



"We Understand Health Care"

**When a Quality Measure Doesn't Measure Quality
A Case Study**

1999

When a Quality Measure Doesn't Measure Quality A Case Study

The explosive growth of interest in Quality Measures/Indicators and Report Cards by the general public is a positive development in the long run, but raises serious concerns given the current state of the art. Do the current measures in use actually measure quality, and if not, why not? This case study examines a commonly utilized indicator - "Surgical Complications" (sometimes termed "Iatrogenic Complications" or "Major Complications") for common procedures - Hip and Knee Joint Replacements.

Use - Measure or Indicator/Proxy?

The general public has become more and more concerned about health care quality, particularly in the last few years. Budgetary and HMO payment constraints increasingly affect the consumer in terms of shorter hospital stays and access to higher cost pharmaceuticals and procedures. Consumers are worried that these constraints are compromising the quality of the care they receive. At the same time, competition in the insurance marketplace has brought the premiums in many areas to within a few dollars of each other - moving the competition into the realm of Quality. The relatively recent availability of "Report Cards" and "Quality Measures" appears to be a much needed solution to their concerns. They want a bottom line answer - to believe that a published report card measure actually measures quality.

The more informed health care professionals recognize the relative infancy of the technology to measure quality and will use the term "Indicator" - a difference is noted, with a need to look deeper to understand why. A difference may be due to quality, but could also be due to:

- A different patient population (severity/risk)
- Different coding and documentation practice for the source data used
- Efficacy of the data indicator as a proxy to actually measure quality
- Inclusion of other dimensions (e.g. costs, staffing, market reputation, etc.)

It is well recognized that given the current state of the art, a given hospital can easily be at the top of one ranking but at the bottom of another due to differences in even the "best" of today's measures. This apparent discrepancy does not deter the commercial publishers of these rankings from asserting that their ranking actually measures quality. The consumer may believe these "experts" with their "objective data", particularly if they are not aware of conflicting results, and make potentially erroneous decisions regarding their care. Others who do observe these conflicting results may not trust any of the measures and lose any potential value.

This paper examines joint replacement surgery (hip and knee). This surgery is performed to increase mobility and reduce pain for patients affected by rheumatoid arthritis or osteoarthritis and/or fractures. Most would agree that "true" measures of quality would include mortality (death) rate and measures of improved mobility and reduced pain. Raw mortality rates can be easily measured, however is rare rare in this population, making it more difficult to risk adjust for patient severity. In addition, the low frequency of occurrence makes it less commercially desirable for use in ranking hospitals since there is little significant variation. Measures of improved mobility and reduced pain are not routinely available today, and are subject to observation and reporting differences. A threshold proxy for these measures is implant life

before a “revision” replacement or repair is needed to again increase mobility or reduce pain. This can be measured to some extent, but requires tracking a patient across multiple years. Future data collection refinements are either in process (e.g ICD-10 coding to replace the current ICD-9) or should be considered.

Today, however, most commercial report cards are based on (and limited by) the administrative data which is easily available nationally - the MEDPAR files based on Medicare billings. These files report age, sex, race, admission status, discharge and mortality status, charges, diagnoses, and procedures for primarily the over 65 population as coded by each hospital. Measures and Indicators are therefore limited to the proxies which can be derived from this data. The most common measure selected is the presence of one or more ICD-9 diagnostic codes which the report card publisher considers to be a “Complication,” with implicit assumptions that it was caused by the hospital or physician and that fewer is better. This may be “risk adjusted” to take patient severity into account to varying degrees.

Composition of the Indicator

Some report cards present data hospital wide, while others present data for a particular specialty (e.g Orthopedics) or for a particular procedure (e.g. Hip or Knee replacement.) More specific measures will account for diagnostic and procedure mix between hospitals, but be more sensitive to individual patient severity/risk issues, data documentation/reporting, and clinical practice differences. The indicator may be a single measure (e.g “Complication” rate) or a composite measure which looks at mortality, complications, staffing, costs, referral patterns, etc. Inclusion of multiple clinical measures such as mortality, complications, and referral patterns attempts to compensate for severity/risk, data/reporting, and practice differences in the measure itself.

Risk Adjustment

Another way to account for differences in patient characteristics and predicted risks is a “risk adjusted” predicted value. Risk adjustment factors which are easily available include the patient’s admitting diagnosis, age and sex, co-morbid conditions (another diagnosis which affects the intensity of the care), etc. Other risk factors which require supplemental data collection include the patient’s weight and height, prior medical history, specific laboratory values, degree of deformity of the skeletal system, etc. The intent is to take into account all of the characteristics of the patient which could affect their likely outcome, so that any remaining variation is due to the actual care provided. Absent such adjustments, any hospital which admits more difficult cases would face lower outcome scores/ranking, perversely constraining access to the patients most in need of care. For example, the hospital with the most aberrant death rate when the Health Care Financing Administration released initial hospital level mortality figures was a hospice caring for the terminally ill. Most current severity adjustors explain 6 - 40% of variation between hospitals. Some, however, explain less than one percent of the statistical variation (R-squared). One commercial site’s risk adjuster only explains 0.8% of the variation in Hip replacement indicators and 0.5% of the variation in Knee replacement indicators. This is not statistically different than saying that “All Hip replacement patients are the same” and that “All Knee replacement patients are the same.” This lacks face validity and ignores the reality of referrals to regional centers of excellence for the obese or hypertensive orthopedic patient.

