

RESEARCH REPORT

OCCASIONAL PAPER

# Risk Evaluation and Mitigation Strategy Assessments: Social Science Methodologies to Assess Goals Related to Knowledge

Elizabeth B. Andrews, Kimberly H. Davis,  
Dana B. DiBenedetti, Barbara H. Forsyth,  
Alicia W. Gilseman, Kelly A. Hollis,  
Sandy A. Lewis, Mark A. Price, Patricia S. Tennis,  
and Laurie J. Zografos

January 2014

**RTI** Press

RTI  
INTERNATIONAL

## About the Authors

All authors are or were members of the staff of RTI Health Solutions (RTI-HS), a business unit of RTI International. RTI-HS provides consulting and research expertise to pharmaceutical, biotechnology, diagnostics, and medical device companies. RTI-HS has office locations in the US, UK, Spain, and Sweden. Unless otherwise mentioned, the authors are based in the RTI-HS home office in Research Triangle Park, North Carolina, USA.

**Elizabeth B. Andrews**, PhD, MPH, FISPE, is vice president of Pharmacoepidemiology and Risk Management.

**Kimberly H. Davis**, MS, is a director in the Surveys and Observational Studies group.

**Dana B. DiBenedetti**, PhD, is a head of the Patient-Reported Outcomes group and is based in Dallas, Texas.

**Barbara H. Forsyth**, PhD, was a senior director of the Patient-Reported Outcomes group when this report was written.

**Alicia W. Gilsenan**, PhD, is a senior director in the Epidemiology group.

**Kelly A. Hollis**, MBA, is a global head in the Surveys and Observational Studies group.

**Sandy A. Lewis**, BSN, RN, is a director in the Patient-Reported Outcomes group and is based in Jenkintown, Pennsylvania.

**Mark A. Price**, MA, MEd, is a director in the Survey and Observational Studies group.

**Patricia S. Tennis**, PhD, MPH, FISPE, is a senior director in the Epidemiology group.

**Laurie J. Zografos**, BS, is a senior director in the Surveys and Observational Studies group.

This publication is part of the RTI Research Report series. Occasional Papers are scholarly essays on policy, methods, or other topics relevant to RTI areas of research or technical focus.

RTI International  
3040 East Cornwallis Road  
PO Box 12194  
Research Triangle Park, NC  
27709-2194 USA

Tel: +1.919.541.6000  
Fax: +1.919.541.5985  
E-mail: [rtipress@rti.org](mailto:rtipress@rti.org)  
Website: [www.rti.org](http://www.rti.org)

RTI Press publication OP-0018-1401

This PDF document was made available from [www.rti.org](http://www.rti.org) as a public service of RTI International. More information about RTI Press can be found at <http://www.rti.org/rtipress>.

RTI International is an independent, nonprofit research organization dedicated to improving the human condition by turning knowledge into practice. The RTI Press mission is to disseminate information about RTI research, analytic tools, and technical expertise to a national and international audience. RTI Press publications are peer-reviewed by at least two independent substantive experts and one or more Press editors.

## Suggested Citation

Andrews, E. B., Davis, K. H., DiBenedetti, D. B., Forsyth, B. H., Gilsenan, A. W., Hollis, K. A., Lewis, S. A., Price, M. A., Tennis, P. S., and Zografos, L. J. (2014). *Risk evaluation and mitigation strategy assessments: Social science methodologies to assess goals related to knowledge* (RTI Press publication OP-0018-1401). Research Triangle Park, NC: RTI Press.

©2014 Research Triangle Institute. RTI International is a trade name of Research Triangle Institute.

All rights reserved. This report is protected by copyright. Credit must be provided to the author and source of the document when the content is quoted. Neither the document nor partial or entire reproductions may be sold without prior written permission from the publisher.

[doi:10.3768/rtipress.2014.op.0018.1401](https://doi.org/10.3768/rtipress.2014.op.0018.1401)

[www.rti.org/rtipress](http://www.rti.org/rtipress)

# Risk Evaluation and Mitigation Strategy Assessments: Social Science Methodologies to Assess Goals Related to Knowledge

Elizabeth B. Andrews, Kimberly H. Davis,  
Dana B. DiBenedetti, Barbara H. Forsyth,  
Alicia W. Gilsonan, Kelly A. Hollis,  
Sandy A. Lewis, Mark A. Price, Patricia S. Tennis,  
and Laurie J. Zografos

---

## Abstract

On May 3, 2012, the US Food and Drug Administration (FDA) announced a public workshop to be held June 7, 2012, to receive input on “survey methodologies and instruments that can be used to evaluate patients’ and health care providers’ knowledge about the risks of drugs marketed with an approved REMS [risk evaluation and mitigation strategy].” The FDA intended to use this input to help develop guidance to industry regarding best practices for such research. In the announcement to the meeting, entitled “Risk Evaluation and Mitigation Strategy (REMS) Assessments: Social Science Methodologies to Assess Goals Related to Knowledge: Public Workshop,” the FDA provided an issue paper summarizing experience with prior REMS assessment surveys and posing a series of questions for which input was requested from the workshop panel and public.

RTI Health Solutions (RTI-HS) participated in the workshop by serving on the invited panel (Alicia Gilsonan and Karol Krotki), presenting to the panel (Kelly Hollis, Sandy Lewis, and Laurie Zografos), and submitting written responses to the FDA to the questions posed in the issues paper accompanying the meeting announcement.

This paper provides a brief background and presents the written responses that RTI-HS submitted to the FDA, with modest revisions to add context and clarity to the response for a wider readership. We also briefly discuss upcoming US and European steps in this field.

---

## Contents

<b>Background</b>	2
<b>FDA Questions and RTI-HS Responses</b>	3
Question 1—Recruitment	3
Question 2—Knowledge Rate	5
Question 3—Questionnaire Design	7
Question 4—Process Issues	11
Question 5—Alternatives	13
Question 6—Design Considerations	14
<b>Discussion</b>	14
<b>References</b>	15
<b>Acknowledgments</b>	Inside back cover

## Background

Monitoring the safety of medicines in humans begins during preapproval drug testing and continues throughout a medicine's marketed life cycle. During each stage, sponsoring companies and regulatory agencies must weigh the recognized and potential benefits of treatment against known and potential safety concerns. Benefits can include reduction of disease symptoms, management of chronic conditions such as hypertension and diabetes, and resolution of infection. Safety concerns may include frequent, familiar, and non-life-threatening side effects such as nausea, headache, and mild rashes; concerns also include rare but serious events such as life-threatening central nervous system infections and serious skin reactions.

In some cases, medications that might otherwise not be approved for marketing because of safety concerns are nevertheless approved because the benefits of treatment outweigh the risks or because the risks are considered acceptable to some patient groups who may have no other treatment alternatives. These circumstances often require the active management or mitigation of risks that goes beyond standard medical information that product sponsors provide through product labeling and promotional material aimed at prescribers, pharmacists, and patients. Active mitigation can include a controlled product distribution system in which patients cannot fill prescriptions without documentation of certain laboratory testing results indicating that they have no contraindications to the product; however, active mitigation can also include purely educational activities.

