

# Quantifying Patient Benefit-Risk Tradeoff Preferences: A Brief Introduction

**F. Reed Johnson, PhD**

Senior Fellow and Principal Economist  
RTI Health Solutions  
PO Box 12194  
Research Triangle Park, NC 27709-2194  
+1.919.541.5958  
frjohnson@rti.org

**A. Brett Hauber, PhD**

Senior Economist and Head  
Health Preference Assessment  
RTI Health Solutions  
PO Box 12194  
Research Triangle Park, NC 27709-2194  
+1.215.345.6519  
abhauber@rti.org



# Quantifying Patient Benefit-Risk Tradeoff Preferences: A Brief Introduction

**F. Reed Johnson, PhD**

Senior Fellow and Principal Economist  
RTI Health Solutions

**A. Brett Hauber, PhD**

Senior Economist and Head  
Health Preference Assessment  
RTI Health Solutions

## Introduction

Several recent and well-publicized events involving withdrawals of drugs from the U.S. market have dramatized the problem of balancing benefits and risks. In all these cases, interventions offering potentially significant therapeutic benefits were found to carry risks of serious, possibly life-threatening adverse events. A decision regarding the need to halt the development or marketing of such therapies clearly requires an evaluation of the balance between benefits and risks given the available evidence at a point in time. The Institute of Medicine (IOM) noted recently that “in both the pre-approval and the post-marketing setting, the risk-benefit analysis that currently goes into FDA decisions appears to be ad hoc, informal, and qualitative” (IOM 2007). Partly in response to such criticisms, the U.S. Congress included a provision in the recent Food and Drug Administration Amendments Act of 2007 requiring FDA to develop and implement a plan for evaluating and developing better strategies for communicating the benefits and risks of new pharmaceuticals to patients and physicians (H.R. 3580 [Public Law 110-85] §904).

Benefit-risk evaluations sometimes are informed by advisory bodies of scientists and clinicians, although balancing benefits and risks involves social judgments for which clinical scientists have no special expertise. Occasionally decisions are influenced by patients, the ultimate stakeholders in the drug review and approval process. The values and risk tolerance of patients are presented to advisory panels and policy makers either individually or through advocacy organizations; however, such anecdotal testimony does not provide systematic evidence of the willingness of well-informed patients to accept observed or expected risks to achieve the therapeutic benefits of these products. Furthermore, it is unclear whether those who advocate for less restrictive or more restrictive access to medications are representative of the population for whom the medication is indicated.

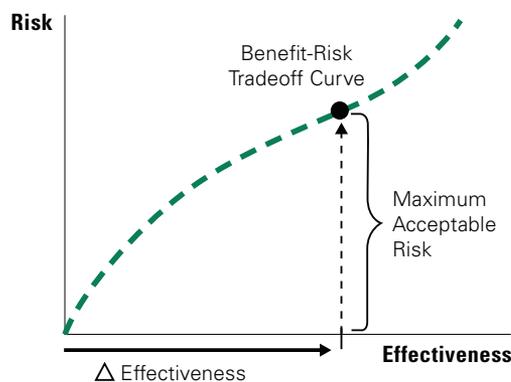
Understanding how patients perceive the benefits of treatment features and their tolerance for possible risks requires a valid and reliable measurement method. Stated-choice (SC) methods, sometimes called discrete-

choice experiments or conjoint-analysis, are the most valid and reliable techniques available for quantifying patient preferences. Two such studies were recently submitted to FDA advisory committees (Johnson et al. 2007a; Johnson et al. 2008). One submission was for re-approval of a product that was withdrawn from the market because of serious safety problems, and the other submission was for approval of the same product for a different indication. Both submissions were successful.

SC methods recognize that products have value because of their characteristics or attributes. People have preferences for each attribute and are willing to accept tradeoffs among different attributes. SC analysis examines these tradeoffs to assess the weight people assign to various treatment attributes. Analysts have used SC to quantify preferences for a variety of market and nonmarket goods and services. These include medical interventions, pharmaceutical treatments, and environmental health risks (Bryan et al. 1998; Johnson et al. 2000; Johnson et al. 1998; Johnson and Desvousges 1997; Ryan and Hughes 1997; Viscusi, Magat, and Huber 1991; Wittink and Cattin 1989).

