214: EMPIRICAL METHODS TO IMPROVE THE RELIABILITY OF QUALITY INDICATORS FOR MEASURING PROVIDER PERFORMANCE

Geppert, J.J., Davies S.M., McDonald K.M., Romano P.S.

Objective:
An important limitation of quality indicators is their imprecision, which complicates the reliable identification of persistent differences among providers in performance. The imprecision in quality indicators arises from two sources. The first is "random" variation (i.e. variation characterized by the central limit theorem and the normal distribution), which is a particular problem for indicators based on a small numbers of patients per provider. The amount of variation due to "random" variation is often large relative to the total amount of provider level variation that is observed in any given quality indicator. A second source of imprecision arises from non-persistent factors that are not sensitive to the size of the sample; for example, a severe winter results in higher than usual rates of pneumonia mortality. Both small samples and other one-time factors that are not sensitive to sample size can add considerable volatility to quality indicators. Also, it is not the absolute amount of imprecision that matters, but rather the amount of imprecision relative to the underlying signal (i.e., true provider level variation) that dictates the reliability of any particular indicator. Even indicators based on relatively large samples with no non-persistent factors at work can be imprecise if the true level of variation among providers is negligible. We applied empirical methods to address both sources of imprecision, thereby improving the reliability of quality indicators based on administrative data.

Methods:
Our approach to account for the imprecision or lack of reliability is a generalization of the idea of applying a "shrinkage factor" to each provider's estimate so that less reliable estimates are shrunk toward the peer group average. These "filtered" estimates are a combination of the provider's own quality indicator, the peer group average, and the provider's quality indicators from past years or other quality indicators. To form the optimal combination one must know the amount of noise and signal variance in each indicator, as well as the correlation across indicators in the noise and signal variance. We estimate the noise variance (and covariance) in a straightforward manner for each provider, based on the number of patients on which each indicator is based. To estimate the signal variance (and covariance) for each quality indicator, we subtract the noise variance from the total variance observed in each indicator across providers (which reflects both signal and noise variance). In other words, the observed variation in quality indicators is sure to overstate the amount of actual variation across providers (because of the noise in the indicators). Therefore, we estimate the amount of true variation in performance based on how much the observed variation exceeded what would have been expected due to sampling error. Importantly, our method does not *assume* that provider performance is correlated from one year to the next (or that performance is correlated across indicators). Instead, we estimate these correlations directly from the data, and incorporate information from past years or other indicators only to the extent that these empirically estimated correlations are large.

Results:
We evaluated the reliability of the AHRQ Inpatient Quality Indicators constructed on the 1997 State Inpatient Data, including five measures of utilization, six measures of in-hospital morality for medical conditions, and seven measures of post-operative in-hospital mortality. We found that the utilization measures tended to be the most reliable, followed by the medical mortality measures, and with the postoperative mortality measures being the least reliable (primarily because few providers performed the procedures and mortality was relatively infrequent). The "signal ratio" is a measure of the fraction of observed provider level variation that appears to be *truly* related to systematic differences across providers, and not random variation. For the utilization measures, the signal ratios generally exceeded 0.70. For the medical mortality measures, the signal ratios ranged from 0.40 to 0.70. For the postoperative mortality measures, the signal ratio generally was less than 0.40.

Conclusions:
All quality indicators, but in particular measures of patient outcomes, have a large share of observed variation among providers that is due to random variation, making it more difficult to

compare performance among providers, and to identify "best practices" or construct summary measures of quality. The empirical methods applied in this project have the potential to greatly improve the reliability of quality indicators based on administrative data, thereby making them more useful for quality improvement, network management, and public reporting.

Davies S, Geppert J, et al. Refinement of the HCUP Quality Indicators. Technical Review Number 4 (Prepared by the University of California San Francisco-Stanford Evidence-based Practice Center under Contract No. 290-97-0013). AHRQ Publication No. 01-0035, May 2001.