The Food and Drug Administration Amendments Act of 2007 granted the FDA authorization to require a risk evaluation and mitigation strategy (REMS) to assure that benefits of an approved drug outweigh risks. This new authorization resulted in increased FDA focus on the active management of identified risks of therapeutic agents, building on prior efforts, such as risk minimization action plans (RiskMAPs).<sup>1</sup> Three levels of risk minimization were specified: (1) elements to assure safe use (ETASU), the most restrictive set of tools to manage risks;

(2) communication plans, targeted communication programs delivered to health care providers or patients (or both); and (3) medication guides, an education sheet for patients. Drug risk minimization programs with only a medication guide may be considered a type of REMS, but the FDA can also require medication guides outside a REMS requirement.

The FDA requires evaluation of the REMS at intervals defined by the legislation (18 months, 3 years, and 7 years following REMS approval) or as negotiated with the FDA. Many of the evaluations have included surveys of knowledge and behavior of health care professionals and patients about the safety messages of the REMS communications and of actions that should be taken to manage risks.

The FDA sponsored the workshop "Risk Evaluation and Mitigation Strategy (REMS) Assessments: Social Science Methodologies to Assess Goals Related to Knowledge" on June 7, 2012, to share information and promote dialogue about best practices to evaluate the effectiveness of REMS programs.<sup>2</sup> In the May 3, 2012, public announcement of the meeting, the FDA provided an issue paper summarizing experience with prior REMS assessment surveys and posing a series of questions for which the agency requested input from the workshop panel and the public.<sup>3</sup>

RTI Health Solutions sought to provide feedback on the specific questions posed by the FDA because of our depth of general expertise, decades of experience in the design and conduct of survey research, and our recent direct experience evaluating REMS and RiskMAPs. Such assessments have included longitudinal surveys of patients enrolled in a RiskMAP for alosetron (a medication used to treat severe cases of irritable bowel syndrome in women who have diarrhea as a major symptom)<sup>4,5</sup>; cross-sectional surveys of physicians and other health care providers in a variety of products including natalizumab (a monoclonal antibody drug used to treat multiple sclerosis and Crohn's disease)<sup>6</sup>; studies of patients in special populations requiring identification through physician practices; and studies with patients recruited by pharmacies based on comprehensive lists of filled prescriptions.<sup>7,8</sup>

Various modes have been used in these studies to collect information, including web-based survey instruments, tablet or other hand-held devices with electronic transmission of data, paper forms mailed or faxed for data entry or scan, and in-person and telephone interviewing. We have introduced as many methods as possible into these studies to reduce potential biases, to elicit truthful and complete responses, and to allow for rigorous analyses. These methods include robust instrument development and testing, careful random selection methods, collection of information about nonrespondents, and the comparison of respondents with nonrespondents and with the full treated population to assess potential biases. Nevertheless, some residual bias may be present in any study that relies on sampling and self-reported knowledge and behavior.

Moreover, resources and timelines available for the design, conduct, and analysis of these studies are limited in comparison with large-scale population-based national surveys. These studies, therefore, represent a best effort to bring scientific rigor to the process of obtaining useful information about the effectiveness of REMS activities within the time constraints of the programs. Our experience in evaluating study results and our interaction with clients who have sponsored the studies have demonstrated in some cases that results are already being used to improve the nature and content of educational programs. In other cases, these studies demonstrate that the REMS programs are working very well.

RTI-HS participated in the workshop by serving on the invited panel (Alicia Gilsean and Karol Krotki), presenting to the panel (Kelly Hollis, Sandy Lewis, and Laurie Zografos), and submitting written responses to six of the seven FDA questions.<sup>9</sup> The seventh question was directed to pharmaceutical companies. The remainder of this paper presents the six FDA questions verbatim. We modified the RTI-HS verbatim responses to improve clarity for this paper, and we added a brief discussion section.

## FDA Questions and RTI-HS Responses

Below (in bold) are six questions and subquestions posed by the FDA. Our responses follow each question or question set. All questions and responses relate to cross-sectional surveys of health care providers or patients (or both) to evaluate the effectiveness of REMS. These surveys collect information on knowledge, understanding, and behavior relating to the safe use of a specific medicine.

### Question 1—Recruitment

1. **What strategies can the applicant use to recruit a sample that is representative of the population that is prescribing/dispensing/taking the drug?**
  - a. **Given that the applicant cannot compel an individual to complete a survey, is it acceptable to enroll a relatively small (making the survey feasible) number of participants who are representative of the totality of the health care provider or patient population and make generalizations from that sample to the larger population?**
  - b. **What is an adequate sample size to be able to confidently extrapolate findings to the entire population prescribing/dispensing/taking a drug?**

### Sampling From the General Population

Obtaining a representative sample of the population using a new drug is challenging because (1) there is no existing list of all patients taking a new drug from which to sample, and (2) the proportion of people taking a new drug within a given time frame is typically extremely small. One might think that taking a random sample of the general population would be the most straightforward approach. However, using this approach would require selecting a very large sample size from the general population to obtain a small number of patients taking the new drug of interest.

For example, even with a relatively common condition such as diabetes, the proportion of individuals in the population who are starting a new antidiabetic drug is likely to be small. In 2001–2002, the prevalence of diagnosed diabetes among those aged 20 to 64 years was 5 percent, and among

individuals aged 65 years and older, the prevalence was 15 percent. Around 80 percent of individuals younger than 65 years of age were treated,<sup>10</sup> and about half of those 65 years and older were treated.<sup>11</sup> Within the population of patients with treated diabetes, only a small proportion will have started a treatment that year, and the vast majority are likely to start on an established first-line medication. Thus, one could expect that only a small proportion of the general population would start a new antidiabetic drug during the first year of availability; approximately 0.1 percent would have started the new drug of interest. Therefore, to recruit 200 people starting the new drug of interest, one would have to select a random sample of at least 200,000 people.

This information is supported by a multinational case-control study,<sup>12</sup> in which the numbers of individuals taking specific medications was low among 1,500 control patients hospitalized for acute conditions. Because the control patients were hospitalized, the prevalence of medication use among the controls in this study may very well have overestimated the prevalence among the general population. Clearly, alternative approaches to targeting the patient population of interest are required.

### Minimizing Bias Involves Selecting the Best Sampling Frame

Because selecting a random sample of the entire population is not realistic, and because no nationwide list exists of all patients receiving any given drug from which to draw a sample, patients must be recruited from some source that represents only a subset of patients of interest. In survey research, this source is described as the **sampling frame**, which should be as similar as possible to the overall population of interest. Conducting studies that evaluate patient knowledge and use of educational materials, even under the best of circumstances, is subject to numerous biases that may not be completely eliminated. However, investigators can adopt several strategies to minimize the biases associated with selecting the target survey population.

