Validation of Metrics—A Comparative Analysis of Predictive- and Criterion-Based Validation Tests in a Qualitative Study

Erin Fordyce[*], Michael J. Stern[†], Sabrina Avripas Bauroth[‡], Catherine Vladutlu[**]

Survey Practice

A primary concern for researchers, when collecting self-reported data from respondents, is supporting the accuracy and reliability of the data collected. Measurement error can be introduced by questions asking for specific, factual information that may be difficult for respondents to recall. It can further increase due to the underreporting of sensitive information (Blattman et al., 2014) often attributed to social desirability. Researchers continue to struggle with finding cost-efficient and burden-free approaches to validate self-reported information. The National Survey of Children's Health (NSCH) collects data on the physical and emotional health of children including current and pre-existing conditions (or disabilities). As part of the NSCH Redesign study, 64 cognitive interviews were conducted with predictive and criterion-based validation tests to validate household screener items as well as items related to medical diagnosis and health insurance status. To assess predictive validity, a test-retest approach was used, whereby a subset of respondents were re-administered items from the household screener and main questionnaire two weeks subsequent to their initial interview. A separate, criterion-based validity test was conducted by asking the remaining respondents to provide documentation to validate the household screener items as well as items related to medical diagnosis and health insurance status. Respondents who did not provide documentation were asked permission to contact the child's primary care provider. This paper addresses important issues surrounding instances where a researcher is asking respondents to provide documentation to validate reported information. For instance, we discuss the impact of requesting various types of documentation on respondent burden and the advantages and disadvantages of requesting documentation versus other means of validating information. In addition, we discuss the effectiveness of conducting re-tests in identifying potential measurement error.

INTRODUCTION

A question that plagues survey researchers is whether the self-reported data they collect is accurate. There are many sources of error that can lead to the misreporting of information and researchers do their best to mitigate error through design, testing, and so forth. As shown in a meta-analysis by Brener

* **Institution:** NORC at the University of Chicago
† **Institution:** NORC at the University of Chicago
‡ **Institution:** NORC at the University of Chicago
** **Institution:** Maternal and Child Health Bureau; Health Resources and Services Administration

et al. (2003), there are numerous cognitive and situational factors that affect respondents when completing a self-administered questionnaire. Cognitive factors include comprehension and the ability to recall information whereas situational factors tend to involve fear of reporting or feeling the need to adhere to socially desirable behaviors.

Researchers have used a variety of validation methods to assess the accuracy of self-reported information. For instance, studies have used administrative data to validate self-reported health coverage (Davern, Call, Ziegenfuss, et al. 2008) as well as reported chronic diseases such as epilepsy (Brooks et al. 2012) and diabetes (Comino et al. 2013). While administrative records serve well for validation of conditions or enrollment in programs, other behaviors are more complicated to validate. For example, Wong et al. (2012) validated self-reports of smoking status through comparing self-reports of smoking status from a survey with urinary cotinine concentrations collected from respondents.

Validity is critical when collecting survey data on health care issues because it is imperative in terms of developing an understanding of patient needs, specifically for identifying gaps in coverage and services provided. Self-reported health data, however, is often questioned regarding its accuracy (Brener, Billy, and Grady 2003) including under or over-estimation (Davern, Call, Ziegenfuss, et al. 2008) due to the sensitive nature of questions asked and social desirability. Still, aside from using administrative records and more invasive methods, questions remain regarding what validation methods are effective and whether one method is sufficient.

Even before a method for validation can be selected, researchers must assess the cost and resource implications for their particular study. This is a challenge for many researchers as they attempt to continue gathering accurate data with available funds. As Podsakoff et al. (2012) mention in regards to obtaining measures of predictor and criterion variables from different sources, the technique may "require more time, effort, and/or cost than the researcher can afford" (p. 549). Therefore, the use of validation measures is a highly desired aspect for research studies if, and when, the resources suffice.

In this paper, we seek to address this issue by examining two validation methods that were implemented to assess the accuracy of self-reported information from a national health study. The first method assessed predictive validity where a test-retest protocol was used. The second method involved a criterion-based test where respondents were asked to provide documentation of enrollment in a program or proof of their child's medical condition. Ultimately, the research questions we sought to answer were:

1. Whether respondents consistently report factual, sensitive information when retested?

2. Whether test-retest and criterion-based metrics are effective means for identifying potential measurement error?

3. Whether the advantages of implementing validation metrics outweigh the disadvantages?

THE STUDY

The validation methods were implemented as part of the National Survey of Children's Health (NSCH) Redesign Study; research was supported and directed by the Maternal and Child Health Bureau (MCHB) in the Health Resources and Services Administration (HRSA), an agency of the U.S. Department of Health and Human Services (HHS). The purpose of the redesign was to assess the impacts on transitioning from a single mode telephone survey to a multimode web and mail survey. The redesign presented a unique opportunity to evaluate the possibility of measurement error which is often introduced by asking for specific, factual information from respondents. There is also the potential for bias being introduced when asking respondents about behaviors perceived as socially undesirable. The NSCH survey asks respondents to report on several specific health-related questions regarding their child, such as height and weight, current and past diagnosed conditions, and health insurance coverage.

