Scoring the Change Process Capability Questionnaire Strategies Items:

Current Plan

National Evaluation Team

May 1, 2017

Introduction

The RFA from AHRQ for EvidenceNOW:

`http://grants.nih.gov/grants/guide/rfa-files/RFA-HS-14-008.html`

described the required measures of practice capacity. One of those measures is the Change Process Capability Questionnaire (CPCQ).

The measurement of CPCQ was described in the RFA as: "The 32-item questionnaire based on the work of Solberg and colleagues can be found in AHRQ's Practice Facilitation Handbook, Module 6 Appendix". These items were identified by a group of successful practice improvement leaders through an iterative Delphi process.

The citation provided in the RFA for the CPCQ measure is:

`https://www.ahrq.gov/professionals/prevention-chronic-care/improve/system/pfhandbook/mod6appendix.html`

The CPCQ is a function of a collection of survey items administered to a single member of the practice. Solberg et al. studied the responses from many members of actual practices and found that the medical leader of a practice provided answers that were as accurate (compared to an audit) as any other aggregation of answers. The original CPCQ survey was a collection of 32-items, but for EvidenceNOW, we determined through the harmonization phase to collect the 14 items related to strategies and the single item related to priorities. This reduction was done for 2 reasons:

1. To reduce response burden
2. To focus on the part of the CPCQ that seemed best associated with improvement and had the least overlap with the Adaptive Reserve

The stem of the CPCQ strategies questions begin with "We would like to learn about the strategies that your practice uses to improve cardiovascular preventive care (e.g., prescribing aspirin for patients at risk for ischemic vascular disease, providing tobacco cessation services for smokers, appropriately managing hypertension, and prescribing statins for high risk patients). These questions should be completed by one senior member of the practice who has good insights into the clinical operations of the practice, such as a lead clinician or an office manager".

The responses to the 14 CPCQ strategies items are on a scale from 1 ("strongly disagree") to 5 ("strongly agree"). There is also an option of "not applicable" (NA) for each item and that value is denoted as 8 in the EvidenceNOW practice survey codebook.

After data cleaning and merging, the question then becomes: **How do you compute a practice-level assessment of CPCQ strategies?**

From *Solberg, et al. (2008) "Measuring an organization's ability to manage change: the change process capability questionnaire and its use for improving depression care." American Journal of Medical Quality 23(3): 193-200*, the description of how to calculate practice-level CPCQ scores is as follows:

"Each response to the CPCQ survey was recoded so that 2 = strongly positive, 1 = positive, 0 =neutral, -1 = negative, and -2 = strongly negative.... All 14 strategies were left in 1 group... "

From this article, the 14 strategies items were summed to create a strategies score. This is evident in Table 4 where Solberg et al. describe the score range from -28 to +28.

We would also like to note that there is a body of literature that describes a different approach to CPCQ scoring. For example, in the following studies, CPCQ is scored in the following ways:

1. *Scholle, et al. "Support and strategies for change among small patient-centered medical home practices." The Annals of Family Medicine 11.Suppl 1 (2013): S6-S13.*

    - The strategies scale includes 18 items. The composite is a sum of items rated "yes, worked well" (1 point), "yes, did not work well" (0.5 point), and "no" (0 points).
    - From Table 4: Range is 0 to 18, with higher scores indicating greater use of strategies.

2. *Morton, et al. "Health IT Enabled Care Coordination: A National Survey of Patient-Centered Medical Home Clinicians." The Annals of Family Medicine 13.3 (2015): 250-256.*

    - We used the Strategies scale of the Change Process Capability Questionnaire (CPCQ) of Solberg et al to assess practices' use of techniques such as rapid-cycle testing and involvement of staff and patients in quality improvement. Each of the 17 scale items was rated "yes, worked well" (1 point), "yes, did not work well" (1/2 point), and "no" (0 points). We summed the results to get a score on a scale of 0 to 17 points; higher scores indicate greater capability to undertake change

3. *Solberg, et al. "Organizational factors and change strategies associated with medical home transformation." American Journal of Medical Quality 30.4 (2015): 337-344.*

    - The second component assesses strategies that have been used to implement improved depression care, and contains 16 items answered as "Yes, worked well" (scored 1); "Yes, but did not work well" (scored 0.5); or "No, not used" (scored 0).

Based on these articles, there is a bit of variability on methods for computing the CPCQ strategies score due to differing number of items and different scales used.

As such, the ESCALATES approach for estimating practice-level CPCQ strategies follows the approach from Solberg et al. (2008) because the EvidenceNOW items match the scaling and number of items for the CPCQ used in Solberg et al. (2008). Below, we will define the steps for scoring the CPCQ.

ESCALATES CPCQ Strategies Scoring Method

1. For each CPCQ strategies item, convert response values ranging from 1-5 to a scale ranging from -2 to 2, with 0 representing 'neutral' (reverse code when appropriate).