Potential sampling frames (e.g., pharmacy networks, patient advocacy groups, patients with selected

physician practices or insurance plans, consumer panels) should be evaluated for suitability for a given study. The ready availability or convenience of a sampling frame that may maximize sample size must be weighed against the representativeness of the source. For example, a patient support group may offer a large number of willing survey participants, but it may be inherently biased for the purposes of the study; for example, a patient support group may include patients who have more severe disease or who have received more education about their medications than the overall population of patients taking the medications.

To ensure inclusion of the appropriate participants in a survey sample, it also is important to consider the accuracy of the data on patient or prescriber characteristics that are used to evaluate participant eligibility in the sampling frame. For example, the convenience of using consumer panels, which rely on patients to self-identify, must be weighed against other more scientifically valid assessments of patient eligibility (e.g., medical condition, medication exposure) that require confirmation by a physician or a pharmacy. Identifying a proper sampling frame is often feasible (e.g., utilizing a national pharmacy database to select a random sample of patients who have picked up a prescription for the medication of interest); therefore, selecting a sample that is likely to be representative of the population of interest is also feasible.

The coverage provided by any given sampling frame is an important consideration. Some evaluation programs may need to include multiple sampling frames in an effort to expand patient representation. However, if the selected sources do not effectively represent the target population, the combination of sampling frames may not be any more representative than a single source, although it may provide access to additional eligible participants.

A preferable strategy to assure appropriate representation of important subgroups is to use stratified sampling within a proper sampling frame with oversampling of groups that otherwise would be too small to enable meaningful analyses. The sampling strategy must be accompanied by the appropriate analyses; these may include weighting

of the results back to the characteristics of the target population.

## Evaluating the Success of the Sampling and Recruitment Strategies

Nonresponse bias is a concern with any voluntary survey. With the advent of the Internet, potential respondents, including both patients and physicians, are often burdened with multiple requests to complete surveys for various reasons. Minimizing nonresponse bias is becoming increasingly challenging, especially among physicians who receive multiple requests to complete REMS assessment surveys, often for drugs in the same class. One method to minimize nonresponse bias is to offer a reasonable incentive that compensates respondents for personal time spent but one that is not considered coercive. Other approaches include reminder letters, e-mails, or telephone calls to nonrespondents; however, because of the time constraints for completion of REMS assessments, following up with nonrespondents is not always feasible.

To determine whether the selected recruitment strategies are successful at enrolling respondent populations who reflect the target population, studies should compare characteristics of responders and nonresponders to determine if they differ in ways (e.g., medication compliance, awareness of risks) that would influence the validity and generalizability of the study results. The availability of data and access to individual-level data on specific characteristics to allow comparisons of responders and nonresponders are often restricted. In studies in which patients are identified by a pharmacy network via a prescription database, comparing a limited number of characteristics (e.g., state of residence, age or age category, sex, and type of health insurance) for responders and nonrespondents is possible. Data on other demographic variables (e.g., socioeconomic status, level of education) would be useful but are often not available. Other data sources (e.g., REMS programs for drugs that require all patients taking a medication to be registered) may provide data on a large number of variables.

Additional analyses that are useful to evaluate responder bias include comparison of the age and sex distribution of respondents to the expected

age and sex distribution of the patient populations taking the drug of interest and descriptive analyses of the proportion of participants who complete the required REMS assessment, stratified by mode of data collection (e.g., paper and pencil, web-based) if multiple modes of data collection are employed.

## Defining an Adequate Sample Size

We recommend using the minimum acceptable precision (i.e., confidence interval) for the primary outcome of interest to determine the target sample size. For example, if the primary outcome measure is prevalence of accurate knowledge defined as the percentage of respondents who correctly answer a single question, the expected true knowledge is 85 percent, and the minimum acceptable width of the 95 percent confidence interval is 10 percentage points, a sample size of 200 would be needed. Identifying the primary metric for the survey is key to determining the sample size.

## Question 2—Knowledge Rate

2. **Is the *knowledge rate* (i.e., the proportion of subjects who demonstrate knowledge of the risk message) the appropriate primary endpoint for a survey?**
  - a. **What factors need to be considered when establishing the threshold for success for educational elements of the REMS?**
  - b. **Should the threshold for successfully meeting a REMS educational goal be set at a knowledge rate of 80% or 90%, or should it vary depending on the risk message? If it should vary, what should the minimum threshold for success be? Should the threshold reflect whether the product is a new molecular entity, an original biologic product, or an older drug?**

Prevalence of accurate knowledge (commonly referred to as *knowledge rate*) is an appropriate endpoint for surveys. However, consideration must be given to whether a single expected knowledge rate should be applied across all REMS programs or whether thresholds should vary based on factors such as the importance of the risk message, the nature of the underlying condition, and the specific patient population. If thresholds are to be set by the regulatory agency or through agreement between

regulator and study sponsor, then how should the study results be interpreted? Study sponsors have legitimate concerns about potential implications of a program's failing to meet an established minimum threshold of knowledge, especially if other aspects of a REMS program are working well, as demonstrated by other measures of behavior and safety.

Many evaluation plans have established prespecified thresholds; however, establishing thresholds *a priori* and applying the same threshold across all REMS programs has inherent difficulties. First, thresholds may be arbitrarily set, not based on evidence of the minimum knowledge necessary to minimize the risks associated with a drug. Furthermore, knowledge may be high among a specific subgroup for whom the risk of the medication is most relevant, but overall knowledge rate may fail to meet the threshold. For example, the results from a recent REMS assessment survey showed that the sample population overall had relatively low awareness about a particular risk; however, knowledge levels were higher among a subgroup of participants for whom the risk was most relevant. If a threshold had been set *a priori*, the overall program may have failed to achieve the threshold, when in fact those participants who were most at risk were knowledgeable about that specific risk.

Results from REMS assessment surveys vary significantly across programs. Both patient and drug program factors can influence such results about levels of patient awareness (Table 1).

Reviewing results pooled across surveys for several different medications can provide an opportunity to evaluate the potential factors associated with knowledge of information communicated in the educational materials and can help to improve design for future REMS programs. For example, we examined anonymized pooled data from six REMS assessment patient surveys ranging in size from 200 to 9,000 respondents.<sup>7</sup> Within each survey, results on awareness of the primary medication risk were stratified by medication guide receipt and review, type of REMS program, disease type, and patient characteristics. We then compared results across surveys to identify patterns.

**Table 1. Selected factors that may influence patient awareness**

Patient Factors	Drug/Program Factors
<ul style="list-style-type: none"> <li>• Age</li> <li>• Education</li> <li>• Health literacy and numeracy</li> <li>• Recency of diagnosis and treatment</li> <li>• Health status (e.g., number of conditions and medicines)</li> <li>• Prior experience with the adverse event</li> <li>• Cognitive impairment</li> </ul>	<ul style="list-style-type: none"> <li>• Severity of underlying condition</li> <li>• Nature of treatment (chronic vs. short-term)</li> <li>• Severity of the adverse events</li> <li>• Recency of drug on the market</li> <li>• Nature of the REMS program (e.g., ETASU vs. only a medication guide)</li> <li>• Complexity of the safety information (e.g., number of risks, terminology)</li> <li>• Counseling by health care provider</li> </ul>

ETASU = elements to assure safe use; REMS = risk evaluation and mitigation strategy.

