American Journal of Biostatistics

Original Research Paper

Methods for Computing Missing Item Response in Psychometric Scale Construction

Ohidul Islam Siddiqui

Institute of Statistical Research and Training (ISRT), Dhaka University, Dhaka 1000, Bangladesh

Article history
Received: 31-08-2014
Revised: 21-02-2015
Accepted: 21-02-2015

Abstract: Therapeutic potential of a new antidepressant drug is evaluated frequently based on multi-item psychometric scales. The total score of a psychometric scale is calculated based on the responses of multiple-items, in which each item is scored on a likert scale. Missing responses in some of the items are inevitable and hence it is a problem in calculating the total score of a scale. Different approaches can be used to handle the missing item responses in constructing the total scores of a psychometric scale. One approach is that if a patient has missing responses in one or more items, his/her total score will be missing; another approach is that the missing item response will be imputed before calculating the scale total score. For the imputation, different methods can be used. Each of the methods has some drawbacks. This paper compares six methods, commonly used in imputing the missing item responses when there are missing responses at one or more items, but not missing more than 50% items of the scale. Simulation studies indicate that substituting the mean of the completed items of a scale for a given patient is generally the most desirable method for imputing both the random and non-random missing items in the psychometric scale construction.

Keyword: Imputation, Missing Data, Non-Response, Psychometric Scale

Introduction

In longitudinal clinical trials of antidepressant drug development, self-rating or clinician-rating multi-item psychometric scales are frequently used to evaluate treatment efficacy. For example, State-Trait Anxiety Inventory (STAI: 20 items; Spielberger *et al*., 1970), the Beck Anxiety Scale (BAS: 21 items; Beck *et al*., 1988), the Patient-Rated Anxiety Scale (PRAS; Sheehan, 1986) and the Self-Rating Anxiety Scale (SAS: 20 items; Zung 1971) are commonly used as self-rating scales in measuring the severity of panic disorder and agoraphobia. The Manifest Anxiety Scale (MAS: 38 items; Taylor, 1953), Hamilton Anxiety Scale (HAMA: 14 items; Hamilton and Guy, 1976) and Clinician Anxiety Scale (CAS: 6 items; Snaith *et al*., 1982) are frequently used as clinician-rating scales to measure anxiety. A frequent problem with data collection using such multi-item scales is missing data (Little and Rubin, 1987). Missing response for even a single item can create major problem in aggregating the scale total score and statistical computer packages may exclude patients if a single item is missing.

In psychometric scale construction, two types of missing data may be identified. First, there can be missing items within a scale, where a patient may have responded some of the items, but has failed to answer to the remaining items on the same scale and this situation is commonly described as ‘missing items’. Secondly, responses of all of the items for a given scale might be missing due to drop out of a patient from the study and it is described as ‘missing scale score’. Much attention has been given to handle such missing scale scores in analyzing longitudinal clinical trial data sets (Little, 1994; Rubin, 1987). However, there are little evidences of literature about methods of handling missing item responses in multi-item psychometric scales commonly used in clinical trials. A few papers point out the extent of missing items or how they handle missing items in their analyses. Therefore, this paper will only focus upon

the issue of handling missing items in analyzing psychometric scale total scores.

Potential bias may be induced due to non-random missing data. Patients who experience a negative impact on their lives due to treatment related toxicity, progressive of disease, are more likely to have missing items. Some patients may also inadvertently miss a response or they may choose not to answer a question for personal reasons. In presence of missing items, one can analyze the data set by (a) ignoring the missing items, (b) omitting the patients with missing items from the analysis, (c) omitting the patients from the particular analysis using the scale that contains the missing item, or (d) finding a desirable method to replace the missing item response with an estimate of what it might be. Each of the above choices in handling missing items has some drawbacks. For example, ignoring the missing items and summing the responses over the remaining items leads to an underestimate of the patients' total scores. Dropping the patients with missing data from the analyses reduce both the power and accuracy of the analyses (Madow *et al*., 1983).

