The Satisfaction Snapshot is a monthly electronic bulletin freely available to all those involved or interested in improving the patient/client experience. Each month the Snapshot showcases issues and ideas which relate to improving patient satisfaction and customer service, improving workplace culture and improving the way we go about our work in the healthcare industry.

The Satisfaction Snapshot features:
- relevant articles from healthcare industry experts
- case study success stories
- tips and tools for quality improvement
- patient satisfaction and other industry research findings
- articles with ideas to help achieve success in your role

If you would like your colleagues to receive the Satisfaction Snapshot please send us their names and email addresses. The Satisfaction Snapshot is published by Press Ganey Associates Pty Ltd. All material is copyright protected. Quotation is permitted with attribution. Subscribers are permitted and encouraged to distribute copies within their organisations.

Subscription to the Satisfaction Snapshot is FREE!

Please direct any comments, suggestions or article submissions to:

Manager of Client Relations
snapshot@pressganey.com.au
www.pressganey.com.au
P: 07 5560 7400
F: 07 5560 7490

No one in health care would base medical decisions on a diagnostic test without evidence of its accuracy or precision, yet many are willing to be held accountable to and assume that a satisfaction survey that looks good and seems to make sense is accurate. Many claims of survey validity are made by the organisations enforcing these surveys, even for well known survey instruments, yet evidence of correct psychometric testing is often never provided.

Patient surveys used as diagnostic measurement tools must be statistically reliable and valid. Reliability refers to the consistency or reproducibility of a measure or the degree to which survey results are free from random error. The more reliable an instrument, the better it reflects a patient’s true opinions and distinguishes among patients with different levels of experience and satisfaction. Validity refers to the extent to which a survey instrument measures what it claims to measure, or the degree to which survey results are free from both random error and systematic bias.

Changing care processes, developing interventions, or creating “report cards” based on questionnaires with untested reliability or validity is a risky venture.

Patient reports are not direct measures of performance and have not been validated using techniques any more sophisticated than focus groups and cognitive interviews.

If a research instrument has not undergone a robust process of development and testing, the credibility of the research findings themselves may legitimately be called into question and may even be completely disregarded.

So what evidence should be provided?

The October 2008 Satisfaction Snapshot provides the guide for asking the correct questions. The article, written by Dr Raymond G. Carey, and published by JCAHO, (the Australian equivalent of ACHS), provides the rationale behind the correct tests that should be undertaken.

If you want to be confident about the survey instrument you use, ask for the evidence that it has been tested against this criteria. If so, you can be confident that the feedback from your patients is accurate and any follow up actions will have the desired impact.
How to Choose a Patient Survey System

Raymond G. Carey, PhD
PATIENT’S PERSPECTIVE

How to Choose a Patient Survey System

Raymond G. Carey, PhD

Patients surveys have evolved through a number of stages. Historically, the typical patient survey was designed by and for a hospital marketing department. Early examples in the scientific literature tended to focus on general perceptions and attitudes towards health care. Later, surveys began to focus on patients’ experiences in specific health care encounters. Recently, much has been made of surveying patient “perceptions” rather than patient “satisfaction”. A defining feature of perception surveys is the inclusion of both evaluative ratings and patient reports (that is, the patient reporting that something either did or did not happen).

Let us say that you are the administrator of a hospital or medical group practice and are convinced of the value of patient perceptions of your health care delivery process. Furthermore, you want to use patient feedback for more than merely to gathering “report card” data to satisfy the demands of purchasers' or accrediting agencies' requirements. You also want to identify opportunities for improvement, to measure the effect of efforts to improve patient satisfaction, and perhaps to use patient satisfaction as a financial incentive for your managers. A single generic survey will not meet all the manager's needs for patient satisfaction data. You will require a set of surveys uniquely designed for inpatients, for outpatients, for emergency room patients, for patients who come for therapy or testing, and for patients making physician office visits. To meet all these needs, you need a high quality survey system embracing an entire family of surveys.

You consider developing your own survey system but realize that it would be too expensive to hire a staff of survey experts to develop customized questionnaires from scratch that would satisfy all your associates. Even if you did develop your own questionnaires, you would lack external benchmark data to evaluate your results in comparison to other providers. Therefore, you decide to choose an outside survey firm that has scientifically sound questionnaires, a national database, and understandable report formats.

