

## Does Hospital Size Affect Our Ability to Accurately Identify High Quality Care in Pay-for-Performance Programs?

### Key Findings

- Smaller hospitals experience much greater variability in performance scores as a result of statistical sampling characteristics than larger hospitals.
- Pay-for-performance methods that identify relative quality from single-year ranks based on composite scores are especially problematic for small hospitals. Small hospitals are more likely to not receive a financial reward when their true rank is actually in the highest quality category. Conversely, they are also more likely to receive a financial reward when their true rank is not in the highest quality category.
- Assessing the likely impact of uncertainty associated with small hospital size and related statistical limitations is important to the successful application of pay-for-performance programs. Minimizing related effects is essential to achieving accurate reflections of performance of small size hospitals.

### Introduction

The Premier Hospital Quality Incentive Demonstration (HQID) is a project sponsored by the Centers for Medicare and Medicaid Services (CMS) and conducted with participating hospitals that are members of Premier Inc., a

national alliance of non-profit hospitals. The demonstration was designed to examine whether a system that explicitly pays-for-performance can lead to system-wide improvement in the quality of care that hospitals provide for selected medical conditions.

For the HQID, high performance and some aspects of low performance are defined in a relative sense by ranking hospitals' performance in providing specific services. Financial bonuses are awarded to hospitals in the top 20 percent of performance rankings for each condition.

Hospitals vary greatly in the annual number of patients seen with acute myocardial infarction (heart attack), heart failure, and pneumonia, the conditions we examined in this study. Statistical theory suggests that smaller hospitals can expect to experience much greater sampling variability in their performance scores than larger hospitals.

## Purpose of the Study

Statistical models were constructed to assess whether hospital size impacts our ability to identify "true" hospital ranks using data generated by pay-for-performance programs patterned after HQID. We used HQID data from Year 1 and CMS' Hospital Compare data to address the following questions:

- How accurately can we expect to identify "true performance scores" for hospitals participating in HQID?
- What unintended consequences might arise in a system that establishes rewards and penalties based on relative performance when relative performance is measured with substantial variation in accuracy?

## Results

There is a dramatic inverse relationship between the size of the hospital and its expected range of ranking positions for its "true" or stabilized mean rank. For the three conditions we examined, the smallest hospitals in the dataset would likely experience five to seven times more uncertainty than the largest hospitals concerning their "true" ranks. This means that the "true" ranks for the smallest hospitals could actually be much higher or much lower. A hospital with 20 or fewer pneumonia patients, for example, would have a percentile-rank that spans 64 percentile points, while a hospital with 1,100 pneumonia patients would have a percentile rank that spans only 10 percentile points.

We conclude that:

- All estimates of rank/percentile need to include adequate measures of uncertainty of those estimates.
- Identifying relative quality from single-year ranks based on composite scores will impact smaller institutions more severely than larger institutions. Smaller hospitals are more likely to be included in the top category of ranks and receive a reward when their true rank is not actually in that category, and to be excluded from it and not receive a reward when their true rank is in the category.
- Increased uncertainty about a hospital's "true" relative quality level—arising from considerable annual variability in achieving the placement required for rewards and recognition—could impact its motivation to achieve higher quality levels.
- The likelihood and consequences of high levels of uncertainty concerning hospitals' relative levels of quality differs by specific medical condition, but in all cases this uncertainty has important implications for policy.
- The above findings are likely to generalize to hospitals beyond this sample.
- The results may understate the degree of uncertainty likely to be found in more mature pay-for-performance programs using simple ranks like the HQID. After several years of operation, the distribution of composite scores for all conditions would be expected to more closely cluster together, making ranks even more volatile.
- We have begun to address the policy-relevant issue of identifying and estimating the likely amount of uncertainty inherent in measuring relative quality through the ranks of composite scores. Assessing the likely impact of this uncertainty for pay-for-performance programs, and identifying ways of minimizing its effects, will be critical to their success.

## About the Study

A Bayesian, hierarchical modeling strategy was used to estimate the uncertainty associated with the ranking of hospitals by their raw composite score values for three medical conditions – acute myocardial infarction, heart failure, and community acquired pneumonia.

Approval was obtained from Premier Inc. and from CMS to use the HQID data for Year 1. Our specific goal was to assess the implications of hospital size for the amount of uncertainty likely to exist in ranks of hospital annual composite scores. This study is not intended to be an evaluation of performance improvement brought about by the HQID.

The hospitals participating in the HQID are not representative of the full population of short-term, general hospitals in the U.S. They include only three critical access hospitals (CAHs) and 44 rural hospitals. Currently CAHs constitute 23 percent of all short-term, general hospitals. Since the goal of the project was to specifically show the influence of small hospital size per se on the likely variability in hospital ranks, we conducted a second set of model runs using additional data gathered from the CMS Hospital Compare Program.

The information in this policy brief is based on Upper Midwest Rural Health Research Center Working Paper #3: Hospital Size, Uncertainty and Pay-for-Performance by Gestur Davidson, PhD and Ira Moscovice, PhD, University of Minnesota, and Denise Remus, PhD, RN, formerly, Premier Inc.

The Working Paper is available at  
[http://www.uppermidwestrhc.org/pdf/hospital\\_size.pdf](http://www.uppermidwestrhc.org/pdf/hospital_size.pdf)

*Support for this Policy Brief was provided by the Office of Rural Health Policy, Health Resources and Services Administration, PHS Grant No. 5U1CRH03717-02-00.*

For more information, contact: Gestur Davidson, (612) 625-5182, or by email at [david064@umn.edu](mailto:david064@umn.edu)

## Partners

Ira Moscovice, PhD, Director  
University of Minnesota  
Rural Health Research Center  
School of Public Health  
2221 University Ave SE, Suite 112  
Minneapolis, MN 55414  
Phone: (612) 624-6151  
<http://www.hsr.umn.edu/rhrc>

Mary Wakefield, PhD, Deputy Director  
University of North Dakota  
Center for Rural Health  
School of Medicine & Health Sciences  
501 N. Columbia Road Stop 9037  
Grand Forks, ND 58202-9037  
Phone: (701) 777-3848  
<http://medicine.nodak.edu/crh>