Imputing missing items increases the number of analyzable patients, maximizes the representativeness of the result. In addition, if the missing items are not missing at random, strategies for handling missing items can be taken so that the bias in the result will be minimized. A research question is whether the available methods for handling missing item responses can reduce the bias potentially introduced by the missing data and which methods most accurately estimate individual total scores. In this study, six methods commonly used in clinical trials for handling missing item responses will be explored. These include case deletion and five methods of imputation. Finally, this paper will address how sensitive are the parameter estimates and their standard errors in using different imputing methods for imputing missing items when missing items are not missing at random.

Methods for Imputing Missing Items Commonly Adopted in Anti-Depression Clinical Trial Data Analysis

Method 1: Treat the Total Score for the Scale as Missing

If a patient has missing responses on one or more items of a scale, the total score of the scale for this patient is treated as missing in all statistical analyses. This method reduces the original sample size and hence the statistical tests may loss power to detect the desire treatment effects and may introduce bias in parameter estimation (Little and Rubin, 1987). However, when data are missing completely at random, this reduced data set represents the full data set and the inference based on the reduced data set can be considered reasonable. This approach may result in misleading and biased conclusions when the missing data are informative or treatment related missing. Moreover, if a patient has one missing item out of several items (let say 20 items) of a scale, his/her total score will be missing and this may not be desirable.

Method 2. Simple Mean Imputation Within Each Patient

In this method, the mean score of the available items' responses of a scale within a patient replaces the missing item response of that scale. Since the items in psychometric scale are correlated with each other, this approach seems to be reasonable. This approach is widely used in handling missing items in psychometric research and is very simple to implement. One disadvantage of this approach is that the imputed values may take fractional values that are intermediate between the discrete points on a categorical scale and hence the total score of the scale will also have a fraction part. But this is not a problem in analyzing the total score using analysis of variance models. In the categorical analyses, the fractional values need to be round off and the fraction part becomes a nuisance.

Method 3. Last Observation Carried Forward (LOCF) Approach on the Missing Total Score

This method has two steps. In the first step, if a patient has at least one missing item of a scale for a given time point, his/her total score of the scale will be missing. In the second step, his/her missing total score will be imputed using LOCF approach. That is, the missing score will be carried forward from the total score of the scale measured at the previous time point. This method assumes that the patient score remains constant over time. It may not be a reasonable assumption in clinical trial, at least for the treatment groups. Another possibility is that some patients may be dropped out from the analysis. For example, if a patient has one missing item of a scale at each of the post-baseline time points, then this patient will be dropped out from the analysis. If a patient has missing item of a scale at baseline, the patient will also be dropped out in the change score analysis of that scale.

Method 4. Last Observation Carried Forward (LOCF) Approach on the Missing Items

In this method, if any item of a scale at post-baseline period is missing, then the value for that item is imputed using LOCF approach based on the corresponding item response measured at the previous time point. Then the total scores of the scale are

calculated and analyzed. This method also assumes that the patient's respond remains constant over time, which may not be true. Moreover, if a patient has missing items at baseline, then his/her total score of the scale at baseline will be missing and in the change score analysis, this patient will be dropped out.

Method 5. Item Mean Substitution

Mean substitution replaces the missing response of an item with the mean value for the item responses from the patients who had responses on that item. This approach retains the original mean but reduces the variance of the responses for that item and it also distorts the true associations among the items (Little, 1995). This method is essentially a random substitution method (i.e., the substitution is generally not related to the patient's other items).

Method 6

If less than certain percentages (e.g., 20 or 25%) of items of a scale have missing responses then the total score of the scale is calculated as:

$$\frac{Tptal\ score\ for\ nonmis\sin g\ items\ X\ Total\ number\ of\ items}{Number\ of\ nonmis\sin g\ items}$$

Otherwise, the total score of the scale is considered as missing and the total score of the scale measured at the previous post-baseline time point is carried forward to impute the current missing value of the scale. In this method, if a patient has more than certain percentages (20 or 25%) missing items at baseline, then the total score for the scale at baseline will be missing and in the change score analysis, this patient will also be dropped out.

