An Analytical Approach to Improving Physician Performance

By Robert Sutter, RN, MBA, MHA, Brian Waterman, MPH, and Michael Udwin, MD, FACOG

In this article...

Examine strategies used to gauge physician performance variability and learn how to develop ways to engage physicians.

Health care reform has strengthened the link between performance and reimbursement, exemplified by value-based purchasing and accountable care organizations. This has led to a heightened responsibility for physician leaders to improve physician performance.

However, providing actionable assessments of physician outcomes is not as straightforward as it may seem. The physician leader is faced with myriad choices regarding how to evaluate physician performance and implement an improvement strategy.

Fortunately there is a thoughtful, analytically based approach that enhances the success of physician improvement initiatives. Physician leaders should answer the following three questions for the clinical conditions under consideration in order to develop a performance improvement strategy:

- What proportion of total performance variability is attributable to physicians?
- Are there statistically significant differences in physician performance?
- How is physician performance distributed across the outcome categories of "better than expected," "as expected," and "worse than expected?"

The answers to these questions will notably enhance establishing relevant, quantifiable performance improvement objectives. Further, this will guide implementation of the best strategy to affect change.

Ultimately, harnessing the data at the physician leader's disposal and conducting analytics to answer the questions will provide the information required to successfully engage physicians and improve organizational performance.

Physician variability

All health care outcomes (length of stay, cost, mortality, etc.) have two dimensions or process factors that contribute to the overall variability in achieved performance:

- Organizational factors such as policies, procedures, staffing, etc.
- Physician practice patterns.

Quantifying what percentage of the total variability of physician practice patterns is strategically valuable information. With this knowledge, a performance improvement strategy can begin to be formulated.

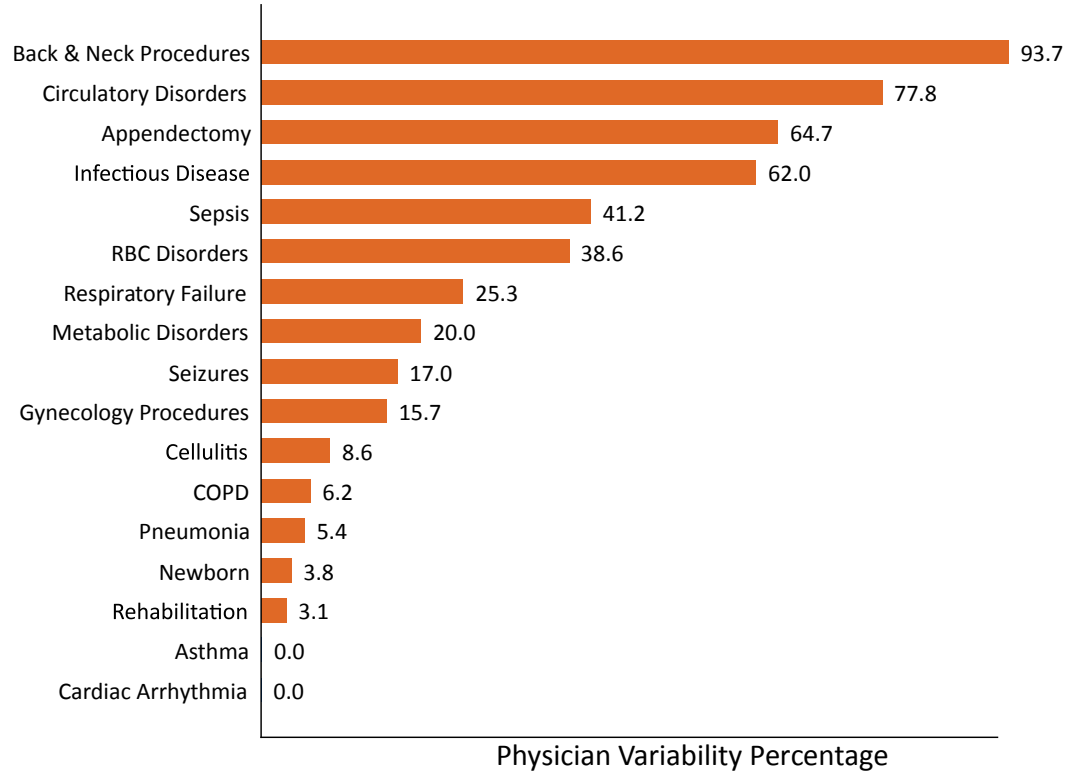
Graph 1 depicts the percentage of total variability in risk-adjusted excess length of stay (defined as observed length of stay minus expected length of stay) that is attributable to physician practice patterns for clinical conditions that have an opportunity for improvement.

Engaging physicians to explore and standardize practice patterns in order to reduce variability in risk-adjusted excess length of stay for back and neck procedures is the strategy of choice for this clinical condition, since physician practice patterns account for 93.7 percent of the total variability.

