or inappropriate wait times**. This would require capturing information about waiting times, and about what happens to the people who are waiting. Current methodologies do not adjust for illness severity. Whether wait times are long or short, it is in the public's interest to know how many people *inappropriately* wait, given their clinical condition.

(2) **Outcomes of care**. Do patients have better comfort, functionality and life expectancy following the episode of care? From the patient's perspective, undergoing surgery, for example, should be associated with better outcomes than not undergoing that surgery, so this is the most important question about the quality of care.

(3) **Information for treatments and drugs about numbers needed to treat and numbers needed to harm**. Treatments and drugs are not equally beneficial for everyone, and so the number needed to treat is the number of people, on average, who must be treated in order for one person to benefit from the treatment. For example, if four patients are treated for migraine, and migraine does not recur in one of these four patients, the number needed to treat (NNT) is 4 (McQuay and Moore, 1997). Similar logic applies to numbers needed to harm (NNH). This information is important for patients and providers since it indicates how likely the treatment or drug is to help or harm the patient. This kind of indicator is valuable for medical and surgical conditions and is commonly used to report on the effectiveness of medication.

(4) **Proportion of appropriate surgery**. Rather than publishing the proportion of less invasive surgeries (for example, laparoscopic cholecystectomy), it is more important to know whether that proportion of procedures was appropriate given the patients' clinical condition.

CONCLUSION

Indicators comprise an important part of accountability and transparency in health care. Patients have a right to know how well their hospitals are treating them, and where they can get the best quality of care. However, choosing which indicators are important and measuring these is burdened with difficulties, ranging from statistically not being able to measure what is of interest to not measuring what matters to patients at all. For example, patient outcomes, which are often used to compare quality, are very complex results of characteristics specific to the individual patient and the care that the patient received. To complicate things further, quality of care is multi-dimensional, being a product of factors directly related to the hospital itself (such as the equipment that is available and the availability of beds) and of factors that are not under direct control of the hospital (such as the individual skills of the medical staff). Indicators that do not directly matter to patients, for example those that measure operational efficiency are often also published, and the effect of publicizing these may even counteract what is in the best interest of patients. Consequently, integral criteria for choosing which indicators best represent what is being measured (i.e., quality of care) must include consideration of how well the indicator would judge quality, and whether publication of the indicator would help to improve quality of care from the patients' perspective. A careful analysis of the validity of each indicator in the specific setting in which it will be used would probably prove worthwhile.

In conclusion, the goal of indicators should be to allow meaningful comparisons to be made between hospitals, which can lead to improvements in quality of care and more informed choices for patients. Currently, these goals are only partially being addressed, and in many instances, do not appear to be part of the goal at all. If quality of care is to be improved, then providing indicators that aim at measuring true quality differences need to be developed and published widely. Otherwise, they will perpetuate more of the same.

RECOMMENDATIONS

Recommendations are:

1. Any evaluation of hospital performance should be based on indicators of the quality of care for a large number of different conditions.

2. Indicators should be chosen based on the following criteria: that they use only discrete, high volume diagnoses (to avoid detecting random errors rather than systematic variation); that data collection is standardised; and that data are validated as markers of systematic variation.

3. Assessment of hospital performance should include an adjustment to reflect differences in case mix and illness severity between hospitals.

4. All organizations should submit severity adjusted information about their outcomes of care, using a common methodology.

5. Quality of care information should be presented in such a way as to facilitate effective and regular use by health consumers in making informed choices about their care.

TABLE OF INDICATORS

Indicators
Mortality after surgery
Cardiac arrest after major surgery
AMI after major surgery
Surgical site infections
Readmission (via ER) of surgical cases with wound infection
Hospitalization of the elderly for falls
Rate of reported misadventures for surgical patients
Decubitus ulcers in elderly patients
Caesarean section rate
Primary caesarean section
Vaginal birth after caesarean section
Use of laparoscopic cholecystectomy
Pediatric admissions treated for asthma
Use of breast-conserving surgery for breast malignancy
Percentage of inpatient days reported as alternate level of care
Percentage of admissions classified as may not require hospitalization
Short stay admissions
Admission via ER
Average acute care and ALC days for discharge to rehabilitative or continuing care
Percentage of mental health inpatients readmitted within one month
Percent of day surgery cases for sentinel procedure groups
Utilization of medical beds for pneumonia and influenza by seniors
Long term complication of diabetes
Percent of stroke patients discharged to inpatients rehabilitation
Percent of stroke inpatients discharged home, referred to home care
Percent of knee replacement patients discharged to inpatient rehabilitation
Percent of knee replacement patients discharged home, referred to home care
Post-admission pulmonary embolism or deep vein thrombosis
Accidental puncture or laceration
Death in low-mortality case mix groups
Birth trauma – injury to neonate
Obstetric trauma – vaginal delivery with instrument
Obstetric trauma – vaginal delivery without instrument
Obstetric trauma – caesarean delivery
Proportion of vaginal deliveries performed with instrument assistance
In-hospital hip fracture of elderly patients

REFERENCES

Adab, P., Rouse, A.M., Mohammed, M.A. and Marshall, T. "Performance league tables: the NHS deserves better", *British Medical Journal* 324: 95 – 98 (2002).