Statistical Method to Compare the Sensitivity of the Methods Used in Imputing Missing Items

Five hundred data sets will be generated. In each data set, there will be ten items to measure a scale total score at each time point. There will be a baseline measure and three post-baseline measures. One hundred fifty patients will be in treatment group and another 150 patients will be in placebo group. At each time point, the item score will be generated from multivariate normal and then the scores of each item will be categorized into five categories (i.e., categories 0, 1, 2, 3 and 4). At post-baseline measures, an extra random number (generated from normal distribution with non-zero mean) will be added to each of the 10 items belonging to treatment group, so that there will be an overall treatment effect at the post-baseline time points. As expected in the psychometric scale, the 10 items' scores at each time point are correlated among themselves. This generated data set will be called 'Complete Data set' from now on.

Two types of missing data mechanisms will be considered here and analyzed separately to evaluate the sensitivity of the six imputing methods stated in this study in evaluating treatment efficacy. First, under the assumption of missing completely at random, an incomplete data set will be created from the complete data set in such a way that about 25% patients will have at least one (but not more 5 items, i.e., up to 50%) missing item responses. Second, under the assumption of informative missing item responses, an incomplete data set will be created from the complete data set so that about 25% patients each of the treatment group will have at least one item (but not more than 5 items, i.e., up to 50%) missing. In doing this, about 25% patients belonging to placebo group who have lower total score at post-baseline measure will have one or more missing items, whereas, another 25% patients belonging to treatment group who have higher total score at post-baseline measures will have one or more missing items. Finally, for each missing data mechanism, the missing items will be imputed using each of the six stated methods. For each of the complete data set, there will be six imputed data sets. Finally, the complete and imputed data sets will be analyzed and compared the parameter estimates and the standard errors with the corresponding estimate and standard error obtained from the complete data set. Separate simulation study will be conducted for each of the two missing data mechanisms.

To analyze the complete and imputed data sets, the analysis of covariance model (ANCOVA) will be used. In the ANCOVA model, the dependent measure will be the change score (= end point total score-baseline total score) and the independent measures will be treatment (as a factor) and baseline total score (as a covariate). The model will be fitted to 500 generated data sets and the means of the estimated 500 unstandardized coefficients and the standard errors for treatment will be calculated. Finally, the test statistics (ratio of mean/standard error) and P-value will be calculated.

Results

Table 1 lists the results obtained from the simulation studies under the assumptions of missing items are missing completely at random and are informative missing. The estimate of regression coefficient for treatment and its standard error obtained from the analysis of the complete data are considered as reference values. When missing items are missing completely at random, the estimate and its

significance level at each of the imputed methods are very close to the corresponding reference values. The similarities in estimates and significance levels indicate that when item responses are missing completely at random, any of the six stated methods can be used to impute the item missing response.

When item missing are informative (i.e., treatment related), the estimate and significance level using imputing method 2 (i.e., Simple mean imputation within each patient) are very close to the corresponding reference values. But the remaining imputing methods are not able to reproduce the results close to the reference values. Methods 1, 3 and 6 even estimated opposite direction treatment effect as compared to the treatment effect at complete data sets.

Table 2 reports the percentages of the 500 imputed data sets are able to estimate the consistent directional treatment effect, as compared to the treatment effects for the corresponding complete data sets. When items are missing completely at random, method 2 is able to re-estimate same (in 98.8% data sets) directional treatment effects, as compared to the treatment effect obtained in the analysis of complete data sets. Methods 3, 4, 5 and 6 are also able to re-estimate consistent (in 90 to 96% data sets) directional effects, as compared to the effects the complete data sets. When items are informative missing, then method 2 can re-estimate same (in 99.4% data sets) directional treatment effects, as compared to the effects in the complete data sets. The imputing methods 4 and 5 can re-estimate consistent directional effects only for about 62 to 69% data sets.