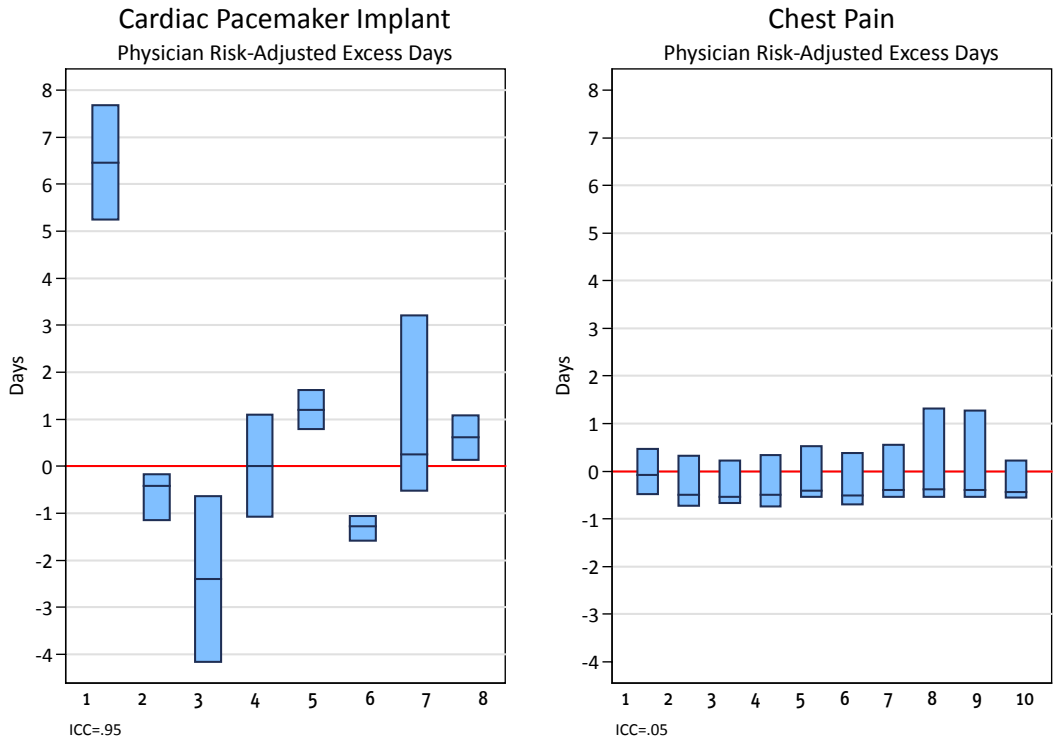
On the other hand, if pneumonia is the clinical condition selected for improvement, focusing the improvement efforts on organizational factors is the strategy that will yield the greatest benefit, since it represents 94.5 percent of the variability in contrast to physician practice patterns, which represents only 5.4 percent.

This information provides the physician leader clarity in selecting the appropriate strategy to address physician opportunities for improvement, Without this information, a physician leader may explore physician practice patterns