Specialty Referral Patterns

The past decade has seen the movement toward “Regional Centers of Excellence” for many

conditions (Transplants, Open Heart Surgery, Neo-Natal Intensive Care, Burns, etc.) Study after study has demonstrated improved quality with increased experience and volume - practice does make perfect. Physicians routinely refer “difficult” cases to their colleagues who have more experience treating these severe cases. The specific condition or combination of conditions may not be routinely reported in easily available administrative data. Many report cards therefore utilize the referral patterns themselves as a proxy for the clinical judgement which resulted in the referral. Some typical proxies include market share of the local market, market share of patients from outside the local marketplace, and overall volume for a particular procedure. Measures which do not utilize such referral proxies assume that the patient differences are accounted for in other risk adjustors or that all patients are the same. It is in exploring a low indicator score (which did not have either a referral adjustment or an effective risk adjuster) for an Orthopedic Center of Excellence that the following was noted.

The Use of One ICD-9 Code Matters

Out of 26 codes used as quality indicators on this measure, one code (997.2 - Peripheral Vascular Complications) accounted for half of the Hip score and 80% of the Knee score on this measure for the time period covered. Inquiries at the hospital uncovered the fact that during this time period, the hospital was participating in a research effort to evaluate the effectiveness of various prophylactic measures to prevent Deep Vein Thrombosis (DVT - clots). The clinical protocol called for (and paid for) Venograms for all Hip and Knee replacement patients. DVT, particularly in the calf, will occur in 50-60% of these patients when left untreated. Treatments include anticoagulants (blood thinners) such as aspirin, warfarin, and various types of heparin and Intermittent Pneumatic Compression Devices which massage the calf and/or thigh. DVT's are a concern, since some clots may progress and cause a Pulmonary Embolism. Treatments were able to reduce DVT to the 25 - 35% range.

Results of this research were widely disseminated to Orthopedic surgeons around the world. Most will now routinely provide prophylactic treatment to all of their Hip and Knee patients, and many will not even utilize a Venogram or Doppler Ultrasound to look for clots unless symptoms occur (swelling, redness, or pain in the leg.) The Hospital utilized the medical record to capture these post-operative DVT's for the study. These same codes were subsequently reported to State and Federal agencies and found their way into external data bases such as the MEDPAR data base used to develop “Quality” measures. We explored the same measures for other hospitals where this research and routine testing occurred and in many cases, these hospitals reported the highest “Complication” rates (and therefore the lowest hospital rankings) in their metropolitan area. They also accounted for 16% of the statistical variation for Hips and 68% of the statistical variation for Knees in a broader sample of Metropolitan areas.

Documenting in the Medical Record and Coding into the Bill

Another of the hospitals which reported DVT rates in research similar to the others involved in the study showed up with low “Complication” rates (and therefore the highest hospital rankings.) They either completed their research prior to the study period for the indicator or did not code the DVTs found as a surgical complication. Coding guidelines in this area are open to interpretation since if the physician describes the condition as a complication, caused by the surgery, and is a more than routine expected condition - the 997.2 would be coded. If the physician considers a 25 - 40% rate to be “routine” and does not indicate that it is a complication, it would not be coded. We also noted that a staff person of the college affiliated

with this hospital developed the “Complication” measurement methodology, which could also make staff at this hospital more sensitive to use of this code. A number of studies have suggested that important complication and co-morbidity risk factors tend to be under coded (i.e. not reported) in the bill, particularly where it does not affect payment, such as Hip and Knee replacements (they are all in a single DRG - 209.)

For other hospitals which do not even do the testing unless the patient is symptomatic (typically 1% or less of patients treated prophylactically), few cases would be coded, since 99% treatment would be considered “routine.” There is an obvious disadvantage to rigorous testing and coding in this area and a 10:1 difference in reporting the identical conditions from one hospital to another. Any “Quality” conclusions drawn from such data are obviously suspect. This code should at minimum be removed from quality measures for Hip and Knee replacements, otherwise, hospitals which provide rigorous testing and fine tuning of care (e.g. minimal treatment if no clots are identified, modifying medication to dissolve the clot, or insertion of a “Greenfield filter” to catch a clot prior to progression to PE) will be perversely considered “Lower Quality.”

Recommendations

- Remove ICD-9 code 997.2 - Peripheral Vascular Complications associated with DVT from current outcome measures for Hip and Knee replacement (ICD-9 procedure codes 81.51 - 81.55) since:
 - It occurs frequently enough that it is routinely worked up and treated prophylactically for most patients
 - It is confirmed in only those facilities which routinely perform venograms and/or Doppler Ultrasounds
 - Some physicians may consider DVT a complication (particularly when researching ways to reduce its incidence) while others do not
 - These factors cause a significant inconsistency in reporting - it is routinely reported in less than 3% of cases, while DVT actually occurs in 30-50% of cases
 - This wide variation in how the exact same clinical circumstances would be reported from one hospital to another invalidates the measure
- Encourage the American Hospital Association (AHA) and the National Center for Health Statistics (NCHS) to develop more detailed guidelines for ICD-9 (and future ICD-10) codes used or likely to be used as Quality Indicators (similar to details produced for codes that are used for payment.) This will produce more meaningful and reliable measures in coming years.
- Any presentation of “Complication” results must responsibly report findings to consumers, and acknowledge the current state of the art is troubled by many issues such as the one described here
- Consumers should also take the responsibility to apply the proper skepticism when reviewing Report Cards and Quality Measures and expect sufficient disclosure of the methodology used to be able to make an informed decision as to their use.

There is an old saying that goes “Everything looks like a nail if your only tool is a hammer.” Unfortunately, this saying is all too familiar in today’s world of “report card” measures for hospitals derived from MEDPAR administrative billing data. We can and should do better.