2. For each practice, sum the 14 CPCQ strategies items to compute a practice-level score from -28 to +28.

3. If there is a missing value for any of the 14 items, set the practice-level CPCQ strategies score to missing.

4. If there is a "NA" value for any of the 14 items, set the practice-level CPCQ strategies score to missing.

NOTE: Missing values and 'Not Applicable' values are both treated as missing. For now, they result in a missing CPCQ score for the practice. However, we outline our approach for handling these missing values in the following section. In short, we plan to implement a multiple imputation procedure to fully utilize observable data and impute missing and 'NA' values.

Approach for handling Missing CPCQ items

Due to how the CPCQ is constructed, the score is highly sensitive to missing data on any of the 14 items that represent the CPCQ strategies construct. If either of the 14 items are missing, there is no clear guidance from the original Solberg study on how to score the CPCQ.

Methods for Missing Data in Likert Scales (Attitude Scales)

- Missing data arises for many reasons, likely nullifying the possibility of missing completely at random (MCAR).

- Many methods are used, the most popular of which are multiple imputation and regression methods.

- Some data reduction techniques are used to impute overall scale scores rather than item scores, but those are discouraged.

- Multiple Imputation seems to work well under MCAR with up to 30% missing data and Missing at Random (MAR) at between 10% and 30% missing data.

- Latent variable methods show promise for gathering more information regarding the mechanism behind item non-response in attitude questionnaires.

ESCALATES approach for handling missing CPCQ items

Our overall approach is to perform multiple imputation using chained equations (MICE) on CPCQ items. We plan to perform 5-10 multiply imputed data sets where missing CPCQ items will be imputed. Items will then be summed to create a practice-level CPCQ strategies score for each multiple imputed data set. Lastly, we will apply Rubin's rules across multiply imputed data sets to obtain parameter estimates and their corresponding standard errors.

When CPCQ is the outcome of interest:

- Under MAR, there are generally no benefits to impute the outcome, and for a low number of imputations the results may even be somewhat more variable because of simulation error. There is an important exception to this. If we have access to an auxiliary complete variable that is not part of the model and that is highly correlated with the outcome, imputation can be considerably more efficient than complete case analysis, resulting in more precise estimates and shorter confidence intervals.

- Because the CPCQ items are highly correlated to each other, we will only use CPCQ items to impute the missing items.

- If more than 30% of the items for a single practice is missing, we will exclude that practice from the multiple imputation procedure

When CPCQ is a predictor:

- Include a comprehensive set of predictors (i.e. all variables in the estimation model).

Literature Search

The following literature were found via Google Scholar and searches for "Likert Scale Missing Data", "Likert Scale non-response", "Attitude Scale Missing Data", and "Attitude Scale non-response". Six published articles from Tourism Management, Journal of the Royal Statistical Society, BMC Medical Research Methodology,

Multivariate Behavioral Research, Journal of Clinical Epidemiology, and Journal of Modern Applied Statistical Methods were selected and short summaries are provided below. We used these 6 articles as our basis for how we should approach scoring the CPCQ when missing data is present. The following literature gives a view of some of the methods available to us and under what circumstances they perform acceptably well.

O'Muircheartaigh C and Moustaki I. Symmetric Pattern Models: a Latent Variable Approach to Item Non-Response in Attitude Scales. Journal of the Royal Statistical Society. 1999;162(2): 177-194.

Summary: The authors propose a novel approach to non-response in attitude scales that incorporates both interpolated responses to missing items in a survey scale and propensity for response. The method is symmetric in that it doesn't take any single item as a dependent variable, treating all survey items equally in the analysis and pattern based in that it takes into account response and non-response patterns.

Background and Objectives: Researchers are concerned with differential attitudes between respondents and non-respondents in attitude scales. The authors aim is to introduce a model that will take attitude into account when measuring propensity to respond and use it to estimate attitudes given non-response.

Methods: The attitude and response propensity factors are assumed to have independent standard normal distributions. Given the factors, response levels (agree to disagree) and non-response are conditionally independent. This model is a mixed model, as is required to handle non-response. With p measured items to analyze and a proportion of non-response for each of 1 to p items, create p pseudo items that are 0/1 indicators for whether the individual has given a response. For the analysis we use all 2p items, the original items provide information about attitude and the pseudo items provide information about response propensity.

Results: The method is applied to 3 datasets, a constructed dataset that shows the method in a simplified example, an ethnocentrism dataset that shows the ability to measure the impact of attitude on response propensity, and a British survey on social attitudes that shows how non-response can be associated with different underlying attitudes.

Conclusions: The method has the capacity to measure both the propensity to respond predict underlying attitudes.

Notable Findings: In the British Social Attitudes Survey, the authors treated both 'don't know' type non-response and skipped questions uniformly.

Shrive FM, Stuart H, Quan H, Ghali WA. Dealing With Missing Data in a Multi-Question Depression Scale: a Comparison of Imputation Methods. BMC Medical Research Methodology. 2006;6(57).

Background: Missing data is a particular issue in self-reported scale. The authors reviewed 6 different imputation techniques applied to the Zung Self-reported Depression Scale.