Graph 1



Graph 2

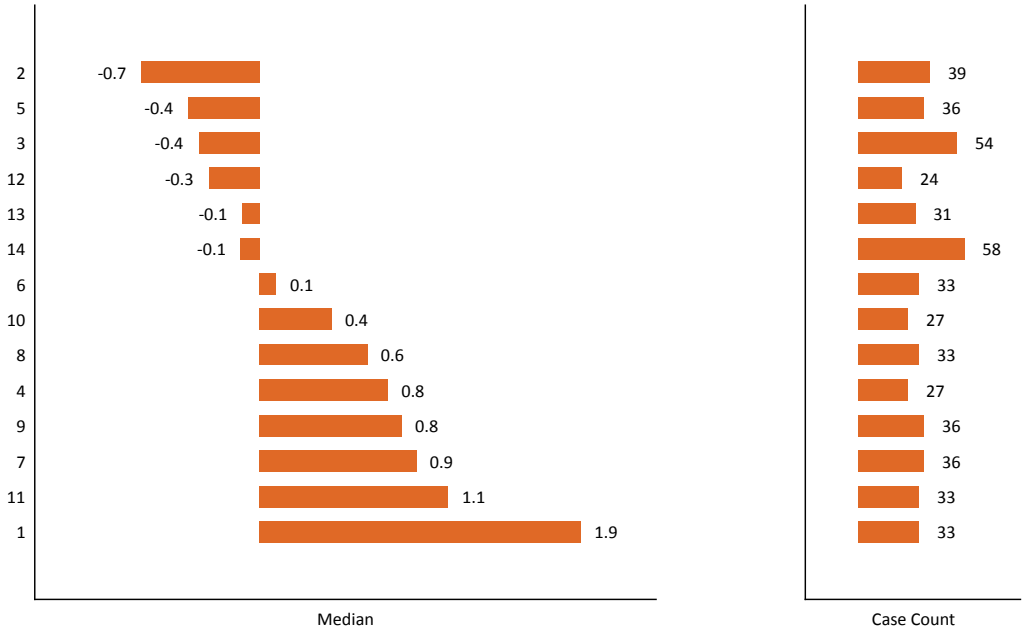


Graph 3

Vascular Procedures

Attending Physician Performance

Risk-Adjusted Excess Days



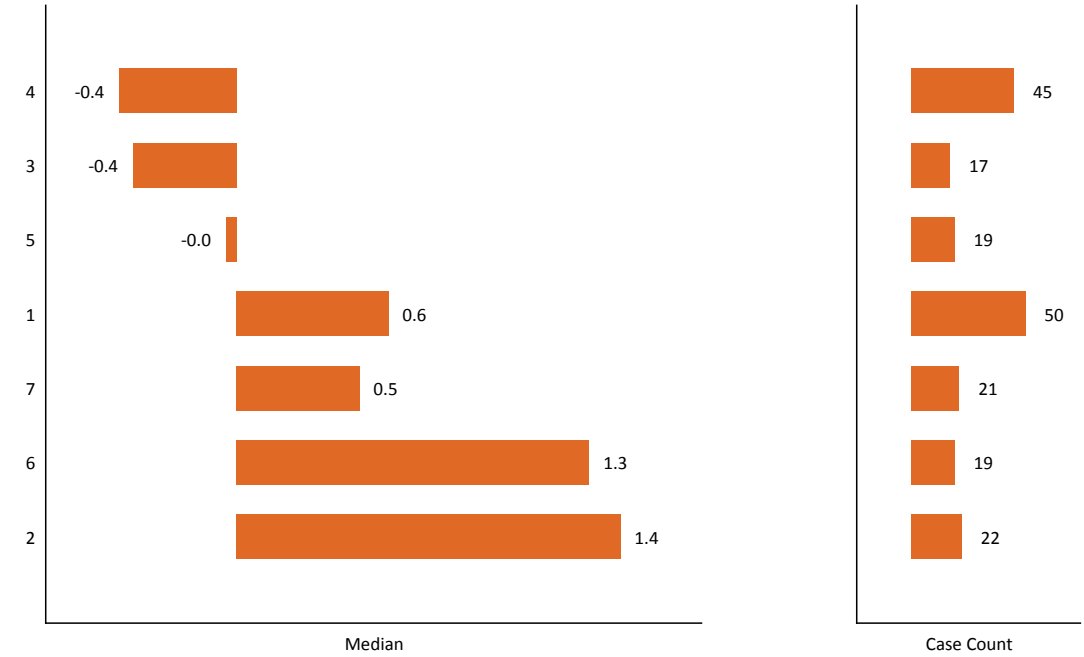
P<0.05

Graph 4

Renal Failure

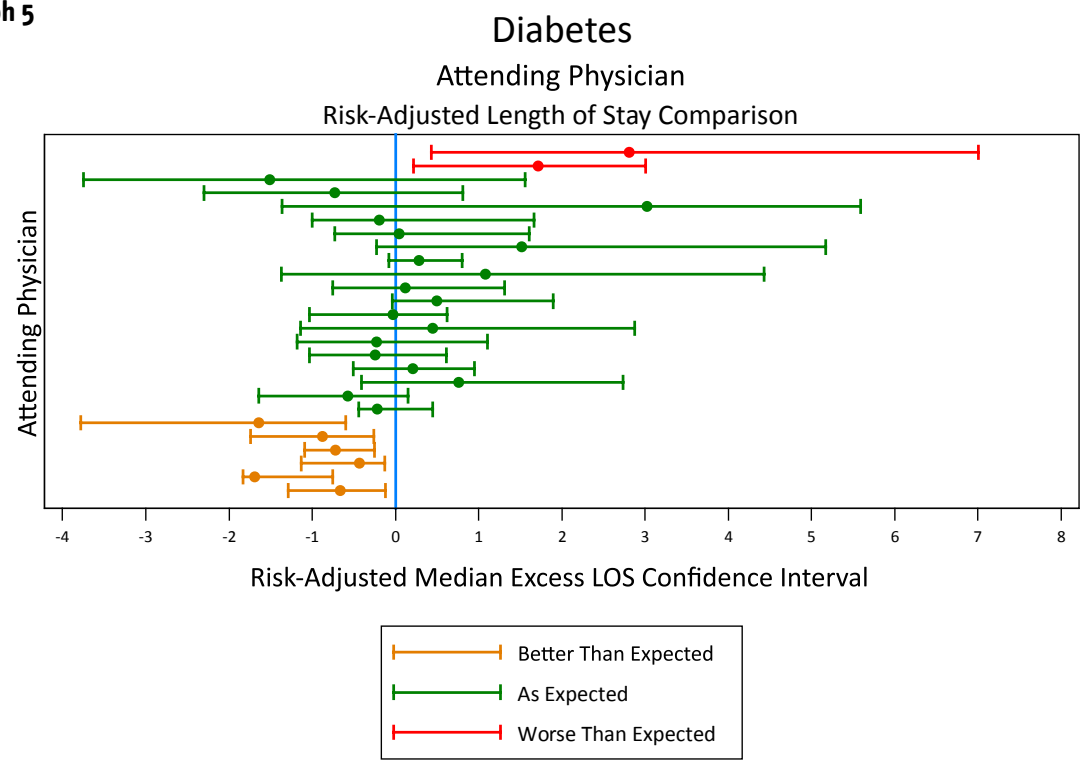
Attending Physician Performance

Risk-Adjusted Excess Days

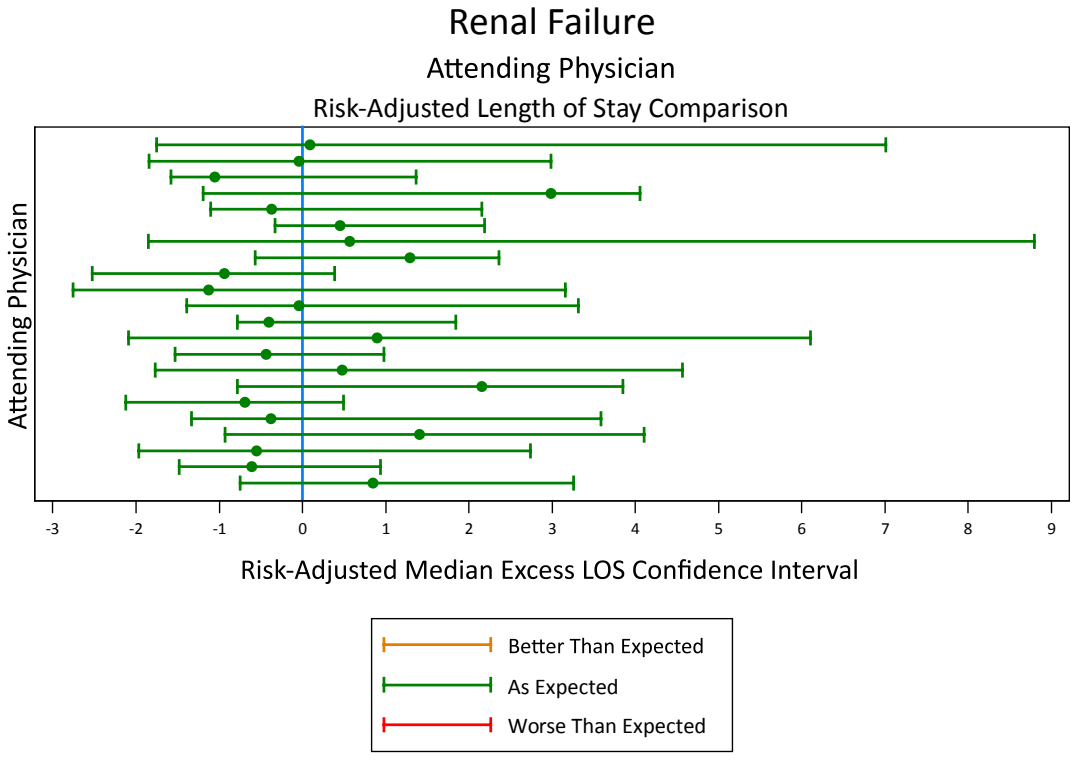


P>0.05

Graph 5



Graph 6



for pneumonia with the hope of performance improvement, only to be disappointed with the results and potentially incurring resentment and disengagement among physicians in the process.