Arah, O.A., and Westert, G.P. "Correlates of health and healthcare performance: applying the Canadian health indicators framework at the provincial-territorial level", *BMC Health Services Research* 5:76, 2005.

Asthon, C.M., Kuykendall, D.H., Johnson, M.L., Wray, N.P. and Wu, L. "The Association between Quality of Inpatient Care and Early Readmission", *Annals of Internal Medicine* 122(6): 415 – 421 (1995).

Avis, M., Bond, M. and Arthur, A. "Satisfying solutions? A review of some unresolved issues in the measurement of patient satisfaction" *Journal of Advanced Nursing* 22(2): 316 – 322 (1995).

Berger, J.S., Fridman, V. and Brown, D.L. "Comparison of Outcomes in Acute Myocardial Infarction Treated With Coronary Angioplasty Alone Versus Coronary Stent Implantation" *The American Journal of Cardiology* 97(7): 977 – 980, 2006

Cook, D., Attia, J., Weaver, B., McDonald, E., Meade, M. and Crowther, M. "Venous Thromboembolic Disease: An Observational Study in Medical-Surgical Intensive Care Unit Patients" *Journal of Critical Care* 15(4): 127 – 132 (2000).

De Luca, G., Suryapranata, H., Ottervanger, J.P. and Antman, E.M. "Time delay to treatment and mortality in primary angioplasty for acute myocardial infarction: every minute of delay counts", *Circulation* 109: 1223 – 1225, 2004.

Dimick, J., Welch, H.G., and Birkmeyer, J.D. "Surgical Mortality as an Indicator of Hospital Quality. The Problem With Small Sample Size", *Journal of the American Medical Association* 292: 847 – 851 (2004).

Frels, C., Williams, P., Narayanan, S. and Gariballa, S.E. "Iatrogenic causes of falls in hospitalized elderly patients: a case-control study", *Postgraduate Medical Journal* 78: 487 – 489 (2002).

Fisher, B., Anderson, S., Bryant, J., Margolese, R.G., Deutsch, M., Fisher, E.R., Jeong, J.H. and Wolmark, N. "Twenty-year follow-up of a randomized trial comparing total mastectomy, lumpectomy, and lumpectomy plus irradiation for the treatment of invasive breast cancer", *New England Journal of Medicine* 347(16): 1233 – 1241, 2002.

Foss, N.B., Palm H. and Kehlet, H. "In-hospital hip fractures: prevalence, risk factors and outcome" *Age and Ageing* 34: 642 – 645, 2005.

Gallagher, B., Cen, L. and Hannan, E.L. "Validation of AHRQ's Patient Safety Indicator for Accidental Puncture or Laceration", http://www.ahrq.gov/downloads/pub/advances/vol2/Gallagher1.pdf

Gardella, C., Taylor, M., Benedetti, T., Hitti, J. And Critchlow, C. "The effect of sequential use of vacuum and forceps for assisted vaginal delivery on neonatal and maternal outcomes" *American Journal of Obstetrics and Gynecology* 185(4): 896 – 902, 2001.

Grobman, W.A., Feinglass, J. and Murthy, S. "Are the Agency for Healthcare Research and Quality obstetric trauma indicators valid measures of hospital safety?" *American Journal of Obstetrics and Gynecology* 195(3): 868 – 874 (2006).

Heylar, C., Flett, J., Hundert, M., Fallon, G., Mosher, G. and Crawford, R. "Benchmarking Comparisons of the Efficiency and Quality of Care of Canadian Teaching Hospitals", *Hospital Quarterly* Vol 1(3), 1998.

Hibbard, J.H., and Jewett, J.J. "Will Quality Report Cards Help Consumers?", *Health Affairs* 16:3, 1997.

Hibbard, J.H., Slovic, P., Peters, E. and Finucane, M. "Strategies for Reporting Health Plan Performance Information to Consumers: Evidence from Controlled Studies", *Health Services Research* 37(2): 291 – 313 (2002).