An important advantage of SC data that incorporate benefit-risk tradeoffs is that they can quantify patients’ tolerance for risks relative to given levels of benefit. Such data identify the benefit-risk trade-off curve as illustrated in Figure 1. For a specified improvement in treatment efficacy,  $\Delta$  efficacy, maximum acceptable risk (MAR)

**Figure 1. Benefit-Risk Tradeoff Curve and Maximum Acceptable Risk**



is the highest level of risk patients would tolerate in return for the benefits offered by the treatment. If the actual risk exposure is less than MAR, then patients experience a positive net benefit from having access to the treatment.

Quantitative estimates of MAR can provide useful information for several areas of product development and marketing.

- *Development strategies for new pharmaceuticals.* If the SC study is undertaken early in the development process, researchers can identify the relative importance that consumers place on product benefits relative to possible risks. The ability to compare benefits and risks directly can lead to better-informed product-development decisions in cases where early data indicate the possibility of an adverse event.
- *Regulatory approval.* FDA is committed to better quantification of benefits and risks and transparent methods for comparing benefits and risks. FDA regulators have expressed interest in quantifying systematically the anecdotal evidence patients and patient advocacy groups currently provide for product evaluations.
- *Risk management.* Risk management professionals often must weigh the potential risks of medical interventions to a small number of patients against the potential benefits of the same interventions to a large number of patients. In addition, interventions designed to minimize the risk to patients require understanding behavior that can lead to adverse events. Prescription drug off-label use and non-adherence can occur when there are systematic differences between physicians' and patients' risk tolerance, and regulators' explicit or implicit judgments regarding acceptable risks. SC methods can quantify such differences among various stakeholder groups and help inform effective risk-management programs.

## Requirements for a Valid Benefit-Risk SC Study

Implementing a valid and reliable benefit-risk SC study requires accurate treatment definition (attributes and levels),<sup>1</sup> attention to format selection (ratings, rankings, or choice), efficient experimental design, and careful statistical analysis. The remainder of this document considers each of these study requirements.<sup>2</sup>

### Treatment Definitions

Measuring stated preferences for medical interventions requires a systematic framework to characterize relevant treatments. Demand for such interventions arises directly from preferences for treatment attributes and indirectly from preferences for the health states realized by their use. Thus, attributes and levels must incorporate the most important health outcomes and treatment attributes associated with medical treatments.

While many pharmaceuticals and medical devices have demonstrated clinical value in alleviating symptoms of disease, such benefits often are accompanied by risks of adverse events. These undesirable outcomes can range from mild symptoms such as transitory drowsiness to potentially life-threatening conditions. Tolerance for adverse-event risks may vary between and among patients and physicians. Thus including such risks is important in situations where therapeutic decisions require determining what level of risk is acceptable for a particular patient or group of patients.

Once identified, outcomes associated with treatments must be defined in sufficient detail such that subjects can distinguish among them. In addition, these outcome definitions must be consistent with the ways that people think about their health. For instance, people often do not think of their health in terms of clinical outcome measures. Rather, they may consider how the severity of symptoms associated with clinical outcomes limits or

<sup>1</sup> An attribute is a qualitative characteristic of the treatment, while a level is one of several values the attribute may have. Color and price are attributes. Blue and \$25 are levels.

<sup>2</sup> A discussion of statistical analysis is beyond the scope of this brief paper.

affects physical, social, and emotional functioning. It is the task of survey developers to determine how subjects think about health outcomes for the intervention of interest and to identify salient attributes and levels. Focus groups and survey pre-testing are a vital part of this process. Table 1 illustrates a possible list of outcome attributes and levels for treatments for vasomotor symptoms.

### Alternative Stated-Choice Formats

Once attributes and levels are determined, they can be combined into treatment profiles that describe particular real or hypothetical treatments. Subjects then evaluate these profiles in a series of SC tasks. Choosing an SC task format is an important step in developing an SC survey. Rating and discrete-choice formats have been used in SC surveys. In a ratings approach, the subjects

are presented with two treatment profiles and asked to indicate how strongly they prefer one to the other. Viscusi, Magat, and Huber (1991) used this approach to measure the value of avoiding an increase in the risk of contracting chronic bronchitis. Alternatively, discrete choice provides subjects with several different treatments simultaneously and simply asks them to identify the most-preferred option in each choice set. Ryan and Hughes (1997) used the discrete-choice format to value women's preferences for miscarriage management. Johnson and colleagues (2000) used both rating and discrete-choice formats to value avoiding cardiovascular and respiratory symptoms.