The physician variability percentage is the intraclass correlation coefficient derived using hierarchical regression techniques. The intraclass correlation coefficient (ICC) answers the question: "To what degree is a patient's outcome dependent upon which physician treats them?" The ICC ranges from zero to one. A high number indicates that a patient's outcome is highly dependent upon the physician treating them while a low number implies that a patient's outcome will essentially be the same regardless of what physician treats them.

Graph 2 depicts two scenarios: The left graph displays physician performance in cardiac pacemaker implantation where the ICC is high -.95; and the right graph represents a low ICC -.05- scenario for chest pain.

Note the stark difference in risk-adjusted excess length-of-stay performance among physicians inserting cardiac pacemakers. Patients treated by physicians 1, 5 and 8 have a median risk-adjusted length-of-stay excess that exceeds zero, whereas patients treated by physicians 2, 3 and 6 have a median risk-adjusted length of stay excess less than zero.

In this scenario a patient's risk-adjusted excess length of stay is dependent upon which physician treats them. On the other hand all the physicians treating patients for chest pain have a median risk-adjusted length-of-stay excess between zero and -1 days. Therefore, a patient's risk-adjusted excess length of stay is not dependent upon which physician treats them.

One can readily see that clinical conditions with a high ICC indicate more variability in physician outcomes than clinical conditions with a low ICC. As a result, clinical conditions with a high ICC represent a greater opportunity to reduce outcome variability by focusing on physician performance compared to clinical conditions with a low ICC.

Physician performance

Determining whether statistically significant differences in physician performance exist provides another piece of information that assists the physician executive in deriving the improvement strategy.

If the goal is to improve risk-adjusted excess length of stay, the question to answer is whether risk-adjusted excess length of stay differs significantly among physicians?

If the answer is yes, then reducing variability for risk-adjusted excess length of stay among physicians will likely yield meaningful improvements in performance. If on the other hand the answer is no, then variability in physician performance does not exist, and any attempt to reduce this variability will likely not yield meaningful results.

To illustrate, let us take two examples of physician performance variability and use the common p-value of <0.05 to determine whether statistically significant differences exist in physician performance.

Graph 3 depicts the median risk-adjusted excess length of stay for attending physicians treating patients undergoing vascular procedures. The graph depicts performance that ranges from a median of -0.7 to 1.9 days.

The one-way test of significance for these data yields a p-value of 0.0001. Since this value is less than 0.05, the answer to the question, "Is there a statistically significant difference in risk-adjusted excess length of stay among physicians" is yes. Therefore, reducing the variability in physician performance will likely yield meaningful improvement.

The scenario depicted in Graph 4 for renal failure depicts variation in median performance that ranges from a 0.4 to 1.4 days, and the one-way test of significance for these data yields a p-value of 0.4873.

Given that this probability is considerably greater than 0.05, the answer to the question, "Is risk-adjusted excess length of stay significantly different among physicians" is no. Simply put, there is insufficient evidence in the data to conclude that there is true variability in physician performance.

A physician leader could not expect that attempts to reduce physician performance variability would yield meaningful improvement results.

Physician performance classification

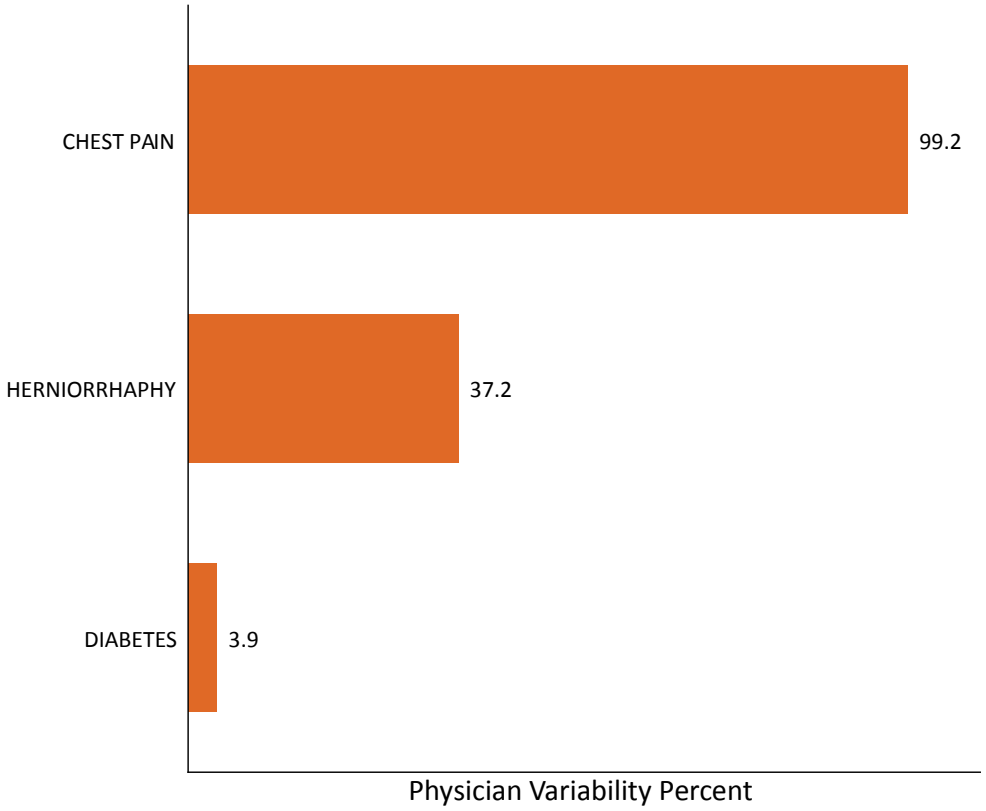
The last piece of information in evaluating physician performance is the distribution of outcomes across the performance categories of "better than expected," "as expected" and "worse than expected."