The six patient surveys covered drugs in five disease areas; two were for acute conditions, one for an intermittent condition, and three for chronic conditions. Awareness of the primary risk associated with each product ranged from 24% to 98% and was lowest for medications treating acute conditions. The highest percentage of correct responses was found in surveys in which the medication was used to treat chronic conditions, had more severe side effects, or had a REMS program that included more elements than just a medication guide. Moreover, knowledge about the medication risk was higher among certain patient categories (e.g., new users of a medication, patients reporting they were counseled by a health care provider) in surveys that included these types of questions.

Numerous other factors can influence results of knowledge surveys and should be considered when regulators and sponsors agree to recommend thresholds for these evaluations. Examples of these factors are as follows:

- Appropriate thresholds may vary based on the population and the type and magnitude of risks (e.g., 70% may be acceptable in one program, whereas 90% is required in another). Because determinations of success may be challenged if a

clear standard is not established, requiring sponsors to provide justification for selected thresholds in their study protocols may be preferable. Thresholds may be set for the overall sample or be different for specific subgroups. For instance, if the primary risk is to women of childbearing potential, should the same knowledge threshold be applied to these individuals as to men?

- In many cases, multiple risks are associated with a particular drug. Therefore, consideration should be given to two issues: (1) whether the threshold should be set on knowledge of only the primary risk or include other important risks as well and (2) whether secondary outcomes should have the same threshold as the primary outcome. Furthermore, consideration should be given to whether knowledge of risk(s) is sufficient or whether thresholds should also be set for awareness of signs and symptoms of potential adverse reactions.
- In addition to the knowledge about specific risks, surveys of patients and health care providers often contain questions regarding appropriate behavior (e.g., laboratory monitoring, what to do if signs or symptoms are present). Such questions may be as important as the knowledge questions. We have observed high awareness of medication risks among physicians for one program without corresponding awareness of the need for laboratory monitoring relating to the specific risk.
- If multiple risks or symptoms are presented in a question, should the response be categorized as incorrect if any portion is incorrect, or should respondents receive “partial credit”?
- Composite scores also may be considered in defining thresholds.

Establishing a standard threshold for all REMS assessment surveys may not provide the flexibility to evaluate the impact of educational materials on those individuals with the highest risk. Rather, programs should be evaluated on their individual characteristics to determine appropriate thresholds, if any. For example, targeting knowledge thresholds for patient surveys on content areas that the patient can control may be better than targeting more general knowledge; an example is knowing what to do if the patient

experiences extreme dizziness and fatigue rather than knowing what potential rare life-threatening illness may be associated with the medication the patient has been prescribed.

Finally, the percentage of individuals who demonstrate correct knowledge is a measure of only one component of a risk minimization strategy. A program may minimize risk adequately through intervention by a physician or other health care provider, despite poor awareness among patients. The results of knowledge-awareness surveys should always be evaluated in the context of the success of the total risk minimization program.

### Question 3—Questionnaire Design

**3. Since most surveys use only true/false and multiple-choice questions, what are the advantages and disadvantages of using other question types (open ended, case vignettes, fill in the blank) to evaluate knowledge?**

Data collection instruments are developed for REMS programs to evaluate knowledge and understanding and, sometimes, behaviors. Their effectiveness depends on meeting the standards of sound survey design and implementation while following standard principles of test development. The American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education<sup>13</sup> have published standards that identify best practices for test design, development, and revision. Updated standards are currently being developed and will include a set of standards for “operations” that mirror the current standards for test design and development.<sup>14</sup> The standards identify four overarching steps in the process of test construction, which have been summarized as follows:<sup>15</sup>

- Identify the primary purposes and the scope for an assessment.
- Develop a test framework and test specifications.
- Develop, test, and select items and scoring rubrics for the assessment.
- Assemble an operational test and evaluate it in preparation for deployment.

These standards indicate that test specifications should include item and response formats, in addition to item content. All of these should be selected based on test objectives.

Extending these principles to the REMS assessment context, no single item type is ideal for all REMS surveys. Rather, the types of items used will likely vary depending on the goals for the specific REMS program and the purposes for including patient, provider, or pharmacy surveys (or combinations) as part of a REMS assessment. The standards mention one advantage of selected response formats—including multiple-choice and true/false formats—which may explain their ubiquity in past REMS surveys. These response formats are versatile and, therefore, effective for a range of program goals and a variety of test purposes.

For patient surveys, a common objective includes assessment of the frequency with which patients (1) receive the required patient educational materials, (2) read the required patient educational materials, (3) demonstrate general knowledge of correct use of the drug, (4) demonstrate knowledge of serious risks associated with the drug, (5) are aware of actions to take if they develop an adverse reaction to the drug, and (6) use a drug appropriately. The first four objectives are well served by alternate choice (e.g., yes/no; true/false), multiple-choice, or matching item and response formats. Documenting patients' experiences receiving and reading required materials and their general knowledge of the contents of those materials requires that the patients remember materials they received and read as well as the key points the materials cover. These objectives do not necessarily require that patients be able to freely recall either the materials by name or key points the materials cover, particularly key points that include unfamiliar vocabulary, technical terminology, or relatively detailed information. The last two objectives may be better assessed through a mix of selected response items (alternative choice, multiple-choice, and matching) and the addition of short answer items (e.g., fill in the blank), because it is reasonable to expect that appropriate use requires at least some free recall.

For prescriber and pharmacist surveys, the common objectives are (1) discerning their understanding of the requirements for distributing the patient materials; (2) assessing their access to and knowledge of any relevant prescriber or pharmacist guide or checklist materials; (3) determining their familiarity with appropriate prescribing; (4) assessing their understanding of elements to assure safe use (ETASU); and (5) measuring their knowledge of important risks, known safety issues, and adverse event reporting requirements. A mix of recognition-based (e.g., multiple-choice) and free-recall-based (e.g., short answer) items may be appropriate for measurement goals such as these that require more detailed knowledge and the ability to act appropriately in the absence of explicit prompts. Note that free-recall-based questions may be challenging to implement in the short REMS assessment surveys and may make interpretation and analysis of results difficult.

Although the testing standards emphasize the priority of matching item and response formats to measurement goals, test developers recognize that considering additional pragmatic factors is important. These include the time required to complete item sets, impacts of extraneous factors (e.g., reading level, writing skills), and time and costs required to ensure consistent scoring.<sup>15-17</sup> These additional pragmatic factors are likely contributors to the prevalent use of selected response formats in past REMS surveys. Many of these pragmatic factors are identified in Table 2, which lists some key advantages and disadvantages for various item and response formats.