But how to choose from among the dozens of survey firms on the market? The hospital trade association and the accrediting agency each provide you with the name of its favorite firm, but without any scientific reasons to support its recommendation. This made you suspect that their recommendations might have been based on personal, political, or pricing considerations, rather than scientific merit. Therefore, you obtain a list of several of the firms with a national presence and are now looking for a set of criteria to help you choose the best one for your needs.

This article is intended to provide you with some basic scientific criteria. Although references and pricing are important factors in the final selection of a
Article-at-a-Glance

Background: An administrator of a hospital or medical group practice who is convinced of the value of patient perceptions of its health care delivery process needs a high-quality survey system embracing an entire family of surveys. This article is intended to provide some basic scientific criteria in choosing the appropriate survey system.

A psychometrically sound questionnaire: A questionnaire’s degree of accuracy needs to be assessed by measuring its reliability and validity.

Avoiding bias in data collection: Interviewer bias, which is often associated with personal interviews and telephone surveys, tends to inflate patient satisfaction scores and, as a result, obscure opportunities for improvement. Nonresponse bias, which is more typical of mailed surveys, occurs when there is a systematic difference between those who respond and those who do not respond.

An enumerative or analytic study? The purpose of the survey determines the sampling method, the survey frequency, and the appropriate report format. An enumerative study is conducted on a static population for a given period and/or location and is designed merely to describe outcomes. An analytic study examines a process over time and seeks to determine why the outcomes were observed and/or whether planned improvements had an impact.

Summary and Conclusion: If administrators intend to hold managers accountable for the results, they should retain a high-quality survey firm to support and direct their efforts to collect the data and interpret the results.

A Psychometrically Sound Questionnaire

Because no patient questionnaire will ever assess patient satisfaction with perfect accuracy, we will never know the true level of patient satisfaction. Therefore, we need to assess the degree of accuracy of a questionnaire by measuring its reliability and validity. If a survey firm cannot offer any documentation to back up its claims of reliability and validity for its scale scores (for example, physician or nursing care scales) - preferably an article in a peer-reviewed journal - this alone is reason not to use the firm.

Reliability is defined as the extent to which measurements are free from random-error variance. A questionnaire with high reliability allows us to be confident that observed scores on a questionnaire accurately reflect a respondent’s true opinions. There are three general forms of reliability: test-retest reliability, equivalent form reliability, and internal consistency. The reliability required for patient surveys is internal consistency among the questions dealing with a particular dimension of patient experience, such as, nursing care or physician care.

To assure good reliability in the sense of internal consistency, the patient survey should have multiple-item scales, not merely stand-alone items, because you cannot estimate the internal consistency of stand-alone items or questions. For example, to assess patient opinion on a particular dimension (for example, nursing) one might use a single question (“How would you rate nursing overall?”) in an effort to keep the questionnaire as brief as possible. Or one might ask several questions regarding various aspects of nursing care; for example, how well nurses answered questions, how promptly they responded to the call button, and so on. The use of two or more questions to describe a dimension of care is referred to as a scale. Combining the ratings of individual questions that make up a scale is called a scale score.

Because internal consistency can be estimated only with scales that have more than one item, people who use one-item measures to assess the level of satisfaction of nursing care, physician care, and so on run the risk of obtaining information that is not highly reliable. That is, one cannot estimate the extent to which the observed score on the one-item measure is related to the true level of patient satisfaction.

Many questionnaires, even those with a national name recognition, suffer from this important failing. Indeed, it would seem that the Joint Commission on Accreditation of Healthcare Organizations (Oakbrook terrace, IL) for example, would be making this mistake if it were to accept only one - item patient satisfaction
Why is a scale with high reliability critical to measuring for improvement? Because scales with high reliability enable you to better distinguish between respondents on a continuum of patient satisfaction. If you cannot accurately detect differences in satisfaction levels, then a successful continuous quality improvement intervention may appear not to have improved the process. One might erroneously conclude that an intervention was a failure, when it actually resulted in an improvement. If a manager’s bonus depended on improvement, he or she will be denied a bonus he or she truly deserved. An unreliable measure used as an indicator of an organization's performance could lead to unnecessary inquiries or investigation by an accrediting agency, such as the Joint Commission.