With the NSCH transitioning from phone to web and mail modes, respondents would be trusted to self-report information without a phone interviewer available to answer questions or note any indications that the respondent might not be accurately recalling the information. Therefore, it was important to implement a method to assess the potential for measurement error and validate the information provided prior to the start of data collection. This validation process would allow researchers to better identify problem questions that are likely to elicit incorrect responses or that ask for information too difficult for respondents to recall.

METHODOLOGY

Sixty-four cognitive interviews were conducted between September and November of 2014. The purpose of these interviews was to conduct both cognitive and usability testing for the revised instrument.

To assess predictive validity, a test-retest approach was used, whereby 31 respondents were re-administered items from the household screener and main questionnaire 2 weeks subsequent to their initial interview. Regarding the criterion-based test, we asked 14 respondents to provide documentation to validate the household screener items and items related to medical diagnosis and insurance status. Respondents who did not provide documentation were asked permission to contact the child's primary care provider.

RESULTS

TEST-RETEST

Respondents completed the entire questionnaire during the cognitive

interviews. A subset of the respondents were then retested 1–2 weeks later on a selection of the measures determined to be the most susceptible to measurement bias. This included questions related to health conditions, health insurance, respondent age, education, and household income. Respondents completed the interview over the phone and received $30 for completing the retest interview.

A total of 26 retests were completed, with no clear differences in responses measured between the initial and subsequent retest. However, an initial review of the data revealed the potential for mode effects influencing the responses to the income questions. During the interview, respondents were asked to provide an exact income amount. If this was not known or they refused, a follow up question requested that they provide an income range. National Opinion Research Center (NORC) found that respondents often switched the question to which they responded, depending on the mode, so there was an increased likelihood that self-reported income responses were not an exact match during the test and retest (Table 1). However, when the exact income amounts were converted to the ranges provided in the income questions, the consistency of the response options increased (Table 2).

Table 1   Exact income questions compared to range income questions.

Strict match vs. non-match		
Questionnaire	Matches	Non-matches
0–5	7	3
6–11	3	3
12–17	2	7
Total	12	13

Table 2   Exact income answers converted to range income question.

Match vs. non-match		
Questionnaire	Matches	Non-matches
0–5	7	3
6–11	4	2
12–17	7	2
Total	18	7

In addition, there were differences in the reported severity of health conditions between the test and retest (Table 3). Respondents were asked to rate the severity of health conditions for their child as mild, moderate, or severe, and there were a number of non-matches between the initial and follow up questionnaires. It is possible that the child's condition either worsened or improved between the initial cognitive interview and the retest.

Table 3   Reported severity of health conditions.

Would you describe the condition as Mild/Moderate/Severe?		
	Matches	Non-matches
Total	8	6

CRITERION

Each respondent was asked, during the initial phone screening, if he or she currently had insurance coverage. Those respondents who answered "yes" were then asked to bring proof of insurance to the scheduled cognitive interview. A majority of the respondents provided proof of insurance (as shown in Table 4 below).

Table 4   Condition verification.

Conditions	(N=14) # of respondents	Health insurance	(N=54) # of respondents
Confirmed*	2	No insurance	11
Refused	4	Proof of insurance	36
Signed consent form	8	No proof of insurance	7
Total	14	Total	54

*Both respondents who provided documentation to verify the child's condition brought in prescription bottles.

Respondents were also asked, during the initial phone screening, whether any children living in the household had a special health care need. These respondents were asked to bring documentation to verify the condition with them to the cognitive interview. Documentation could include prescription bottles, a doctor's note, etc. Those respondents, who did not provide documentation at the time of the interview, were then asked to sign a provider consent form allowing NORC staff to contact the child's primary care provider. NORC staff then followed up with the providers to have them sign a form to verify the condition(s). As shown in Table 4 below, a majority of respondents elected to sign the provider consent form. NORC contacted eight providers to get a confirmation of diagnosis or treatment for all medical conditions that were reported by the respondent. NORC staff faxed to the providers information about the study, the signed consent form, and a form that could be returned with the necessary information. Most providers required a follow-up call from NORC to collect the information.

Of the contacted providers, five confirmed that the patient had been diagnosed or treated for the conditions they reported. Two providers had a record of the child but did not have a record of the indicated conditions. The final provider did not have a record of ever treating the respondent's child. Table 5 below shows the results.

Table 5   Provider follow up.

Provider follow up for signed consent forms (N=8)

	# of respondents
Physician confirmed diagnosis	5
Did not confirm diagnosis	1
Physician reported diagnosing other/related condition	2
Total	8

DISCUSSION

There were a few limitations to the validation methods implemented for the redesign. First, a small number of cases were assigned to the criterion-based group. Several of these respondents reported extenuating circumstances for why they could not provide documentation for the child's condition(s). For instance, two respondents were fathers and, due to custody disputes, did not feel comfortable signing any documentation regarding their child's health records. Further, certain conditions (e.g., Down syndrome) can have several underlying conditions (e.g., language disorder) which confused respondents as to whether they should answer yes for both. This was evidenced in the criterion-based validation. For example, a provider reported diagnosing the child with a condition not specifically reported by the respondent (spina bifida); however, the respondent did report an underlying condition associated with the diagnosed condition (migraines/severe headaches).

