Bayesian models are increasingly fit to large administrative data sets and then used to make individualized recommendations. In particular, Medicare’s Hospital Compare webpage provides information to patients about specific hospital mortality rates for a heart attack or Acute Myocardial Infarction (AMI). Hospital Compare’s current recommendations are based on a random-effects logit model with a random hospital indicator and patient risk factors. Except for the largest hospitals, these recommendations or predictions are not individually checkable against data, because data from smaller hospitals are too limited. Before individualized Bayesian recommendations, people derived general advice from empirical studies of many hospitals; e.g., prefer hospitals of type 1 to type 2 because the observed mortality rate is lower at type 1 hospitals. Here we calibrate these Bayesian recommendation systems by checking, out of sample, whether their predictions aggregate to give correct general advice derived from another sample. This process of calibrating individualized predictions against general empirical advice leads to substantial revisions in the Hospital Compare model for AMI mortality, revisions that hierarchically incorporate information about hospital volume, nursing staff, medical residents, and the hospital’s ability to perform cardiovascular procedures. And for the ultimate purpose of meaningful public reporting, predicted mortality rates must then be standardized to adjust for patient-mix variation across hospitals. Such standardization can be accomplished with counterfactual mortality predictions for any patient at any hospital. It is seen that indirect standardization, as currently used by Hospital Compare, fails to adequately control for differences in patient risk factors and systematically underestimates mortality rates at the low volume hospitals. As a viable alternative, we propose a full population direct standardization which yields correctly calibrated mortality rates devoid of patient-mix variation.

(This is joint research with Veronika Rockova, Paul Rosenbaum, Ville Satopaa and Jeffrey Silber).