The expected component of this measurement is the predicted risk-adjusted outcome based on the severity of illness among the physician's patient population. In our earlier examples, we derived risk-adjusted excess length of stay by subtracting a patient's expected length of stay from their observed length of stay. Negative differences represent a shorter length of stay than expected, zero differences represent a length of stay that is equal to expected, and positive differences signify a length of stay that is longer than expected.

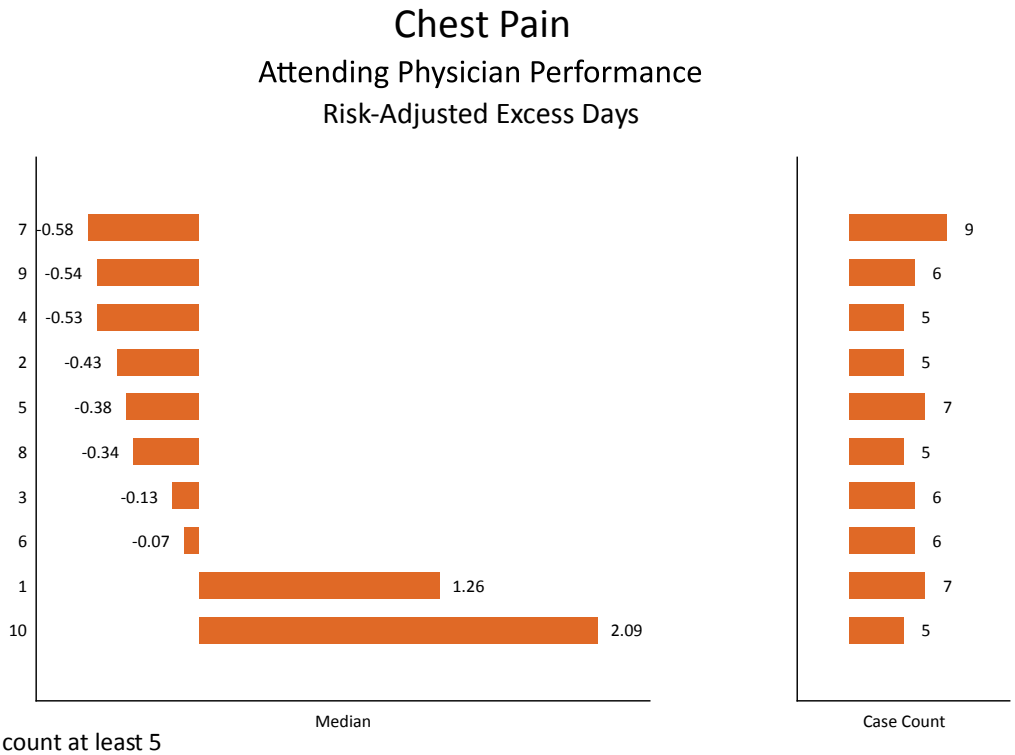
By statistically summarizing these differences across a group of patients attributed to a specific physician, a physician leader can determine whether systematic departures from risk-adjusted expected length of stay are present.

Classifying physician performance involves deriving risk-adjusted confidence intervals and comparing each physician's confidence interval to an