Hofer, T.P., Bernstein, S.J., Hayward, R.A. and DeMonner, S. "Validating quality indicators for hospital care", *Joint Commission Journal on Quality Improvement* 23: 455 – 467 (1997).

Hofer, T.P., and Hayward, R. "Identifying Poor-Quality Hospitals: Can Hospital Mortality Rates Detect Quality Problems for Medical Diagnoses?", *Medical Care* 34(8): 737 – 753 (1996).

Jha, A.K., Li, Z., Orav, E.J. and Epstein, A.M. "Care in U.S. Hospitals – The Hospital Quality Alliance Program", *The New England Journal of Medicine* 353(3): 265 – 274 (2005).

Knutson D.J., Kind, E.A., Fowles, J.B. and Adlis, S. "Impact of Report Cards On Employees: A Natural Experiment", *Health Care Financing Review* 20(1): 5 – 27 (1998)

Lakhani, A., Coles, J., Eayres, D., Spence, C. and Rachet B. "Creative use of existing clinical and health outcomes data to assess NHS performance in England: Part 1 – performance indicators closely linked to clinical care", *British Medical Journal* 330: 1426 – 1431 (2005).

Mant, J. and Hicks, N. "Detecting differences in quality of care: the sensitivity of measures of process and outcome in treating acute myocardial infarction", *British Medical Journal* 311: 793 – 796 (1995).

McKee, M. and Sheldon, T. "Measuring performance in the NHS", *British Medical Journal* 7128(316), 31 January 1998.

McQuay, H.J. and Moore, R.A. "Using Numerical Results from Systematic Reviews in Clinical Practice", *Annals of Internal Medicine* 126: 712 – 720, 1997.

Miller, M.R., Elixhauser, A., Zhan C., and Meyer G.S. "Patient Safety Indicators: using administrative data to identify potential patient safety concerns", *Health Services Research* 36:6(2): 110 – 132, 2001.

Ontario Hospital Association, Hospital Reports, (http://www.oha.com/oha/reprt5.nsf)

Pennsylvania Health Care Cost Containment Council (PHC4). "Paying for Performance – The Business Case" *fyi* 28, 2004 (http://www.phc4.org/reports/fyi/docs/phc4fyi28.pdf)

Rubenstein, L.Z., Powers, C.M. and MacLean, C.H. "Quality Indicators for the Management and Prevention of Falls and Mobility Problems in Vulnerable Elders", *Annals of Internal Medicine* 135(8:2): 686 – 693 (2001).
http://www.annals.org/cgi/content/full/135/8_Part_2/686#FN

Schneider, E.C. and Epstein A.M. "Use of Public Performance Reports. A Survey of Patients Undergoing Cardiac Surgery", *Journal of the American Medical Association* 279(20): 1638 – 1642 (1998).

Scott, I., Youlden, D. and Coory, M. "Are diagnosis specific outcome indicators based on administrative data useful in assessing quality of hospital care?", *Quality and Safety in Health Care* 13(1): 32 – 38 (2004).

Thomas, J.W., and Hofer, T.P. "Accuracy of Risk-Adjusted Mortality Rate As a Measure of Hospital Quality of Care", *Medical Care* 374(1): 83 – 92 (January 1999).

Thomson, R.G., McElroy, H. and Kazandjian, V.A. "Maryland Hospital Quality Indicator Project in the United Kingdom: an approach for promoting continuous quality improvement", *Quality in Health Care* 6: 49 – 55 (1997).

Vassallo, M. Sharma, J.C., Briggs, R.S.J. and Allen, S.C. "Characteristics of early fallers on elderly patients rehabilitation wards" *Age and Ageing* 32: 338 – 342 (2003).

Veronesi, U., Cascinelli, N. Mariani, L., Greco, M., Saccozzi, R., Luini, A., Aguilar, M. And Marubini, E. "Twenty-year follow-up of a randomized study comparing breast-conserving

surgery with radical mastectomy for early breast cancer" *New England Journal of Medicine* 347(16): 1227 – 1232, 2002.

Weissman, J.S., Ayanian, J.Z., Chasan-Taber, S., Sherwood, M.J., Roth, C., and Epstein, A.M. (1999). "Hospital Readmissions and Quality of Care", *Medical Care* 37(5), 490 – 501.

Witt, J. "A Finger on the Pulse: Comparative Models for Reporting the Quality of Hospital Care", *Atlantic Institute for Market Studies*, June 2005.

Atlantic Institute for AIMS Market Studies

YES! I want to support AIMS.

(An official tax receipt will be provided for your donation.)