Although the various SC task formats appear similar, studies have shown that subjects often use different simplifying decision rules for different formats. These decision strategies can produce somewhat different results for different SC formats. Therefore, the SC elicitation method should be context-specific, and study objectives should play a role in format selection (Huber, 1997). Figure 2 illustrates a rating task based on the vasomotor symptom-treatment attributes in Table 1.

**Table 1. Treatment Attributes and Levels for Vasomotor Symptoms**

Treatment Feature	Levels
Severity of Daytime Hot Flashes	<ul style="list-style-type: none"> <li>No daytime hot flashes</li> <li>Mild: a fleeting warm sensation with no sweating that does not disrupt normal daily activity</li> <li>Moderate: a warm sensation with sweating that does not disrupt normal daily activity</li> <li>Severe: a hot sensation with sweating that can disrupt normal daily activity</li> </ul>
Frequency of Daytime Hot Flashes	<ul style="list-style-type: none"> <li>None (0 times) during the daytime</li> <li>1–2 times during the daytime</li> <li>3–6 times during the daytime</li> <li>More than 6 times during the daytime</li> </ul>
Frequency of Night Sweats	<ul style="list-style-type: none"> <li>None (0 times) per night</li> <li>1–3 times per night</li> <li>4 or more times per night</li> </ul>
Duration of Hot Flashes and Night Sweats	<ul style="list-style-type: none"> <li>1 year</li> <li>2 years</li> <li>4 years</li> <li>7 or more years</li> </ul>
Risk of Heart Attack Within 10 Years	<ul style="list-style-type: none"> <li>38/1,000</li> <li>50/1,000</li> <li>65/1,000</li> </ul>

### Experimental Design

Full-factorial experiments generate data based on all possible combinations of attribute levels. Such designs typically are impractical for SC surveys because subjects' cognitive and time limitations do not allow consideration of a large number of profiles. For example, a full factorial design of the treatment attributes in Table 1 contains 5 attributes, 3 with 4 levels and 2 with 3 levels, resulting in 576 ( $4^3 \times 3^2$ ) possible health-outcome profiles. In addition, subjects do not rate these options individually. Rather, subjects compare two or more options at a time. Considered in pairs, the number of possible pairwise combinations is 165,600, which obviously is impossible to evaluate.

Most current SC applications use an algorithm to reduce the number of comparisons to the smallest number necessary to efficiently quantify trade-off preferences (Dey 1985; Huber and Zwerina 1996; Kuhfeld, Tobias, and Garratt 1994). Such efficient designs can be produced using an iterative computer algorithm (Zwerina, Huber, and Kuhfeld 1997).

**Figure 2. Example of a Ranking Task for Vasomotor-Symptom Treatments**

Considering the different results and risks associated with Treatments A and B,  
which would you prefer if these were the only options available?

	Results of Treatment A	Results of Treatment B
<b>Intensity of daytime hot flashes</b>	Mild	Severe
<b>Frequency of daytime hot flashes</b>	1–2 times a day	More than 6 times a day
<b>Frequency of night sweats</b>	None	1–3 times a night
<b>Duration of hot flashes and night sweats</b>	7 years	1 year
<b>Risk of heart attack within 10 years</b>	38/1,000 (3.8%)	65/1,000 (6.5%)

<b>Check the box that best describes your opinion</b>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	<b>A is much better</b>	<b>A is somewhat better</b>	<b>A and B are about the same</b>	<b>B is somewhat better</b>	<b>B is much better</b>

### A Numerical Example<sup>3</sup>

The SC experimental design determines a sequence of profile evaluations for each subject. Responses to these choices form the data necessary for quantifying the relative importance of each attribute level. Our earlier example considered a set of 5 attributes that described vasomotor-symptom treatment profiles. For simplicity, consider 4 one-year symptom profiles and 1 side-effect risk:

- **Symptom Severity**

- Severe: Severe symptoms with more than 6 hot flushes during the day and 4 or more night sweats per night
- Moderate: Moderate symptoms with 3–6 hot flushes per day and 1–3 night sweats
- Mild: Mild symptoms with 1–2 hot flushes a day and 1–3 night sweats
- None: No hot flushes or night sweats

- **10-year Heart Attack Risk**

- 38/1,000
- 65/1,000

<sup>3</sup> This example was adapted from Johnson et al. 2007b.