Table 3 lists the percentages of imputed data sets (among the data sets which have consistent directional treatment effects) can also able to maintain consistent significance level of the treatment effects, as compared to the significance levels of treatment effects in the corresponding complete data sets. Significance levels are grouped here into two levels (<0.10 and >0.10). When items are missing completely at random, imputing method 2 can maintain consistent (in 95.6% data sets) significance levels of treatment effects, as compared to the significance levels of treatment effects for corresponding complete data sets. However, in the other imputing methods, lower percentages of data sets can maintain the consistent significance levels, as compared to the levels in the corresponding complete data sets. When items are informative missing, Method 2 can also maintain consistent (in 98.6% data sets) significance levels, as compared to the levels in the corresponding complete data sets. In using the other imputing methods, the imputed data sets cannot maintain the original significance levels.

Table 1. Simulation results based on 500 simulated data sets assuming different missing data mechanism

Assumed missing data mechanism	Missing data imputed method	Mean of 500 estimated unstandardized coefficients for treatment	Mean of 500 estimated standard errors for treatment	Z-statistic	P-value
Items' responses are missing completely at random	Complete data	2.076	1.251	1.659	0.097
	Method 1	2.602	1.315	1.978	0.107
	Method 2	2.085	1.269	1.643	0.100
	Method 3	1.966	1.250	1.572	0.115
	Method 4	2.051	1.042	1.968	0.049
	Method 5	2.084	0.965	2.158	0.030
	Method 6	2.067	1.253	1.648	0.099
Items' responses are informative missing	Complete data	2.076	1.251	1.659	0.097
	Method 1	-6.363	1.133	5.615	0.000
	Method 2	2.082	1.254	1.660	0.096
	Method 3	-4.347	1.078	4.030	0.002
	Method 4	0.402	1.150	0.349	0.726
	Method 5	0.272	1.126	0.242	0.808
	Method 6	-1.028	1.186	0.867	0.386

Table 2. Percentages of imputed data sets have the same directional treatment effect[1] as compare to the treatment effect in the complete data set

	Complete data versus imputed data					
Missing data mechanism	Method 1	Method 2	Method 3	Method 4	Method 5	Method 6
Missing completely at random	67.40%	98.8%	90.40%	96.4%	96.2%	91.6%
Informative missing	5%	99.4%	5%	69.0%	62.8%	28.6%

[1]Sign of the estimated regression coefficient obtained from the complete data is compared with the sign of the estimated coefficient in the corresponding imputed data set

Table 3. Percentages of imputed data sets have the same directional treatment effect[1] and significance level[2] as compare to the treatment effect and the significance level obtain from the complete data set

	Complete data versus imputed data					
Missing data mechanism	Method 1	Method 2	Method 3	Method 4	Method 5	Method 6
Missing completely at random	33.4%	95.6%	45.6%	76.2%	84.2%	62.8%
Informative missing	0.0%	98.6%	0.0%	19.8%	10.6%	1.2%

[1]Sign of the estimated regression coefficient obtained from the complete data is compared with the sign of the estimated coefficient in the corresponding imputed data set

[2]If P-value for the regression coefficient of treatment in complete data is <0.10 (or >0.10) and in the imputed data, it is also <0.10 (or >0.10) then it is called same significance level

Conclusion

In analyzing clinical trial data sets, there has been an interest in imputing methods for coping with missing item responses. Decision about the appropriate imputing method involves accurate estimation of individual total score. In the statistical analysis plan of the study protocols, different imputing methods (seems to be arbitrarily selected) are proposed to handle missing item responses. In some protocols, even, no statement about handling missing item responses are given. The reasons of not stating any method of handling missing items might be that there is a little research on how to treat missing data in multi-items scale, or little knowledge in the impacts of using different imputing methods on the scale total scores. In the present study, we provide some general information about the differential effects of the available imputing methods on the final statistical analysis for treatment efficacy and the potential limitation of those methods.