Graph 7



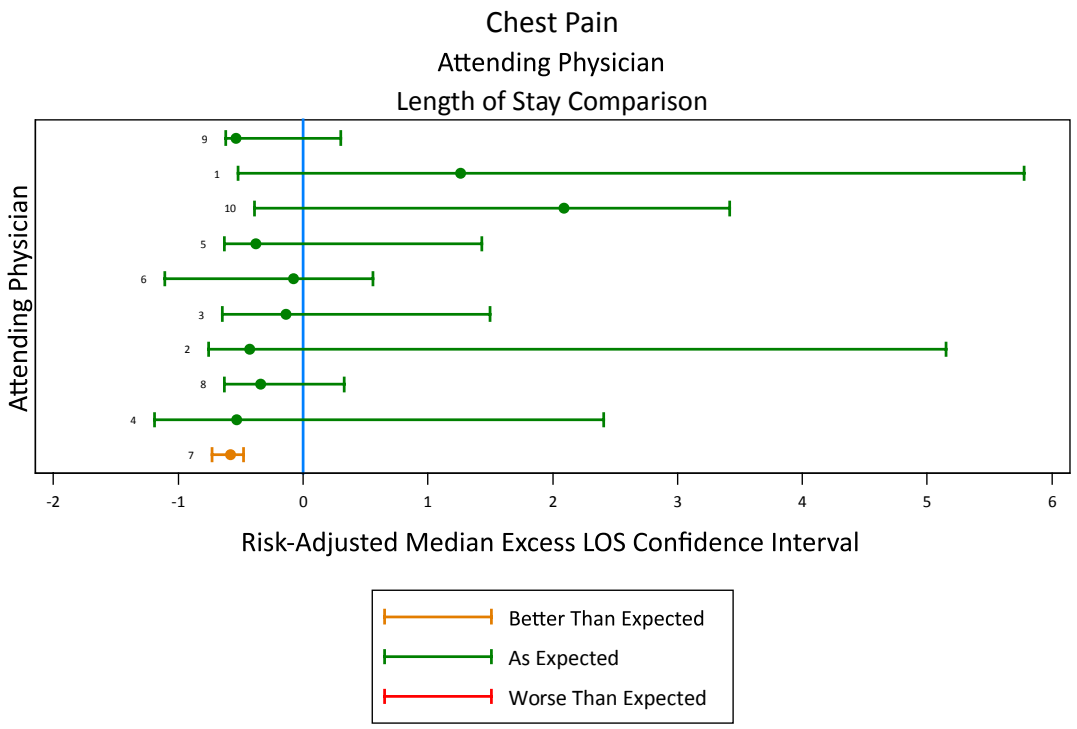
Graph 8



Physician case count at least 5

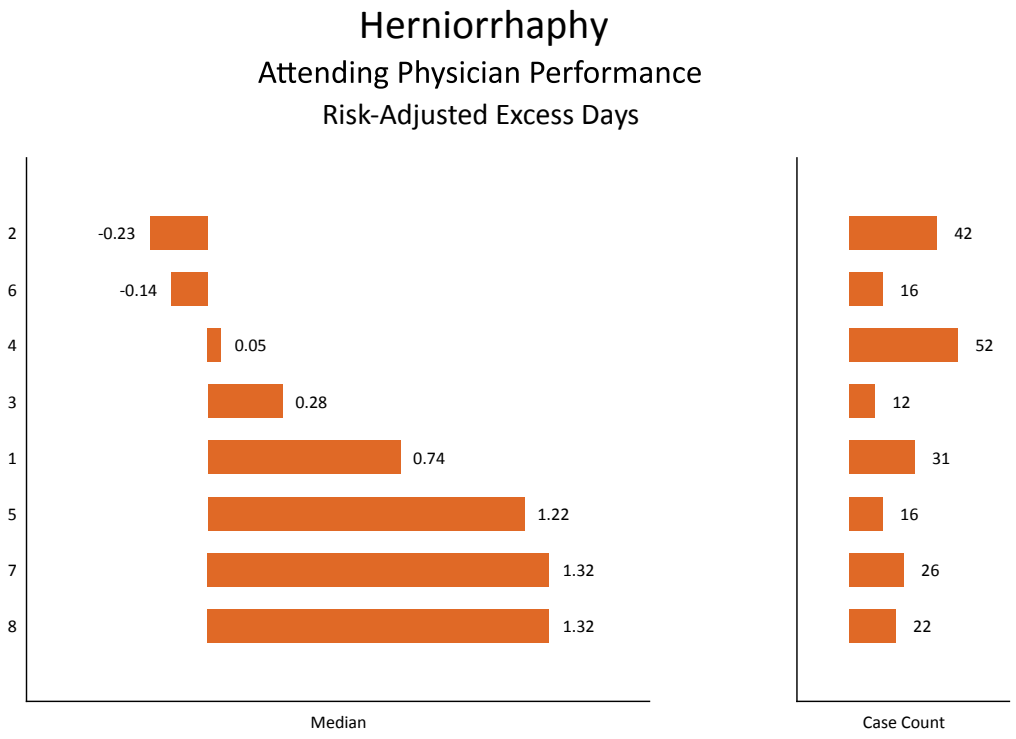
P<0.05

Graph 9



Confidence Level = 0.95

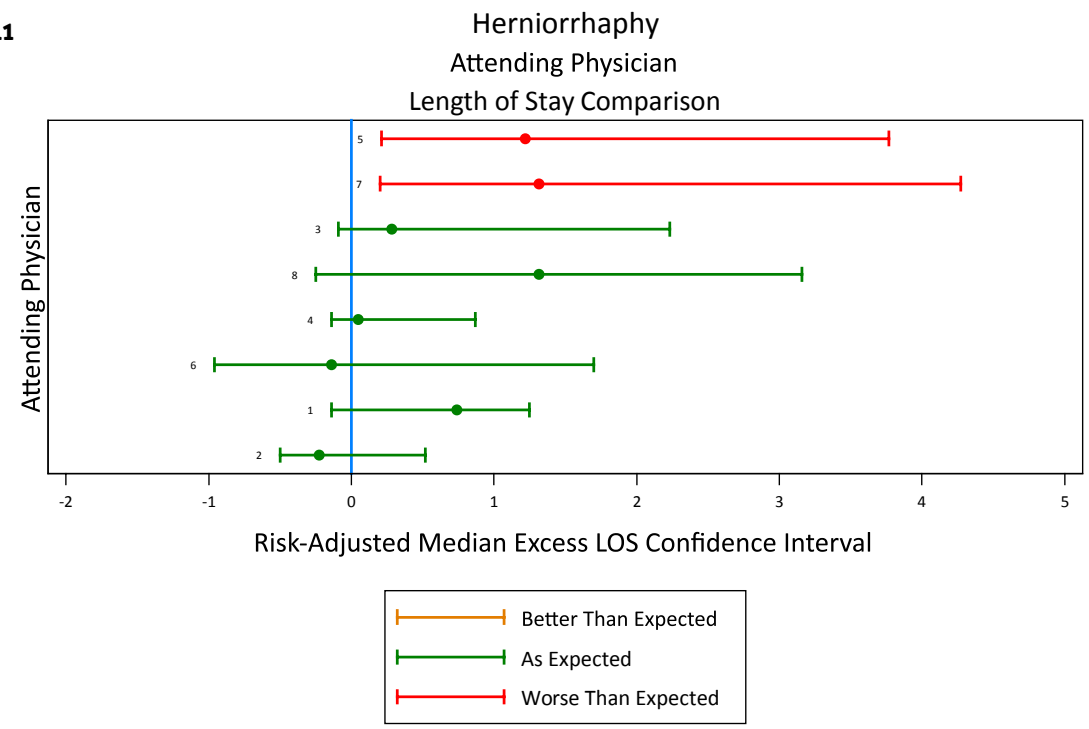
Graph 10



Physician case count at least 12

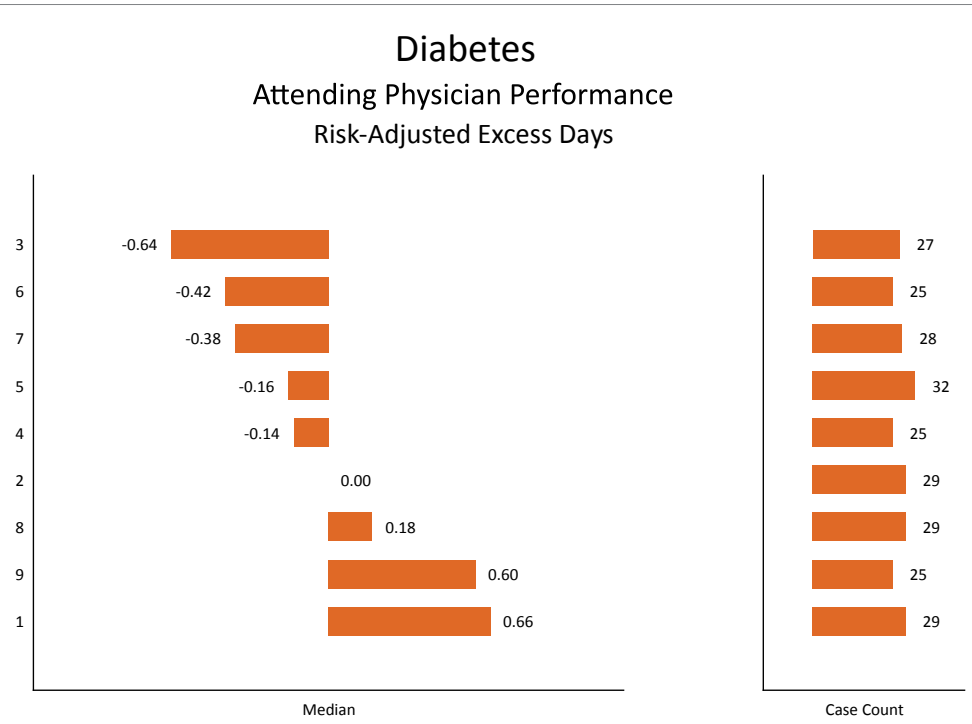
P>0.05

Graph 11



Confidence Level = 0.95

Graph 12



Physician case count at least 24

P>0.05

appropriate reference. Before we delve into examining performance using confidence intervals, let's first understand what a confidence interval is.

A confidence interval is a range of values that is likely to occur given the variability present in the data. Performance measurements are not static; they vary from one time period to the next. To accurately measure physician performance, this variability must be accounted for in the measurement system. When deriving confidence intervals, one can select the level of precision desired. Typical levels are 99 percent, 95 percent, and 90 percent with the most common being 95 percent.[1]

The following example demonstrates how best to interpret confidence intervals and derive an improvement strategy. Graph 5 depicts a 95 percent confidence interval of the median risk-adjusted excess length of stay. The left-most bar of the interval is the lower confidence limit (LCL), the right-most bar is the upper confidence limit (UCL) and the median is depicted by the dot. The appropriate reference here is the vertical blue line located at the zero value which represents performance that is as expected.

The physicians at the bottom of the graph, highlighted in orange, are performing better than expected, since their confidence intervals do not intersect the reference line, and the UCL lies to the left of the reference line.

The physicians depicted in green are performing as expected, since their confidence intervals intersect the reference line. And the physicians depicted in red are performing worse than expected, since their confidence intervals do not intersect the reference line and the LCL lies to the right of the reference line.

In this scenario, a viable improvement strategy is to examine the practice patterns of the physicians with better-than-expected performance and disseminate the findings to the other physicians. A standardized clinical protocol could also be developed based on the findings; and garnering the cooperation of the other physicians in using the protocol will likely yield meaningful improvement.

Graph 6 shows a scenario where all physicians are performing as expected. Hence, there are no physicians who can be used as role models for performance improvement purposes. The improvement strategy in this case consists of researching and implementing best practices and practice guidelines.