AIMS is an independent economic and social policy think tank. Our objective is to broaden the policy debate to make Atlantic Canadians, and Canadians more generally, aware of the full range of options for resolving our economic and social problems, and the consequences of those options for our quality of life. To that end, AIMS is an active voice in public policy discussions, publishing practical analysis and policy recommendations. In order to maintain our independence,

To maintain our independence, AIMS takes no money from government. Our work depends entirely on the support of people like you.

I want to become:
- an individual supporter ($100 minimum)
- a corporate supporter ($1000 minimum)
- an in-kind supporter (event space/sponsorship, telecommunications, equipment, supplies)

Name: ———————————————————————————————

Title: ———————————————————————————————

Organization:———————————————————————————

Address: ———————————————————————————————

Telephone:———————————————— Facsimile: ————————————

E-mail: ———————————————————————————————

I am paying by: VISA Mastercard Cheque (enclosed)

Credit card #: ———————————————— Expiry Date:————————————

Name on Credit card: ———————————————— Signature:————————————

Please send or fax this form to 2000 Barrington Street, Suite 1302, Halifax, NS B3J 3K1
Telephone: (902) 429-1143 Facsimile: (902) 425-1393 E-mail: aims@aims.ca
For more information please check our website at **www.aims.ca**

Other AIMS work on health care:

A Finger on the Pulse by Julia Witt

The Good News: Pharmaceuticals and the Cost of Health Care by Julia Witt

Going Public on What is Private – the Canadian Health Care Consensus Group

A First Look at the Numbers – the Canadian Health Care Consensus Group

Definitely NOT the Romanow Report, by Brian Lee Crowley, Brian Ferguson, David Zitner, and Brett J. Skinner

Payment Is Powerful: Overcoming Canada's Shortage of GPs by Increasing Family Practice Compensation, by Ida Rayson

Operating in the Dark: The Gathering Crisis in Canada's Public Health Care System, by Brian Lee Crowley, Dr. David Zitner, and Nancy Faraday-Smith

Expenditure on Medical Care in Canada, by Brian Ferguson

The Non-Sustainability of Health Care Financing under the Medicare Model, by Brett J. Skinner

Medicare and User Fees: Unsafe at Any Price? by Carl Irvine and Dr. David Gratzer

Alice in Borderland: Why Canadians Cannot Afford to Be Complacent about US Drug Re-importation, by Brian Ferguson

This Won't Hurt a Bit: Why the Vaccine Crisis Shows We Shouldn't Believe What Health "Planners" Tell Us about How to Reform Drug Policy, by Brian Ferguson and Julia Witt

Why Not "Europeanize" Health Care? AIMS in The National Post, by Brian Lee Crowley and Johan Hjertqvist

Issues in the Demand for Medical Care: Can Consumersand Doctors Be Trusted to Make the Right Choices? By Brian Ferguson

Better Medicine, by David Gratzer

How Should We Decide What to Cover under Medicare? by Julia Witt

Profits and the Hospital Sector: What Does the Literature Really Say? by Brian Ferguson

Improving Canadian Health Care: Better Ways to Finance Medicare, by Brett J. Skinner

Public Health, State Secret, by Dr. David Zitner and Brian Lee Crowley

AIMS Books

Retreat from Growth: Atlantic Canada and the Negative-Sum Economy, by Fred McMahon

Road to Growth: How Lagging Economies Become Prosperous, by Fred McMahon

Looking the Gift Horse in the Mouth: The Impact of Federal Transfers on Atlantic Canada, by Fred McMahon (photocopies only)

Commentaries

Defining Atlantica: Bridges to Prosperity by Charles Cirtwill

Still More Equal than Others: Capped Equalization is still too much by Bobby O'Keefe

Health Care: Towards significant changes by Claude Castonguay

Moving on Up: The transition from poverty to prosperity by Charles Cirtwill

Taking the caller off hold: Move forward with telecom deregulation by Ian Munro

A new Golden Rule: The three Cs of local government by Charles Cirtwill

Taxing Incentives: How Equalization Distorts Tax Policy in Recipient Provinces, by Kenneth J. Boessenkool

Fiscal Equalization Revisited, by Professor James M. Buchanan, Nobel Laureate

Testing & Accountability: The Keys to Educational Excellence in Atlantic Canada by Charles Cirtwill, Rod Clifton, and John D'Orsay

2000 Barrington St., Suite 1302
Halifax NS B3J 3K1

Telephone: (902) 429-1143
Facsimile: (902) 425-1393
E-mail: aims@aims.ca
Web site: www.aims.ca