A reliability coefficient applies to an individual patient’s score. There are several methods to estimate the degree of reliability, which is generally reported as a coefficient between zero (no reliability) and one (perfect reliability). The most common method is Cronbach’s alpha coefficient\(^9\). There is no rigid rule to determine an acceptable level of reliability. A criterion offered by Helmstader is whether a measure is intended to compare groups or to compare individuals. For measures of individuals, such as personality tests, .90 is an appropriate minimum. For scales that compare groups, not individuals, a minimum of .50 is generally considered acceptable\(^10\). As explained by Ghiselli et al, if “the reliability coefficient of ratings assigned by a single rater is .50, and we wish to have ratings with a reliability coefficient of .90... we need to employ [at least] nine such raters.”\(^11\) The typical patient survey sample is considerably larger.\(^11\)

Validity, on the other hand, refers to the degree to which a scale measures what it is designed to measure. Unlike indices of reliability, such as the Cronbach alpha coefficient, no one statistic provides an overall index of validity of scale scores. For example, if you wish to determine whether the scales you developed to measure distinct and separate aspects of care (for example, accessibility and quality) are viewed in the same way by respondents, then you are asking whether the questionnaire has construct validity. One way to evaluate construct validity is to conduct a factor analysis. Factor analysis examines the relationships between a large set of variables and tries to explain these correlations using a smaller number of variables. Generally, in a set of items, those that are highly correlated with each other are represented by a single factor or dimension. Each variable (survey item) can be assigned to a factor based on the strength of “factor loadings”. Normally, factor loadings above 0.40 are used to discriminate which variables belong to a factor or dimension of care. This approach is far superior to using judgment (or face validity) to group items into scales.

Although factor analysis is probably the most common way to establish the construct validity of scales, other methods\(^11\) are available. One such method is the demonstration of convergent and discriminant validity. Surveyors can establish evidence of such validity by correlating scale scores designed to measure the same dimension and then correlating scale scores designed to measure different dimensions.

A valid scale should correlate with variables with which it is expected to correlate but not with scales with which it should not correlate. Another approach to obtaining construct-related evidence is through the use of multitrait-multimethod, which requires the measurement of multiple constructs (or traits) through multiple measurement devices (or methods). In summary, if the scales or dimensions that a survey purports to contain are supported by neither factor analysis nor an alternative method, the scales are probably unreliable and invalid.

---

\(^*\) Mathematically, the effect of measurement at the group level can be seen in the formula where \(r_{cc}\) is the reliability coefficient of the measure, \(r_{cc}\) is the reliability coefficient of the magnitude desired and \(k\) is the number of raters.

\[
k = \frac{r_{cc}(1-r_{cc})}{r_{cc}(1-r_{cc})}
\]
Criterion-related validity is also important. Multiple regression studies showing the extent to which scale scores predict (or explain the variance in) perceived quality of care and patient loyalty are good evidence of this type of validity. At a minimum, the survey firm should provide the multiple R-square from a regression model. This represents the percentage of variance in the dependent variable (perceived quality, loyalty, and so on.) explained by the scales. A good survey will account for 50% or more of the variance in global evaluations of “overall quality of care” or in patients’ expressed willingness to return to, or to recommend, a facility. Simple correlations should not be accepted as a substitute for regression models.

A lengthy discussion of reliability and validity is beyond the scope of this article. The main idea is that the survey firm you choose should provide you with a list of Cronbach Alpha coefficient values and factor loadings for a survey's scales so that you can examine the documentation for the firm's claims of reliability and construct validity. Then you can decide for yourself whether the survey is psychometrically sound. Readers seeking a detailed illustration of how reliability and validity ought to be developed for patient surveys are referred to an article by Carey and Seibert5 or to the book, Measuring Customer Satisfaction, by Bob E. Hayes7.