Physician performance improvement scenarios

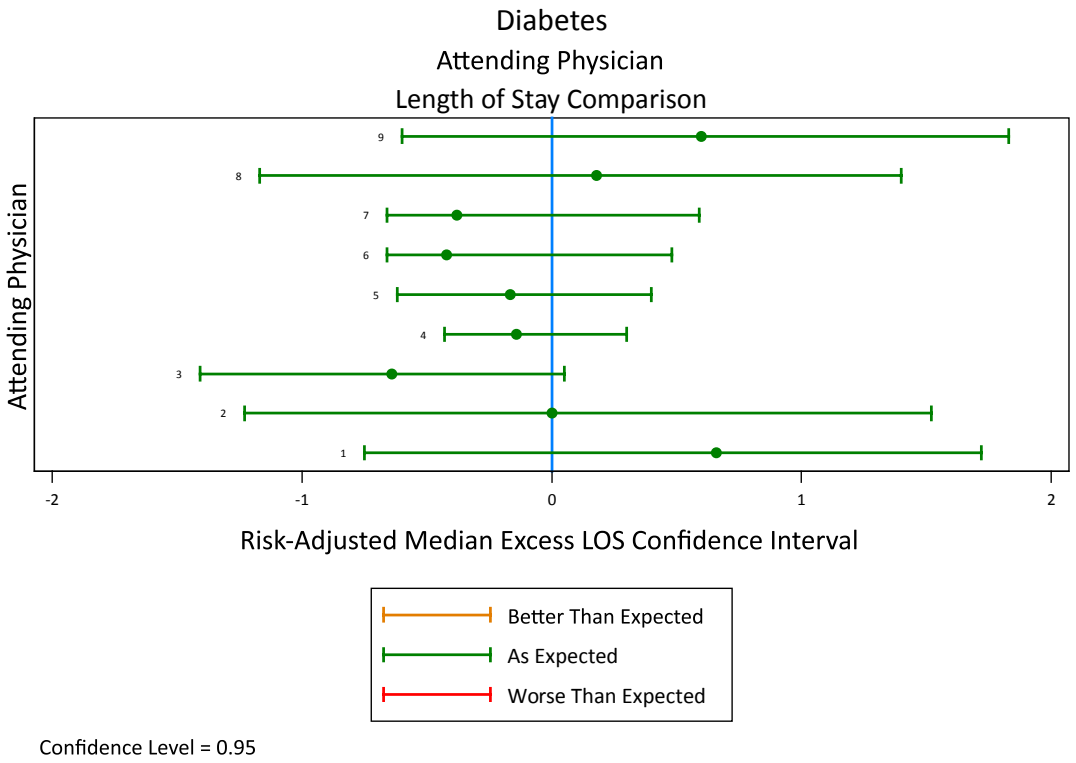
Here's a look at how to apply the concepts to the three common scenarios depicted in Graph 7. Although these are by no means a comprehensive set of scenarios that physician leaders will encounter, our hope is that they will demonstrate application of the concepts in a manner that enables physician leaders to apply them to whatever situation they encounter.

Chest pain

In this scenario physician practice patterns account for 99.2 percent of the total variability in risk-adjusted excess length of stay (defined as observed length of stay minus expected length of stay). Consequently, there is a substantial opportunity for engaging physicians to explore and standardize practice patterns in order to reduce variability in risk-adjusted excess length of stay.

Graph 8 answers the question, "Is risk-adjusted excess length of stay significantly different among physicians?" In this scenario the answer is yes, since the P-value is less than 0.05. Therefore, reducing variability in risk-adjusted excess length of stay among physicians will likely yield

Graph 13



meaningful improvements in organizational performance.

This is a common scenario where the physician case counts are relatively small. In some instances statistically significant differences will not be detected due to the small case counts. However, when statistically significant differences are detected a large difference in physician performance is suggested.

Graph 9 reveals one physician with "better-than-expected" performance. The performance improvement strategy in this scenario is two fold:

- Examine the practice pattern of physician No. 7 to determine the components that contribute to excellent performance and disseminate the findings to other physicians.
- Facilitate a dialogue regarding the findings and pursue the development and implementation of a standardized clinical protocol.

These actions will promote reduction in the variability of physician performance, migrate more physicians to the better-than-expected performance category, thereby improving organizational performance.

Herniorrhaphy

In this case physician practice patterns account for 37.2 percent of the total variability in risk-adjusted excess length of stay. This means that 62.8 percent of the variability is the result of organizational factors that are not under the physician's control.

Therefore, if the goal is to reduce overall variability, the primary focus of the performance improvement project should first entail identifying the organizational factors that are contributing to the variability. Physician practice patterns would be a secondary consideration.

Furthermore, Graph 10 reveals that there is not a statistically significant difference in physician performance since the P-value is greater than 0.05. Therefore, there is no meaningful variability in risk-adjusted excess length of stay performance among physicians—so efforts to reduce physician variability are futile.

Physician performance, however, is distributed among the performance categories as depicted in Graph 11. Since none of the physicians is performing better than expected, role models cannot be used to identify best practices. And since the two physicians that are performing worse than expected only account for 19.4 percent of the case volume, focusing attention on improving these physicians will not likely yield an improvement in organizational performance.

Consequently, if there is a desire to improve physician performance, then developing and implementing a standardized clinical protocol based on evidence-based practice is the strategy of choice.

Diabetes

This situation is a common one where physician practice patterns contribute a negligible amount to the overall variability in performance. In this scenario it is 3.9 percent.

In order to reduce the variability in risk-adjusted excess length of stay the entire focus of performance improvement should be on organizational factors—such as case management, use of evidence-based protocols, etc.— which contribute 96.1 percent of the total variability in performance.

In addition, Graph 12 underscores the lack of opportunity to reduce physician variability. Since there is no significant difference in performance—P>0.05—there is no meaningful variability to reduce. And Graph 13 reveals that all the physicians are performing as expected. So the overall improvement strategy in this scenario should be focused on improving organizational factors that adversely impact the variability in performance.

Imagine convening a group of physicians and encouraging them to reduce the diabetes risk-adjusted length-of-stay variability among them without answering these questions first:

- What proportion of total performance variability is attributable to physicians?
- Are there statistically significant differences in physician performance?
- How is physician performance distributed across the outcome categories of "better than expected," "as expected," and "worse than expected?"

You would not realize that physicians account for only 3.9 percent of the variability and that there is not any significant variation in performance to reduce. Therefore, you would be spending your time and that of your medical staff in a fruitless manner and may indeed impair your relationship with them.

Consequently, it is worth the time and effort to answer these questions before initiating a performance improvement project. You will be positioned to spend your medical staff's time in a more productive manner.

Robert Sutter, RN, MBA, MHA, is a consultant with Truven Health Analytics in Ann Arbor, MI.

bob.sutter@truvenhealth.com

Brian Waterman, MPH, is senior product manager with Truven Health Analytics in Ann Arbor, MI.

Michael Udwin, MD, FACOG is medical director with Truven Health Analytics in Ann Arbor, MI.

Reference

1. Martin JG, Altman DG: Statistics With Confidence: Confidence Intervals and Statistical Guidelines. *British Medical Journal*. 1989.