Findings from the cognitive interview process and the validation methods used provided valuable insight in response to the research questions posed:

Do respondents consistently report factual, sensitive information when retested? Responses provided in the retest were shown to be consistent with those provided in the original survey. However, it is recommended that researchers be cognizant of the potential differences in responses due to mode effects and other factors. Questions should be carefully formatted across modes to avoid these effects. Additional analysis may be required after data collection, as was done with the income questions from this study. Also, respondents may interpret questions differently, and therefore, it should be evident what information you are requesting. Instruction text and definitions should be clear to the respondent to improve consistency in interpretation. For example, after the cognitive interview process, it was decided that the instruction text "Has a doctor or other health care provider ever told you that your child has…" would be repeated throughout the series of questions asking about diagnosed conditions. The purpose was to remind respondents that they should only report a condition if a doctor or health care provider made the diagnosis. This approach would help ensure that respondents did not include other instances where perhaps a sports coach or school nurse suggested that the child may have a condition. We were looking for medically confirmed diagnoses only.

Are test/retest and criterion-based metrics effective means for identifying potential measurement error? The test/retest and criterion metrics were found to be efficient and effective for this study. However, it is suggested that researchers experiment with several validation methods to improve efficiency in data collection, allow sufficient time for Institutional Review Board (IRB) and (Educational Records Bureau (ERB) reviews, and provide several options for respondents to provide documentation for the validation criteria when possible.

Do the advantages of implementing validation metrics outweigh the disadvantages? This is highly dependent on the process used and the resources available. For the redesign study, the research team planned well in advance for the validation process. It is not something that can be implemented at the last minute because of the planning required. Materials, including consent forms and retest questionnaires, have to be prepared as well as the IRB and ERB forms. For this study, staff was needed to follow up with respondents for the retest interviews as well as contacting the health care providers should they not respond to initial contact attempts. Another issue to note is the potential impact on response rates. Are respondents going to participate if they are asked to bring documentation to the interview or be recontacted at a later date for a follow up interview? A concern for the redesign was whether respondents would be willing to sign a consent form for researchers to contact the child's health care provider.

CONCLUSION

Collecting self-reported data poses several challenges for researchers, most notably the potential for measurement error. But with advance planning and meticulous survey design, researchers can better minimize this error. The validation methods (test-retest and criterion-based) used for the NSCH redesign proved to be efficient and effective at identifying measurement error. Moving forward, researchers will have to utilize the most effective validation strategy that meets the needs of their particular study. This may require the use of more than one validation method and additional resources which is a compromise research teams will have to discuss and weigh in the early planning stages of their project(s).

ACKNOWLEDGEMENTS

REFERENCES

Blattman, C., J. Jamison, T. Koroknay-Palicz, K. Rodrigues, and M. Sheridan. 2016. "Measuring the Measurement Error: A Method to Qualitatively Validate Sensitive Survey Data." *Journal of Development Economics* 120: 99–112.

Brener, N.D., J. Billy, and W.R. Grady. 2003. "Assessment of Factors Affecting the Validity of Self-Reported Health-Risk Behavior among Adolescents: Evidence from the Scientific Literature." *Journal of Adolescent Health* 33 (6): 436–57.

Brooks, D.R., R. Avetisyan, K.M. Jarrett, A. Hanchate, G.D. Shapiro, M.J. Pugh, D.R. Berlowitz, D. Thurman, G. Montouris, and L.E. Kazis. 2012. "Validation of Self-Reported Epilepsy for Purposes of Community Surveillance." *Epilepsy & Behavior* 23 (1): 57–63.

Comino, E.J., D.T. Tran, M. Haas, J. Flack, B. Jalaludin, L. Jorm, and M.F. Harris. 2013. "Validating Self-Report of Diabetes Use by Participants in the 45 and up Study: A Record Linkage Study." *BMC Health Services Research* 13 (1): 481.

Davern, M., K.T. Call, J. Ziegenfuss, et al. 2008. "Validating Health Insurance Coverage Survey Estimates: A Comparison between Self-Reported Coverage and Administrative Data Records." *Public Opinion Quarterly* 72 (2): 241–59.

Podsakoff, P.M., S.B. MacKenzie, and N.P. Podsakoff. 2012. "Sources of Method Bias in Social Science Research and Recommendations on How to Control It." *Annual Review of Psychology* 63: 539–69.

Wong, S.L., M. Shields, S. Leatherdale, E. Malaison, and D. Hammond. 2012. "Assessment of Validity of Self-Reported Smoking Status." *Statistics Canada* 23 (1): 47–53.