In this case, the following 8 (4 × 2) possibilities describe all the possible profiles. The 8 alternatives in Table 2 represent the full-factorial design. In a rating format or two-alternative choice format, there are 28 possible profile pairs. These could be divided into four blocks or survey versions so that each subject would evaluate 7 pairs.

**Table 2. Menopause Treatment Profiles**

Profile	Symptom Severity	10-Year Heart-Attack Risk
1	Severe	3.8%
2	Severe	6.5%
3	Moderate	3.8%
4	Moderate	6.5%
5	Mild	3.8%
6	Mild	6.5%
7	None	3.8%
8	None	6.5%

We can quantify the relative importance of each symptom profile and risk level from the answers to the tradeoff questions using ordered probit or logit for rating questions and conditional logit for choice questions. Ordered-probit analysis of the pairwise rating data for vasomotor symptoms produces the following estimates:

$$U = 0.56 \cdot \text{None} + 0.41 \cdot \text{Mild} + 0.37 \cdot \text{Moderate} + 0.55 \cdot \text{Risk}_{3,8}$$

where  $U$  is the implicit SC utility for each profile.<sup>4</sup> None, Mild, and Moderate are 0/1 dummy variables, and Severe is the omitted category.  $\text{Risk}_{3,8}$  is a dummy variable for 3.8%, and 6.5% is the omitted category. The relative importance of each symptom profile and risk level is the estimated preference weight for that outcome. Omitted categories are assigned a weight of 0.

Table 3 shows the difference in SC utility from more efficacious treatments that improve symptoms from Severe to each of the better categories. Because Severe is the omitted category with weight = 0, the change in  $U$  is just the estimated preference weight.

<sup>4</sup> We use the term “SC utility” in the conventional economic sense, not the specialized sense of a standard-gamble or time-tradeoff utility that ranges between zero for death and one for perfect health.

**Table 3. Efficacy, Satisfaction, and Maximum Acceptable Risk**

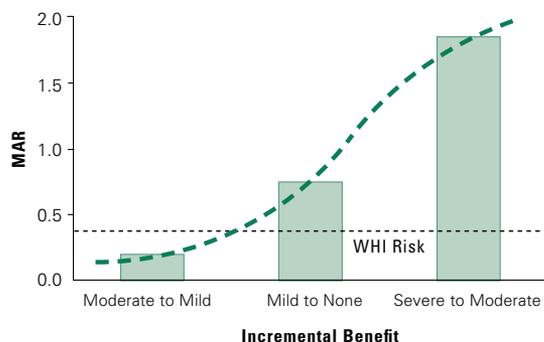
Treatment Efficacy	Change in SC Utility	Maximum Acceptable Risk
Severe to Moderate	0.37	1.85%
Severe to Mild	0.41	2.05%
Severe to None	0.56	2.80%

MAR for each treatment is the increase in risk that would exactly offset the increase in SC utility from better efficacy. Offering a treatment that improves outcomes from Severe to Moderate increases patients’ SC utility by 0.37. According to the preference-weight estimates, each percentage-point increase in risk decreases patients’ SC utility by  $0.55/(6.5 - 3.8) = 0.20$ . Thus if the treatment has a risk of  $0.37/0.20 = 1.9\%$ , then the risk causes patients’ utility to decrease by 0.37 and they are no better off than they would be without treatment. If treatment risk is more than 1.9%, then patients would be worse off with the treatment than without it. If treatment risk is less than 1.9%, then patients would be better off with the treatment than without it.

Figure 3 indicates that patients are willing to accept a relatively large increase in heart-attack risk to improve vasomotor symptoms from Severe to Moderate, but only a small increase in risk for an improvement from Moderate to Mild. The risk tolerance for an improvement from Mild to None falls somewhere in between. The increase in heart-attack risk originally reported in the Women’s Health Initiative study was about 0.4% (Rossouw et al. 2002). These results suggest that women are willing to accept risks above clinically observed levels only for significant treatment benefits.

