How to assess the quality of care between hospitals: An example from the Berlin Myocardial Infarction Registry (BHIR)

Authors: B. Maier, E. Vettorazzi, S. Behrens, A. Fried, J.F. Kersten, H. Theres, W. Thimme, K. Wegscheider

Purpose: The ranking of hospitals according to their quality of care with respect to process or outcome parameters is one of several approaches to improve hospital performance. The statistics behind these rankings are frequently rather simple when crude arithmetic means or simple one-way analyses of variance are applied. These simple approaches do not take into account the hierarchical nature of the data and the different challenges the hospitals have to face for differences in patient mix and environment. Our study was aimed at showing that it is possible to compare the quality of care between departments of cardiology on a statistically sound basis with adjustment for differences in patients' characteristics.

Methods: The BHIR is an ongoing prospective acute myocardial infarction registry including all Troponine positive ACS Patients who reach one of the participating hospitals within 48 hours after symptom onset. The analysis is on 1767 patients included in 11 Berlin hospitals in 2004 or 2005. Hospital mortalities were compared by fitting a two-level random effects model with patient characteristics as covariates to the data. The resulting mortalities are Empirical Bayes (EB) estimates adjusted for differences in patient populations between hospitals. This method was chosen to establish fair comparisons and to account for the extra variability caused by random effects.

Results: Hospital populations revealed large significant differences in many respects, e.g. in mean age (span: 12 years), percentage of females (range 27%-52%), CHF on admission (2%-48%). Crude hospital mortalities also differed considerably between departments (3.1%-21.7%). After Bayesian shrinkage only one hospital remained with a hospital mortality significantly above average. After adjustment for the differences in patients' characteristics, the range was reduced to 5.3% – 12.2%. Based on these figures we were not able to demonstrate any significant differences between hospitals. The order of the hospitals was considerably shuffled as compared to the order of the crude means.

Conclusion: The analysis demonstrates that the naïve comparison of hospitals by crude means (here: mortalities) may be unfair and misleading. A statistical analysis that takes population differences and random effects into account may result in different conclusions. Two-years of data collection may not be enough to demonstrate relevant differences between average-size German city hospitals. We suggest a minimum of three years of data collection for mortality comparisons.