Of course, the REMS survey context differs in important ways from a typical testing environment. Notably, test-takers are frequently highly motivated to score well. In addition to testing standards, well-established survey design principles and practices also inform effective REMS survey questionnaire design and item development.

Table 2. Key advantages and disadvantages for alternative item and response formats

General Response Format	Item Type(s)	Advantages	Disadvantages
<b>Selected response</b>	True/false and multiple choice	<ul style="list-style-type: none"> <li>• Low burden to respondents</li> <li>• Easy to administer many items in a relatively short time, allowing fuller sampling of content domain</li> <li>• Low dependency on high reading levels and well-developed writing skills</li> <li>• Relatively easy to develop scoring rubrics for response formats that include one correct response</li> <li>• Relatively easy to ensure consistent scoring</li> <li>• Relatively fast scoring; can be scored by machine</li> <li>• Minimal per-item costs for scoring</li> <li>• Generally easy to administer in a variety of modes, including telephone, mail, and Web survey modalities</li> <li>• Straightforward approaches for assessing differential item function</li> <li>• Relatively straightforward analysis and reporting</li> </ul>	<ul style="list-style-type: none"> <li>• Item development requires well-articulated measurement goals and skilled item writers</li> <li>• Foil selection is a key factor affecting item difficulty</li> <li>• Clear question specifications are required for appropriate foil development and selection</li> <li>• Weighted scoring rubrics can be difficult to develop when response formats include options that are partially correct</li> <li>• Difficult to assess higher-order thinking skills, which may be important for surveying prescribers and/or pharmacists</li> <li>• Response selection prone to effects of guessing</li> </ul>
<b>Constructed response</b>	Short answer	<ul style="list-style-type: none"> <li>• With careful development, may be able to assess deeper respondent understanding</li> </ul>	<ul style="list-style-type: none"> <li>• Takes somewhat longer to answer, increasing burden and/or reducing the number of items and breadth of construct sampled per unit of time</li> <li>• Respondents are more likely to react to extraneous aspects of item or response task</li> <li>• Consistent scoring requires carefully developed and well-circumscribed scoring rubric (e.g., a set of acceptable responses)</li> <li>• Resulting data are likely to be somewhat more difficult to analyze and report than data from selected response format</li> </ul>
<b>Constructed response</b>	Extended response	<ul style="list-style-type: none"> <li>• Particularly effective for assessing relatively deep understanding</li> </ul>	<ul style="list-style-type: none"> <li>• Takes longer to administer and/or answer</li> <li>• Relatively costly and slow scoring</li> <li>• Effective scoring rubric requires effectively anticipating respondent answers</li> <li>• Unanticipated scoring factors often require adjustments, retraining, and rescoring</li> <li>• Analytic scoring requires identifying key dimensions for scoring and developing scoring rubrics for each</li> <li>• Holistic scoring requires identifying a single effective scoring rubric that is effective across possible item dimensions</li> <li>• Reliable and valid scoring requires detailed scoring rubric, careful training, monitoring, and retraining as needed</li> <li>• Respondents may react to extraneous aspects of item or response task</li> <li>• Generalizability susceptible to person-by-task interaction</li> <li>• Relatively difficult to assess differential item functioning</li> <li>• Depending on scoring approaches, data may be particularly difficult or time-consuming to analyze and report</li> </ul>

(continued)

**Table 2. Key advantages and disadvantages for alternative item and response formats (continued)**

General Response Format	Item Type(s)	Advantages	Disadvantages
<b>Vignette item formats</b>	—	<ul style="list-style-type: none"> <li>• Versatile: can be used with a variety of response formats, depending on measure models and measurement goals</li> <li>• Face validity: typically perceived as more authentic</li> </ul>	<ul style="list-style-type: none"> <li>• Relatively difficult to administer as part of a telephone survey</li> <li>• Dependent on reading level when administered in written format</li> <li>• Development requires detailed item specifications that (1) map vignette details to components of correct response and (2) ensure item developers avoid including misleading or superfluous information</li> <li>• Respondents may react to extraneous aspects of item or response task</li> <li>• Requires sufficient pretesting to avoid differential item function across cultural and sociodemographic groups</li> </ul>

Source: Modified from Andrews et al., 2012.<sup>9</sup>

Social desirability is an important questionnaire design issue, because REMS surveys often focus on measuring respondent knowledge. Carefully constructed instructions placed strategically throughout a questionnaire can help to diffuse the potential impact of social desirability bias by clarifying that the intent is to evaluate a REMS *program* rather than the individual *respondent*. Also, survey methodologists typically recommend avoiding open-ended, constructed response items because perceived burden may influence the level of detail respondents are willing to record. Finally, depending on how they are scaled and scored, data from open-ended constructed response items may be difficult, time-consuming, and costly to analyze and report.

Research on survey design and survey methodology plus experience have informed our guidance on question wording and design (see Table 3). Item developers should take care to either avoid or clearly define any medical and other technical terminology, select concrete and unambiguous item wording, and clearly specify the reporting time frame or reference period. Item developers should carefully consider whether and when to offer an “I don’t know” response choice. Offering “I don’t know” as an option can reduce guessing but also can have negative effects on accurate assessment if respondents overuse it.<sup>18-20</sup> Furthermore, evidence suggests that forced-choice items are processed more deeply and more carefully

**Table 3. Good item-development practices**

- Rely on closed-ended, selected response formats as often as feasible given measurement goals
- Avoid check-all-that-apply question structures
- Keep language simple and provide descriptions for medical terminology that may be unfamiliar to the respondent population
- Use simple sentence structure
- Present questions using concrete, unambiguous language
- Limit recall period
- Minimize the burden to the respondent, by keeping the instrument short and using constructed response formats judiciously
- Pay attention to the order of questions to ensure that general or the most important questions are asked early
- Use cognitive testing methods throughout questionnaire development, to ensure that wordings are interpreted as intended and also understood consistently across respondents

than mark-all-that-apply items that cover the same content.<sup>21</sup>

Considerable evidence supports the fact that, in survey measurement, response probability is related to the order in which responses are presented.<sup>22,23</sup> Furthermore, order has different effects when items are presented visually than when they are presented

orally (e.g., by telephone). In electronic data collection, randomizing the order of responses across respondents is one approach for reducing the impact of response order on survey estimates.

## Question 4—Process Issues

### 4. Please discuss process issues related to these surveys:

#### a. Given issues of recall, should the lag time between the REMS communication and the survey administration be standardized?

The lag time between a REMS communication and survey administration should not be standardized given the variability associated with the unique circumstances of the selected treatment and the study population. For patients, the nature of the condition and length of time on treatment should be taken into account. For example, the lag time for acute conditions requiring a short treatment course (e.g., antibiotics) might reasonably be shorter than the lag time for a chronic condition for which treatment is used over a long period of time. In addition, if the survey is intended to ask patients about their reading of the medication guide, the time frame should allow for patient recall of the process of receiving and reading the medication guide.