Figure 3 clearly indicates that risk tolerance increases with the level of benefit offered. The dashed line suggests the possible shape of the benefit-risk tradeoff curve from Figure 1. If the horizontal axis were scaled with a continuous efficacy variable instead of categorical benefit levels, we could estimate the dashed benefit-risk tradeoff curve directly.

**Figure 3. Maximum Acceptable Risk for Vasomotor Symptom Treatment**



## Hypothetical Bias

The most serious limitation of SC methods is that they employ judgments among hypothetical alternatives. Hypothetical choices do not have the same clinical, financial, and emotional consequences of actual choices. Ideally, we would infer patients' preferences from observing real health-care decisions. Unfortunately, such data generally is nonexistent for the domains of interest or actual choices are so constrained by institutional, informational, and financial factors that observed choices are a poor indicator of patient preferences. A controlled experiment, even if it involves constructed alternatives, may be a far more valid source of data than actual decisions. We can limit the problem of hypothetical bias by constructing choice tasks that mimic realistic clinical choices as closely as possible.

In addition, SC judgment tasks encourage subjects to explore their preferences for various attribute combinations. This process of explicitly trading off well-defined attributes encourages subject introspection despite the hypothetical context. Because each subject provides answers to multiple tradeoff questions, SC allows analysts to devise internal checks for attentiveness and consistency. Such tests can identify subjects who are insufficiently attentive to the tradeoff tasks or whose stated preferences do not conform to basic requirements of logic and consistency for other reasons. For example, a significant challenge in applying SC methods to evaluate benefit-risk tradeoff preferences is that some subjects may have difficulty understanding

how risk is quantified. Devising effective ways of helping subjects conceptualize risk is an essential requirement for obtaining valid and reliable benefit-risk preference data.

## Using Stated Preferences to Inform Decisions Involving Benefit-Risk Tradeoffs

Understanding patient and physician perceptions, preferences, and choices is important for reducing the incidence of adverse outcomes while making treatment benefits available to the largest possible number of patients. A carefully designed and skillfully implemented SC survey can produce valid and reliable estimates of patients' and other stakeholders' risk tolerance. These estimates may be useful, in conjunction with traditional forms of evidence, for informing product-development decisions, licensing decisions, and clinical choices among alternative therapies.

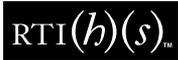
Unlike product-development and licensing decisions, clinical decisions involve individual patients who vary with respect to the likelihood they would experience given benefits or harms, but also vary with respect to their risk tolerance. It is possible that particular patients may face a relatively large chance of experiencing a harmful side effect, but have a high tolerance for risk, while other patients have a very small chance of experiencing a harmful side effect, but be intolerant of even small treatment-related risks. SC data can help identify the factors that influence patients' risk tolerance. This information may help physicians and patients identify treatments that are most consistent with both their health condition and their attitudes toward risk bearing.

When attributes and levels are carefully selected, SC data can provide useful answers to a variety of "what if" questions. For example, understanding what kinds of side-effect risks are of greatest concern to patients may help identify appropriate strategies for modifying drug formulations, designing more useful labels, or helping physicians communicate more effectively with their patients. SC researchers can then predict how such changes would affect patient and physician choices.