Avoiding Bias in Data Collection

Two primary types of bias can affect survey data: - interviewer bias and nonresponse bias. Interviewer bias is often associated with personal interviews and telephone surveys. For example, a patient who is interviewed by a nurse before being discharged may be reluctant to respond negatively for fear of being considered rude or confrontational. Alternatively, discharged patients contacted by telephone at home may be reluctant to answer questions about physician care for fear that their comments will be shared with their personal physician and possibly damage their relationship with the physician. Interviewer bias tends to inflate patient satisfaction scores and, as a result, obscure opportunities for improvement.

Nonresponse bias which is more typical of mailed surveys occurs when there is a systematic difference between those who respond and those who do not respond. There is no consensus on the percentage required for a representative response rate. Some researchers, such as Babbie12 argue that a 50 percent or better response rate is necessary to generalize from your data. Barkley and Furse13 came to the same conclusion based on a study of 19,556 patients from 76 hospitals. They found that the results of individual scale scores and subsequent improvement priorities had a 50-50 chance of being different when the first 30% of responses were compared with all respondents.

More importantly, there was no way for an individual hospital to know whether its satisfaction scores would be higher or lower with additional respondents, that is, a low response rate is as likely to result in overstated patient satisfaction as in understated patient satisfaction!

Should you choose a firm that offers to collect your data by phone, by mail, or suggests that you distribute your questionnaires directly to patients? It depends. If you put a high priority on quick turn-around, or if you are conducting a special-purpose assessment (for example, of recipients of a new service) and want to ask open-ended questions that allow your surveyor to obtain fuller explanations from patients, then use the phone method. However, if expense is a high priority, mailing surveys is the recommended method of data collection.

Mailed can obtain a 50 percent response rate or better with less than half the expense of a phone survey, and without the interviewer bias typical of telephone surveys. To achieve this response rate, however, you should have a brief questionnaire (15 minutes or less to complete) with an aesthetically appealing format, a motivating cover letter from the chief executive officer, a prepaid postage return envelope, and at least one effort to follow-up on the original mailing. In addition, the questionnaire and cover letter should be mailed within a short time (one to two weeks) of the patient encounter.

(Personally distributing questionnaire to patients either in the hospital, outpatient setting, or physician offices is the least expensive method of data collection, but results in highly inflated responses that are virtually useless for identifying improvement opportunities). In any case, one method should be chosen for all the data - you do not want to mix your data collection methods and then combine the data into one database.

Although achieving a high response rate is sometimes used as an argument for conducting telephone surveys instead of the much less expensive mailed surveys, telephone surveyors have been known to inflate their reported response rate by reporting what is called the “cooperation rate” instead of the actual response rate.
For example, if you mail questionnaires to a sample of 300 patients and obtain 180 returned questionnaires, your response rate is 60 percent. If you try to contact the same 300 people by phone and are able to actually contact only 200 patients of which 180 agree to be interviewed, then your response rate is still 60 percent, not 90 percent (180/200). Ninety percent is really a “cooperation rate”, not a response rate, because you must consider all 300 patients in the denominator when computing your response rate.

**An Enumerative or Analytic Study?**

The purpose of your survey - enumerative or analytic - will determine three aspects of your survey: the sampling method, the frequency of the surveys, and the report format. An *enumerative study* is done on a static population for a given period and/or location to merely describe outcomes. For example, if you are only trying to identify and prioritize opportunities for improvement, you are conducting an enumerative study - the most common use of patient surveys. An *analytic study* examines a process over time, and seeks to determine why the outcomes were observed and/or whether planned improvements had any impact. It requires ongoing surveying, not just an annual or a biannual survey.

An enumerative study is like a snapshot; an analytic study is like a video. Depending on whether you are using your patient survey as an enumerative or analytic study, a survey firm should be able to guide you in the choice of sampling and the frequency of surveys, and provide the appropriate report format for both purposes.

Sampling for an enumerative study will usually be stratified random sampling driven by your plans for analysis. For example, if you want to analyze inpatient data by nursing unit, emergency room patients by shift, or physician office visits by site - then you must sample by unit, shift, site and so on. The number in your sample will be determined by the degree of confidence you want to have in the survey results. Most hospitals and other organizations will use this method of sampling, rather than a simple random sample drawn from all patients.