Furthermore, recall may be higher among new users of a chronic medication than among experienced users because experienced users have already made the decision to continue on the medication, so recall of safety messages may no longer be relevant to them. Also, lag times may have practical implications on the ability to recruit patients; that is, shorter time frames may limit the pool of eligible participants and preclude certain methods of recruitment. Once a time frame is established, its appropriateness should be evaluated in cognitive interviews.

#### b. Should pretesting/validation be required, to reduce the likelihood of a poorly worded question that was not recognized during survey development?

Despite the attention and proficiency devoted to questionnaire design, pilot testing the questions, response choices, and instructional text with the target population is critical to identify and eliminate

potential sources of measurement error.<sup>24</sup> The cognitive interview process, which was developed in the 1980s, is a rigorous qualitative research methodology to optimize survey instructions, question wording, response options, and questionnaire format by evaluating the cognitive process that respondents use to answer questions, including item comprehension, information retrieval, and response selection.<sup>25,26</sup> Hence, conducting cognitive interviews before REMS survey implementation to pretest the data collection instruments will optimize data quality and contribute to the success of any REMS assessment program.

Cognitive pretesting ensures that respondents interpret the questions and formulate responses as easily and consistently as possible. Interpretation should not only be consistent across respondents, but also consistent with the researcher's intentions for measurement.<sup>27</sup> Sets of cognitive interviews are conducted iteratively so that potential problems can be addressed and the adequacy of modifications demonstrated before fielding the survey.

Cognitive pretesting/debriefing interviews are typically conducted face-to-face with individual members of a target population (e.g., patients with a certain condition, patients taking a particular medication, physicians with experience treating a certain patient population with a specific medication or treatment). This type of one-on-one, in-person interview explicitly focuses on the cognitive processes that participants use to understand and answer questions. For example: What does the participant think the question is asking? What do specific words or phrases in the question and response choices mean to the participant? Are the response categories distinct? Both covert processes and overt, observable processes can be studied in these interviews.<sup>28</sup>

Specifically, cognitive debriefing interviews provide insight into the following issues:

- Identifying items that might be misinterpreted, permitting researchers to identify items that are not optimally worded, either because they are difficult to understand or because interpretations differ across patients

- Difficulties associated with information recall, such as asking about a time period that is either too long or too short for the concept being measured
- Difficulties or biases in response selection, including response categories that are not mutually exclusive and exhaustive and response categories that are socially undesirable<sup>29</sup>
- Assessment of face validity (do the questions appear to measure what was intended?)

Sponsors who field REMS surveys without a cognitive pretest in the population of interest risk inclusion of poorly written or understood instructions, questions, and response options. A poorly written or confusing question can cause a respondent to select the wrong answer when the respondent may really know the correct answer, which could lead to inaccurate interpretation of results.

In addition to identifying potentially problematic wording with survey items and response choices, pretesting REMS survey instruments in target populations (e.g., patients, physicians, and other health care providers such as pharmacists, nurses, and staff at centers where medications are administered by infusions) provides other valuable insights. In a recent cognitive pretest of a patient REMS survey, we evaluated alternative questions and response sets regarding patients' understanding of potential risks associated with a particular injection. Patients generally understood all alternatives, but found one of the alternatives to be frightening, which led them to overestimate the actual probability of the risk. Using an alternative question in the final survey reduced the likelihood that the survey itself was influencing the patients' continuation of treatment.<sup>28</sup>

Additionally, cognitive interviews capture insights not available otherwise. A recent study showed that participants taking a specific chronically used medication were unaware of key safety information about that medicine. Cognitive interviews revealed that these patients (who often had coexisting conditions and were taking 12 to 14 different medications) did not read written materials they received because there was simply "too much to go through." This insight led the sponsor to consider alternative means of communicating with this patient population.<sup>28</sup>

Cognitive pretesting is critically important in special population groups. Some patient groups may have cognitive, sensory, or communication challenges associated with their disease. During cognitive pretesting of a patient REMS survey instrument, RTI-HS moderators were able to listen to responses and observe the reactions to new questionnaire items among patients with varying severity levels of a neurologic condition. These patients had specific recommendations for wording to simplify and clarify questions: very concise wording; a small number of response choices; and capitalizing, bolding, or underlining key words and phrases within the items. In light of these observations, we added to the questionnaire a new item asking patients to rate the extent to which the disease had affected their cognitive abilities—to enable stratification of study results by patient self-reported cognitive function.

Also, the interviewer gleaned important information during this cognitive testing by observing the facial expressions and body language of patients as they reacted to the draft items. Patients' nonverbal reactions are often insightful and give interviewers cues for further probing. The importance of nonverbal communication emphasizes the need for in-person interviews.

Several books have been written on the cognitive interview process, including best practice techniques.<sup>24</sup> However, one important best practice is to ensure that the interviews are conducted by interviewers who are trained in questionnaire development and qualitative research methodologies and highly experienced.<sup>24,29</sup> The most effective cognitive interviewers are those with both technical expertise (e.g., in questionnaire design, survey development, and the survey topic) and highly developed interpersonal skills (e.g., being flexible, spontaneous, and calm in unexpected situations).<sup>24</sup> Cognitive interviewing is both a science and an art. The expertise and level of experience of the interviewer determine the value obtained from pretesting and therefore influence the quality of the instrument and robustness of the study results. Cost and time constraints should not be reasons for an instrument developer to circumvent the process of cognitive pretesting.

**c. On average, how long does it take to design, test, recruit participants, conduct, analyze, and report the results of a survey?**

REMS timelines can reasonably be divided into two different phases. Phase one encompasses protocol and questionnaire development and submission of the protocol and questionnaire to the FDA, and phase two encompasses data collection and management, analysis, and reporting.

Timelines for phase one activities are fairly standardized. Typically, protocol and questionnaire development takes 6 to 8 weeks and cognitive testing takes an additional 10 to 12 weeks. For the 18-month assessments, protocols and questionnaires must be submitted to the FDA for review. The FDA has 90 days to review and comment on the materials.

Timelines for phase two vary depending on the feedback received from the FDA and whether additional changes and testing of the questionnaire are required as well as the recruitment and data collection method employed (e.g., site-based vs. pharmacy network) and the availability of patients (e.g., asthma patients vs. transplant patients). Allowing 2 to 6 months for data collection and an additional 3 months for analysis and reporting of results is reasonable.

Therefore, the total timeline ranges from 12 to 18 months on average.

**d. Please comment on appropriate incentives for patients and health care providers to complete surveys.**

Offering incentives to patients and health care providers who complete surveys facilitates increased response rates. Incentives may vary depending on the recruitment method and burden on the participant and should be commensurate with the level of effort required. The incentive amount is generally reviewed by an institutional review board to assure that it will not be considered coercive. For example, a typical incentive for patients to complete a 15-minute survey is a \$25 gift card to the patient's pharmacy. Based on our experience, payments to physicians may range from \$75 to \$100 for a 15-minute questionnaire, depending on the physician specialty. Some legal

requirements forbid pharmaceutical companies from paying physicians for completing a survey, which significantly diminishes the response rate.