## References

- Bryan, Stirling, Martin Buxton, Robert Sheldon, and Alison Grant. 1998. "Magnetic Resonance Imaging for the Investigation of Knee Injuries: An Investigation of Preferences." *Health Economics* 7:595-603.
- Dey, Aloke. 1985. *Orthogonal Fractional Factorial Designs*. New York: Halstead Press.
- Huber, Joel. 1997. What We Have Learned from 20 Years of Conjoint Research: When to Use Self-Explicated, Graded Pairs, Full Profiles or Choice Experiments. Sawtooth Software Annual Conference:1-16. (<http://www.sawtoothsoftware.com/download/techpap/WhatweMS.zip>)
- Huber, Joel, and Klaus B. Zwerina. 1996. "The Importance of Utility Balance in Efficient Choice Designs," *Journal of Marketing Research*, 33 (August 1996): 307-317.
- Institute of Medicine. 2007. *The Future of Drug Safety: Promoting and Protecting the Health of the Public*. Institute of Medicine (US).
- Johnson, F. Reed, William H. Desvousges, Melissa C. Ruby, David Stieb, and Paul De Civita. 1998. "Eliciting Stated Preferences: An Application to Willingness to Pay for Longevity." *Medical Decision Making* 18 (Suppl):S57-S67.
- Johnson, F. Reed, and W.H. Desvousges. 1997. "Estimating Stated Preferences With Rated-Pair Data: Environmental, Health, and Employment Effects of Energy Programs." *Journal of Environmental Economics and Management* 34:79-99.
- Johnson, F. Reed, Melissa Ruby Banzhaf, and William H. Desvousges. 2000. "Willingness to Pay for Improved Respiratory and Cardiovascular Health: A Multiple-Format Stated-Preference Approach." *Health Economics* 9:295-317.
- Johnson, F. Reed, Semra Özdemir, Carol Mansfield, Steven Hass, David Miller, Corey Siegel. 2007a. "Crohn's Disease Patients' Risk-Benefit Preferences: Serious Adverse Event Risks Versus Treatment Efficacy." *Gastroenterology* 133(3):769-79.
- Johnson, F. Reed, George Van Houtven, Semra Özdemir, Steven Hass, Jeffrey White, Gordon Francis, David Miller, J. Theodore Phillips. 2008. "Multiple Sclerosis Patients' Risk-Benefit Preferences: Serious Adverse Event Risks versus Treatment Efficacy." *Journal of Neurology* (in press).
- Johnson, F. Reed, Semra Özdemir, A. Brett Hauber, Teresa L. Kauf. 2007b. "Women's Stated Willingness to Accept Perceived Risks for Vasomotor Symptom Relief." *Journal of Women's Health* 16(7):1028-40.
- Kuhfeld, Warren F., Randall D. Tobias, and Mark Garratt. 1994. "Efficient Experimental Design with Marketing Research Applications." *Journal of Marketing Research* 31:545-557.
- Rossouw, Jacques E., Garnet L. Anderson, Ross L. Prentice, et al. 2002. "Risks and Benefits of Estrogen Plus Progestin in Healthy Postmenopausal Women: Principal Results from the Women's Health Initiative Randomized Controlled Trial." *Journal of the American Medical Association* 288:321-333.
- Ryan, Mandy, and Jenny Hughes. 1997. "Using Conjoint Analysis to Assess Women's Preferences for Miscarriage Management." *Health Economics* 6:261-273.
- Viscusi, W. Kip, Wesley A. Magat, and Joel Huber. 1991. "Pricing Environmental Health Risks: Survey Assessments of Risk-Risk and Risk-Dollar Trade-offs for Chronic Bronchitis." *Journal of Environmental Economics and Management* 21:32-51.
- Wittink, Dick R., and Philippe Cattin. 1989. "Commercial Use of Conjoint Analysis: An Update." *Journal of Marketing* 53:91-96.
- Zwerina, Klaus, Joel Huber, and Warren F. Kuhfeld. 1996. "A General Method for Constructing Efficient Choice Designs." SAS Technical Document TS-694E. [http://support.sas.com/techsup/tnote/tnote\\_stat.html](http://support.sas.com/techsup/tnote/tnote_stat.html).





**RTI HEALTH SOLUTIONS®**

**RTI HEADQUARTERS:  
Research Triangle Park, NC, USA**

3040 Cornwallis Road  
Post Office Box 12194  
Research Triangle Park, NC 27709-2194 USA  
Telephone: +1.919.541.6000  
Facsimile: +1.919.541.5985

**Ann Arbor, MI, USA**

3005 Boardwalk Street, Suite 105  
Ann Arbor, MI 48108-5218  
Telephone: +1.734.213.5372

**Waltham, MA, USA**

1440 Main Street, Suite 310  
Waltham, MA 02451-1623  
Telephone: +1.781.434.1700

**Barcelona, Spain**

Trav. Gracia 56, Atico 1  
08006 Barcelona, Spain  
Telephone: +34.93.241.7766

**Manchester, UK**

Williams House  
Manchester Science Park  
Lloyd Street North  
Manchester, M15 6SE  
Telephone: +44.161.232.3400

**Ottawa, Canada**

220 Laurier Avenue West, Suite 350  
Ottawa, ON K1P 5Z9  
Telephone: +1.613.742.8225

---

[rtihealthsolutions@rti.org](mailto:rtihealthsolutions@rti.org)  
[www.rtihs.org](http://www.rtihs.org)

LEADING RESEARCH...  
MEASURES THAT COUNT