Methods: 1580 participants, 20 question scale with values from 1-4 for each. The sum is converted into a scale. Missing completely at random was simulated by random deletion of responses, missing at random and missing not at random were also simulated. Imputation methods compared were:

- Multiple imputation
- Single regression
- Individual mean
- Overall mean
- Participant's preceding response
- Random value

Results: At 10% missing, all methods except random selection produce $kappa > 0.80$. At 30% missing or unbalanced missing, MI maintained a high kappa statistic, and individual mean and single regression produced substantial agreement ($kappa > 0.70$).

Conclusions: Multiple imputation is a clear leader in accuracy for dealing with missing data, though even in adverse missing situations single regression and individual mean produced results that were acceptably accurate and potentially more interpretable than those produced through use of multiple imputation.

Notable Findings: Three methods remain reasonably accurate in the event that response is related to a differential attitude.

Ryan C and Garland R. The Use of a Specific Non-Response Option on Likert Type Scales. Tourism Management. 1999;20: 107-113.

Abstract: The authors argue the use in analyzing patterns in non-response may lead to useful information, justifying the inclusion of a non-response option in attitudinal questionnaires.

Background: Scales in attitude testing are well established as a methodology in social and behavioral research. Research from the 70s indicates that many non-responses are due to not having an opinion on an issue. This may differ from not wanting to respond to a question.

Objectives: Investigation of sources of non-response to items on a questionnaire.

Results: In a survey on customers' loyalty to a bank, questions with higher non-response had higher variances in response than questions with lower non-response with indications that non-response was due to a lack of knowledge or information pertaining to the question at hand. In a review of attitudes on tourism where respondents were grouped into 'Moderate Enthusiasts', 'Extreme Enthusiasts', and 'Cautious Supporters' a cluster analysis grouped the survey takers with higher non-response into the 'Cautious Supporters' category. The third survey was on New Zealander's attitudes towards Australia and the Northern Territory as a holiday destination. The survey was implemented with and without a no response option and was found to have little effect on the mean values of the questions but did have effects on the variance for individual items.

Conclusions: Non-response options may be valuable to allow for lack of knowledge or information on a subject in a survey and may in itself be useful for measuring patterns of non-response.

Notable Findings: One article reviewed (Goyder 1987) suggests that non-response patterns of 20-40% are unavoidable. Another article (Schuman and Presser 1981) suggests that giving non-response options increases non-standard responses by as much as 25%.

Eekhout I, Enders CK, Twisk JWR, de Boer MR, de Vet HCW, Heymans MW. Including Auxiliary Item Information in Longitudinal Data Analyses Improved Handling Missing Questionnaire Outcome Data. Journal of Clinical Epidemiology. 2015;68: 637-645.

Background: Previous studies suggest that missing values in questionnaires are best handled at item score levels.

Objectives: To show two methods for handling incomplete data in longitudinal studies. A latent growth model can be built with either the item scores or a summary of a parcel of available items as auxiliary variables.

Results: Inclusion of the item information improved precision of regression coefficients and standard errors over an analysis that did not include the item information.

Conclusions: Parcel summaries efficiently improve accuracy in a longitudinal survey setting (where the number of variables can approach the sample size). This can be used to complete impaired scores due to incomplete survey response.

Notable Findings: Parcel summaries as a form of data reduction simplify model estimation over including item information and are an efficient way of including item information as auxiliary variables in latent growth models.

Leite W and Beretvas NS. The Performance of Multiple Imputation for Likert-Type Items with Missing Data. 2010;9(1): 64-74.

Background: Multiple imputation has emerged as one of the preferred techniques for dealing with missing data; however, MI is often performed with the assumption that variables are multivariate normally distributed. This assumption is violated when using the Likert-type responses.

Objectives: To evaluate the performance of multiple imputation as a missing data technique on data derived from Likert-Scale surveys. Simulation was performed where MI assuming multivariate normality was used on data where Likert-type data points were missing.

Results: MI when applied to normally distributed data performed acceptably in MCAR and MAR simulations with 10% missing data, but only in MCAR simulations with 30% missing data, MI did not perform acceptably in any simulation with 50% missing data. MI performed the same when applied to non-normally distributed data, indicating that it is robust to violations of the normality assumption.

Conclusions: Whether MI is a suitable method for dealing with missing data is more dependent on the mechanism of missingness and prevalence of non-response than on the distribution of the response.

Notable Findings: These simulation results with unacceptable levels of bias at 30% missing at random data appears to contradict the results drawn from Shrive et al, 2006.

Mazza GL, Enders CK, Ruehlman LS. Addressing Item-Level Missing Data: A Comparison of Proration and Full Information Maximum Likelihood Estimation. Multivariate Behavioral Research. 2016;50(5): 504-519.

Background: It is common to prorate scale scores by replacing missing values with averages of the available items.

Objectives: To encourage researchers to use methods other than proration, and to introduce a Full Information Maximum Likelihood Estimation method for handling missing data at the item level.

Results: Using simulations, the authors conclude that item-level missing data handling increases power relative to scale-level missing data handling.

Conclusions: FIML methods that include all but one items as auxiliary variables performed roughly the same as item-level imputation, and considerably better than proration.

Notable Findings: N/A