## Question 5—Alternatives

**5. Given the issues with surveys that we have observed, what are the alternatives to knowledge surveys to assess the effectiveness of the educational elements of the REMS? If any, what are the advantages and disadvantages of the alternatives?**

One alternative approach to directly assessing physician knowledge is to assess physicians' behavior through analysis of health insurance claims or electronic medical record databases. Although such approaches do not directly measure physician knowledge, they measure the effectiveness of the educational elements of the REMS in that they can show whether the REMS has produced the desired effect (e.g., reduced the at-risk population receiving a medication).

For example, one might use this approach to assess whether physicians are prescribing pharmaceutical or biological products in accordance with elements in the education program. These sources provide data on real-world clinical practice in large populations. By using these data retrospectively, rather than collecting knowledge data from health care providers, the data capture the results of physician knowledge and understanding without being influenced by the physician's awareness of the prospective data collection process.

The data fields generally available to researchers who have obtained appropriate permissions are patient diagnoses (reason for health encounter), procedures applied (e.g., vaccinations, radiographs), medications dispensed (in claims) or prescribed (in electronic medical record databases), and the dates of occurrence of each. Such data are useful for assessing whether patients have diagnoses for which specific pharmaceuticals or biologicals are contraindicated or not recommended<sup>30</sup> and may provide an indirect measure of whether dosages or durations of use are inappropriate or inconsistent with recommendations and labeling.

Although this approach is likely to be more representative of real-world practice than a survey approach, the main disadvantage is that some behaviors of interest may not be measurable in such data sources. For example, the disposition of any medication after it is prescribed or dispensed cannot be examined in such data sources, and direct evidence of abuse or diversion or specific doses used on a daily basis cannot be assessed. Despite known limitations of these data sources, however, they provide an objective measure of behavior, which is a useful complement to knowledge surveys. Our experience suggests that the combination of knowledge surveys and drug utilization studies is increasingly being requested in risk minimization programs in the European Union.

## Question 6—Design Considerations

### 6. What are alternative methods to assess behavior, burden, and access for a REMS? If any, what are the advantages and disadvantages of the alternatives?

As described in our response to Question 5, health claims and electronic medical record databases provide large, relatively more representative data sources (compared with voluntary surveys) on the process and results of clinical decision making. Such data sources, although unlikely to be used to assess burden or access, can be used to assess the behaviors associated with some specific aspects of REMS. In particular, the impact of a REMS or other regulations or recommendations may be evaluated if the REMS, regulations, or recommendations change over time. In such a situation, the change in patient characteristics or physician or clinic type following development of the new REMS, regulations, or recommendations can provide an indirect measure of the regulatory changes.

For two examples of such assessments that followed regulatory actions, see Weatherby and colleagues<sup>31</sup> and Shrank and colleagues.<sup>32</sup> In the Weatherby and colleagues study,<sup>31</sup> after the FDA issued a “Dear Doctor” letter about contraindicated medications to be avoided with cisapride, the concomitant prescribing of cisapride and contraindicated medications was evaluated among health claims data. In the Shrank and colleagues study,<sup>32</sup> after an FDA communication expressing concerns about the

efficacy of ezetimibe, the researchers identified more than 800,000 users of this medication from a national pharmacy database and evaluated the nonpersistence of ezetimibe and the appropriateness of a medication switch after discontinuation of ezetimibe.

The advantages of using this approach are that (1) the research is conducted retrospectively and independently of the process that generates the data (i.e., insurance claims) and (2) the sample size is large and likely to be more representative than with alternative approaches. The primary disadvantage is that some variables of interest are not available through claims data.

---

## Discussion

According to the most recent Prescription Drug User Fee Act V Reauthorization,<sup>33</sup> the FDA has the following performance goals:

- By the end of fiscal year (FY) 2013, FDA will
  - develop and issue guidance on criteria for requiring a REMS.
  - hold one or more public meetings to obtain stakeholder input on standardizing REMS to reduce the burden on the health care system.
  - initiate one or more public workshops on methodologies for assessing REMS.
- By the first quarter of FY 2014, FDA will
  - issue a report of its findings (related to standardization) and identify at least one priority project in each of the following areas, including a work plan for project completion:
    - pharmacy systems
    - prescriber education
    - providing benefit-risk information to patients
    - practice settings.
- By the end of FY 2014, FDA will issue guidance on methodologies for assessing REMS.

The FDA has established several working groups to assure the achievement of these performance goals. We expect that the product of these groups, along with the discussions held during the June 7, 2012, workshop and written feedback provided in response to FDA questions, will help shape the 2014 guidance

on methodologies for assessing REMS. In June 2013, the European Medicines Agency (EMA) issued draft guidance on the evaluation of risk minimization measures<sup>34</sup> and invited public comment. Although US REMS programs and EU risk minimization measures may be highly customized for local health care systems and laws, the principles and methods for

evaluation are similar. Therefore, we can expect that the FDA and EMA recommendations will cover many of the same concepts addressed here. We hope the guidance documents will improve research practices in REMS evaluations, which ultimately should lead to more effective risk minimization programs.

---

## References

1. US Department of Health and Human Services. Food and Drug Administration Center for Drug Evaluation and Research (CDER) Center for Biologics Evaluation and Research (CBER). Guidance for industry. Development and use of risk minimization action plans [Internet]. 2005 March [cited 2013 August 28]. Available from: <http://www.fda.gov/downloads/RegulatoryInformation/Guidances/UCM126830.pdf>
2. Food and Drug Administration. Risk Evaluation and Mitigation Strategy (REMS) Assessments: Social Science Methodologies to Assess Goals Related to Knowledge: public workshop [Internet]. 2012 June [cited 2013 August 28]. Available from: <http://www.fda.gov/Drugs/NewsEvents/ucm292337.htm>
3. US Department of Health and Human Services. Food and Drug Administration. Docket No. FDA-2012-N-0408. Risk evaluation and mitigation strategy assessments: social science methodologies to assess goals related to knowledge. Issue paper. 2012 June 7. [cited 2013 August 28]. Available from: <http://www.fda.gov/downloads/Drugs/NewsEvents/UCM301966.pdf>
4. Miller DP, Bennett L, Hollis KA, Tennis P, Cook SF, Andrews E. A patient follow-up survey programme for alosetron: assessing compliance to and effectiveness of the risk management programme. *Alimentary Pharmacol Therap*. 2006;24(5):869-78.
5. Tennis P, Andrews E, Hickman P, Miller D, Hollis K, Cook S. The relationship between dosing of alosetron and discontinuation patterns reported by patients participating in a follow-up programme. *Aliment Pharmacol Ther*. 2007;25: 317-22.
6. Hollis K, Calingaert B, Price M, Andrews EB, Moy L, Kooijmans M, et al. Tysabri TOUCH prescriber program risk management program (RiskMAP) surveys. *Pharmacoepidemiol Drug Saf*. 2010;19(Suppl 1):S59-60.
7. Hollis K, Zografos L, Price M, Gilsenan A, Andrews EB, Calingaert B. A comparison of factors influencing patient knowledge: results across selected REMS surveys. *Pharmacoepidemiol Drug Saf*. 2011;20(Suppl 1): S213-4.
8. Gilsenan A, Hollis K, Zografos L, Calingaert B, Andrews EB. Examination of respondent bias and patient characteristics in multiple REMS assessment surveys. *Pharmacoepidemiol Drug Saf*. 2011;20(Suppl 1):S208.
9. Andrews E, DiBenedetti D, Forsyth B, Gilsenan A, Hollis K, Lewis S, et al. Risk evaluation and mitigation strategy assessments: social science methodologies to assess goals related to knowledge. Response to Docket FDA-2012-N-0408 [Internet]. 2012 [cited 2013 August 28]. Available from: <http://www.regulations.gov/#!documentDetail;D=FDA-2012-N-0408-0024>
10. Mainous AG 3rd, Diaz VA, Saxena S, Baker R, Everett CJ, Koopman RJ, et al. Diabetes management in the USA and England: comparative analysis of national surveys. *J R Soc Med*. 2006 Sep;99(9):463-9.
11. McDonald M, Hertz RP, Unger AN, Lustik MB. Prevalence, awareness, and management of hypertension, dyslipidemia, and diabetes among United States adults aged 65 and older. *J Gerontol A Biol Sci Med Sci*. 2009 Feb;64(2):256-63.