However, for certain types of analytic studies (namely those using X Bar and R or X Bar and S control charts), you will want to use what Deming called “*rational*” or “*judgment*” sampling. Judgment sampling entails selecting a series of subgroups from a population based on the opinion of those who have expert knowledge of a process being studied.

Judgment samples are drawn on a continuous basis, while proportionate random samples are usually drawn at fixed points in time. Judgment samples are less costly than other approaches, require less data, and enable an improvement team to take action on a dynamic process as it unfolds over time. However, the results of a single judgment sample cannot be generalized to the population from which the sample was drawn. It is unlikely that a survey firm could draw judgment samples for you.

The survey's frequency will also be determined by its purpose. Quarterly analysis is usually sufficient to identify and prioritize opportunities for improvement. An analytic study is much more labor intensive and requires analysis of daily or weekly data points.

Report formats for enumerative studies will often be characterized by bar charts, comparing your data to an appropriate database. The bar charts should show where your patients fit in a relation to those in other facilities. Is your facility consistently above or below the norm (average) of other facilities? What level of confidence do you have in the findings? Is your facility one or two standard deviations above or below the norm? The bar charts should clearly identify whether your findings represent an opportunity for improvement or an area of excellence.

However, if you are using your survey as an analytic study, the results should be displayed in run or control charts. These charts will tell you whether the variation in a process in which you are interested is common-cause variation (a stable process) or special-cause variation (an unstable process). For example, if you have introduced an improvement plan or a new clinical pathway, the control chart will tell you whether the process has improved, deteriorated, or remained the same after the intervention.

The choice of sampling and report formats is complicated, and therefore it is critical that the survey firm you choose is prepared to offer you guidance on these issues. A more extended discussion of enumerative and analytic studies, and the associated sampling and report formats can be found in *Measuring Quality Improvement in Healthcare* by Carey and Lloyd.\(^{15}\)

**Comparative Reference Data**

Comparative reference data (or benchmark data) help place your survey results in perspective. A number that is hanging out in thin air, (for example 66 percent of patients said they were able to get an appointment in...
a week) is uninterpretable. Is 66 percent good, satisfactory, or unsatisfactory? You have to have some basis for comparison to answer the question. Using your own internal benchmark data is often useful. For example, you can compare your own nursing units, emergency room shifts, physician groups or sites - or look at the results of any one of these entities over time. It is relatively easy to obtain high quality internal benchmark data because you can ensure that everyone is using the same instrument and collecting data in the same way and in the same time intervals.

External benchmark data are more difficult to obtain. To make appropriate comparisons, data from other hospitals or facilities must be obtained by using the same valid and reliable instrument, from similar patients or similar hospitals or facilities, with a representative response rate. If any of these requirements are not met, the external benchmark data will be questionable. It is not just the number of facilities represented in your vendor’s database that is important, but also the quality of the data. Do all sites use the same methodology? Are sites that use variant methods or get low returns excluded from the database?

If Nothing Meets Your Needs

What if none of the surveys offered meet your special needs? Either you feel that certain items critical to your situation are missing, or you want to change key words in questions, for example, physician to clinician. It is important to realize that if you add or remove any items, modify the wording of items, or change the order in which they appear in the survey, then the reliability and validity of the scales will be affected to some unknown degree. If you choose to make any of these modifications, then you must also be willing to take the time and expense to do a pilot study with your revised questionnaire and obtain new estimates of reliability and validity. A high-quality survey firm will be sufficiently flexible to adapt its standard survey to meet your specific needs.

Summary and Conclusion

Administrators should not survey patients unless they plan to use the results. If they intend to hold managers accountable for the results, then the administrators should retain a high-quality survey firm to support and direct their efforts to collect the data and interpret the results. A high-quality firm will provide documentation of the reliability and validity of its questionnaires, as well as hot-line professional assistance in sampling techniques, data collection procedures, and the interpretation of its report findings.

Managers tend to see what they want to see. When survey findings even appear to be negative, managers tend to be defensive and look for reasons why the survey is invalid. At other times they are elated over what appear to be “improving trends” that in reality are random data variation due to unreliable questionnaires, poor sampling, or an unrepresentative response rate. Obtaining valid and reliable patient survey feedback is a complex and difficult undertaking. You will avoid chaos by taking time to choose your survey firm carefully.

References