

Listening for Signals in Survey Data Is Just the Beginning

When evaluating the health of an individual, we typically consider several factors such as blood pressure, blood sugar levels, cholesterol, aerobic fitness, daily exercise, diet, and family history, and if there is a diagnosis, care is taken to isolate and manage comorbidities. We take this level of care because we recognize that the body is a set of interconnected systems, each with its own specific function contributing to the whole. Breakdowns in one area create signals of ill health. The detective work is in following that signal to a complete diagnosis of a broader issue.

The same can be said about patient experience and the diagnosis of the root cause of service breakdowns. Rarely is one indicator, or even a few, the true source of a service breakdown or a safety event. The key to improvement is piecing together the right directional indicators to triangulate the source of systematic service issues. Patients who have negative experiences are typically reacting to the system of care that an organization has created and maintains. In fact, multiple negative experiences within the same visit can have a compounding negative effect. The same can be said of service recovery. Making changes in certain areas can have a ripple effect across the whole experience and can make patients more forgiving of other issues. Without the full complement of directional variables, these insights can be overlooked.

One or two items may point to an underlying issue within the average patient's experience at a health care organization. However, isolated data points do not provide directions for action. Here is a common example.

In the medical practice setting, there is more to a patient's experience than just communication with the provider. Even though provider communication tends to correlate highly with the likelihood to recommend that provider, the other facets of the experience are also vital to a patient and will affect their likelihood to recommend.

A provider may be a great communicator; however, if patient safety protocols are not followed, a patient may become concerned about the health risk they may be exposed to and become less likely to return. If a practice's Likelihood to Recommend scores decrease, the assumption would be that this measure only reflects providers' communication skills. Providers will be retrained in communication skills when the actual reason, patient safety, is not addressed. Therefore, after the training exercise is completed, it is not likely that an increase in provider communication skills will increase Likelihood to Recommend, as the patient's real concern has not been addressed.

- Press Ganey surveys/questions are designed to go beyond the signal and provide both a diagnosis and a direction for creating a change in the care experience.
- Signals can be indicators of an underlying issue with patient care but are rarely enough to get the full picture.
- A clear picture of patient experience is a collection of moments of care that contribute to the overall care experience.
- The key to quality improvement is finding the root cause of suboptimal experiences of care and focusing improvement efforts on the specific directional components.

Reducing Survey Length Can Increase Noise

There is a long history of studying the best way to elicit sound and detailed responses. One major consideration in designing any test or measure is its length—specifically, the number of items used to score and assess a measure of interest. Generally, a longer measure is more valid and reliable than a shorter one (Anastasi and Urbina 1997).

Every item will carry two sources of variability: true score and random measurement error. True score variances are differences in scores given to an item due to actual differences in what is being measured. Random measurement error variances are differences caused by any factors that randomly affect measurement of the variable across the sample. In the case of the latter, because the error variance carried by each item is random, the aggregation of scores from multiple

items has the effect of cancelling out this error and increasing the precision in scores obtained. This is defined as the statistical reliability of that measure. This relationship between item count and reliability is defined by the Spearman-Brown formula (Spearman 1910; Brown 1910). Since statistical reliability is the specific prerequisite for validity in any type of measurement, its consideration should be paramount. If the number of items is too few, statistical reliability can collapse and render that measure invalid.

The Right Combination of Brevity and Coverage

All of Press Ganey's standard surveys are tested to ensure that they meet high scientific standards of validity and reliability. This ensures that clients can rely on the data collected with these tools and the resultant benchmark comparisons. We also test to ensure that our surveys provide a complete picture of the patient experience so that important aspects of care are not missing.

For example, the 13 standard medical practice items explain approximately 79% of the variation in patients' likelihood to recommend their care provider and approximately 74% of the variation in patients' likelihood to recommend the practice. This means that with those 13 items, you will understand nearly all the possible contributing factors to those outcomes.

Through our psychometric processes, we also find that once you drop below this number of items, the likelihood of missing a contributing factor increases. Therefore, validity and reliability begin to decrease. In the medical practice space, our general recommendation is that the overall score is the most complete measure of the average patient's experience in the typical medical practice setting. The questions chosen as standard on a given survey, as well as their order and section location, are determined through a deliberate process that involves secondary literature research, focus groups, pilot tests, factor analysis, and other statistical metrics. The questions in a given section are designed to work in concert; they have value not only as individual measures but also because they are hypothesized to operate together in identifying the patient's satisfaction with larger related concepts (registration, facility, personnel issues, etc.) that our research has identified as the strongest determinants of patient experience with a given type of health care service.

- Single items or small groups of items can be more prone to error (false positives/false negatives, inflation/deflation) than collections of items.
- Multi-item indicators collectively reduce error in any one single item.
- Just as we want to guard against making any measure too long and burdensome, we should also be cautious about having it be too short.

- Press Ganey uses robust survey science to ensure that its tools are the optimal combination of length and breadth.
- Press Ganey rigorously tests its surveys to limit survey fatigue, maximize response rate, and deliver valid and reliable results on which to base improvement efforts.

References: Anastasi, Anne, and Susana Urbina. 1997. *Psychological Testing*. Upper Saddle River, N.J.: Prentice Hall. Brown, William. 1910. "Some experimental results in the correlation of mental abilities." *British Journal of Psychology* 3, 296-322. Spearman, Charles C. 1910. "Correlation calculated from faulty data." *British Journal of Psychology* 3, 271-295.