12. Mockenhaupt M, Viboud C, Dunant A, Naldi L, Halevy S, Bouwes Bavinck JN, et al. Stevens-Johnson syndrome and toxic epidermal necrolysis: assessment of medication risks with emphasis on recently marketed drugs. The EuroSCAR-study. *J Invest Dermatol*. 2008 Jan;128(1):35-44.
13. American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education. *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association; 1999.
14. American Psychological Association. *Updated Standards for Educational and Psychological Testing* is released for public comment [Internet]. 2011. [cited 2013 August 28]. Available from: <http://www.apa.org/science/about/psa/2011/01/testing.aspx>
15. Mislevy RJ, Knowles KT. *Performance assessments for adult education: exploring the measurement issues*. Washington, DC: National Academy Press; 2002.
16. Ebel RL. Proposed solutions to two problems of test construction. *J Educ Meas*. 1982;19:267-78.
17. Crocker L, Algina J. *Introduction to classical and modern test theory*. Belmont, CA: Wadsworth; 2006.
18. Shuman H, Presser S. *Questions and answers in attitude surveys: experiments in question form, wording and context*. New York: Academic Press; 1996.
19. Beatty P, Herrmann D. A framework for evaluating “don’t know” responses in surveys. *Proceedings of the Section on Survey Research Methods*. Washington DC: American Statistical Association; 1995. p. 1005-10.
20. Beatty P, Herrmann D, Puskar C, Kerwin J. “Don’t know” response in surveys: is what I know what you want to know and do I want you to know it? *Memory*. 1998 Jul;6(4):407-26.
21. Smyth J, Dillman DA, Christian LM, Stern MJ. Comparing check-all and forced-choice question formats in web surveys. *Public Opin Quart*. 2006;70(1):66-77.
22. Malhotra N. Order effects in complex and simple tasks. *Public Opin Quart*. 2009;73(1):180-98.
23. Groves RM, Fowler FJ, Couper MP, Lepkowski JM, Singer E, Tourangeau R. *Survey Methodology*. 2nd ed. New Jersey: John Wiley and Sons, Inc.; 2009. Section 5.3, Effects of different data collection methods on survey errors. p. 160-181.
24. Willis GB. *Cognitive interviewing: a tool for improving questionnaire design*. Thousand Oaks, CA: Sage; 2005.
25. Sudman S, Bradburn NM, Swartz N. *Thinking about answers: the application of cognitive processes to survey methodology*. San Francisco: Jossey-Bass, Inc.; 1996.
26. Willis GB, Royston P, Bercini D. The use of verbal report methods in the development and testing of survey questionnaires. *Appl Cogn Psychol*. 1991;5:251-67.
27. US Department of Health and Human Services. *Guidance for industry. Patient-reported outcome measures: use in medical product development to support labeling claims* [Internet]. 2009 December [cited 2013 August 28]. Available from: <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM193282.pdf>
28. DiBenedetti DB, Price MA, Andrews EB. Cognitive interviewing in risk minimization survey development: patient and healthcare professional surveys. *Expert Rev Clin Pharmacol*. 2013;6(4):369-73.
29. DeMuro CJ, Lewis SA, DiBenedetti DB, Price MA, Fehnel SE. Successful implementation of cognitive interviews in special populations. *Expert Rev Pharmacoecon Outcomes Res*. 2012 Apr;12(2):181-7.
30. Tennis P, Toback SL, Andrews E, McQuay LJ, Ambrose CS. A postmarketing evaluation of the frequency of use and safety of live attenuated influenza vaccine use in nonrecommended children younger than 5 years. *Vaccine*. 2011 Jul 12;29(31):4947-52.
31. Weatherby LB, Nordstrom BL, Fife D, Walker AM. The impact of wording in “Dear doctor” letters and in black box labels. *Clin Pharmacol Ther*. 2002 Dec;72(6):735-42.

32. Shrank WH, Choudhry NK, Tong A, Myers J, Fischer MA, Swanton K, et al. Warnings without guidance: patient responses to an FDA warning about ezetimibe. *Med Care*. 2012 Jun;50(6): 479-84.
33. Food and Drug Administration. PDUFA reauthorization performance goals and procedures: fiscal years 2013-2017 [Internet]. [cited 2013 August 28]. Available from: <http://www.fda.gov/downloads/ForIndustry/UserFees/PrescriptionDrugUserFee/UCM270412.pdf>
34. European Medicines Agency. Guideline on good pharmacovigilance practices (GVP). Module XVI – Risk minimisation measures: selection of tools and effectiveness indicators' (EMA/204715/2012) [Internet]. 2013 June 6 [cited 2013 August 21]. Available from: [http://www.ema.europa.eu/docs/en\\_GB/document\\_library/Scientific\\_guideline/2013/06/WC500144010.pdf](http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2013/06/WC500144010.pdf)



## **Acknowledgments**

Our thanks for the reviewing and editing provided by Martin Meyer, Craig Hollingsworth, Nanthini Ganapathi, Charlotte Scheper, and Patricia Smith.

RTI International is an independent, nonprofit research organization dedicated to improving the human condition by turning knowledge into practice. RTI offers innovative research and technical solutions to governments and businesses worldwide in the areas of health and pharmaceuticals, education and training, surveys and statistics, advanced technology, international development, economic and social policy, energy and the environment, and laboratory and chemistry services.

The RTI Press complements traditional publication outlets by providing another way for RTI researchers to disseminate the knowledge they generate. This PDF document is offered as a public service of RTI International.