Simple mean imputation within each patient (i.e., Method 2) using the available items' response of a scale seems to the appropriate imputation method in imputing missing items (when more than 50% items have valid response) of a psychometric scale. This method seems to be robust in case of items are missing at random, or items are informative missing. The simulation studies also indicate that when items are missing completely at random, each of the six imputed methods can produce consistent results in terms of estimating treatment effects and their standard errors. However, when missing are informative, the choice of imputing method is crucial. In this situation some imputing methods even change the direction of the treatment effects. Therefore, for imputing missing data in any clinical data set, one can use simple mean imputation within each patient (so called subject Mean) in case of whether items are missing at random or not. Fairclough and Cella (1996) also conclude that simple mean imputation within each patient using completed items is generally the most unbiased and precise approach in imputing missing item responses in scoring quality of life subscales.

Other imputation methods are also available for imputing missing responses. For example, regression imputation substitutes the missing response by a predicted value that is based on the regression of other items. The regression formula is based on the patients who have complete data. The method assumes that the responses of other items of a scale are related to the missing item response. The regression imputation method closely approximates the simple mean imputation within each patient as the number of items in the scale increases. Wilks (1938) presented a proof showing that when correlated items are used, the correlation between any two weighted combinations of the items approaches 1.00 as the number of items increases. Another imputation method called Hot-Deck imputation is similar to the regression approach. This method replaces the missing variable for one individual with the score from a group of similar people. To find a similar person, one can use discriminant analysis (Roth and Switzer, 1995). Hot-Deck imputation is popular in survey research. Another method called multiple imputation is commonly used for imputing missing scores (missing response due to dropout from a study) of a scale. This method replaces each missing value by a vector of length M> = 2 imputed values and creates M completed data sets. Then standard complete-data methods are used to analyze each data set. The M complete-data inferences can be combined to form one inference that properly reflects uncertainty due to nonresponse under that model. Single imputation methods treat the imputed values as known and cannot reflect the sampling variability under the model for nonresponse, i.e., imputation uncertainty. Multiple imputation rectifies this disadvantage (Little and Rubin, 1987). However, for imputing missing item response in multi-item scale, the multiple imputation method may not be viable in practical point of view.

Acknowledgment

The author is grateful to the editor and anonymous referees whose suggestions improved this study.

Ethics

This article is original and contains unpublished material. The corresponding author confirms that all of

the other authors have read and approved the manuscript and no ethical issues involved

References

Beck, A.T., N. Epstein, G. Brown and R. Steer, 1988. An inventory for measuring clinical anxiety: Psychometric properties. J. Consult. Clin. Psychol., 56: 893-897. DOI: 10.1037/0022-006X.56.6.893

Fairclough, D.L. and D.F. Cella, 1996. Functional Assessment of Cancer Therapy (FACT-G): Non-response to individual questions. Quality Life Res., 5: 321-329. DOI: 10.1007/BF00433916

Hamilton, M.C. and W. Guy, 1976. ECDEU Assessment Manual for Psychopharmacology. National Institute of Health, Psychopharmacology Research Branch, Rev. Rockville, MD: U.S.

Little, R.J.A., 1994. A class of pattern-mixture models for normal incomplete data. Biometrika, 81: 471-483. DOI: 10.1093/biomet/81.3.471

Little, R.J.A., 1995. Modeling the drop-out mechanism in repeated-measures studies. J. Am. Stat. Assoc., 90: 1112-1121. DOI: 10.1080/01621459.1995.10476615

Little, R.J.A. and D.B. Rubin, 1987. Statistical Analysis with Missing Data. 1st Edn., Wiley, New York.

Madow, W.G., D.B. Rubin and I. Oklin, 1983. Incomplete Data in Sample Surveys. 1st Edn., Academic Press, New York, ISBN-10: 0123639018, pp: 495.

Roth, P.L. and F.S. Switzer, 1995. A Monte Carlo analysis of missing data techniques in a HRM setting. J. Manage., 21: 1003-1023. DOI: 10.1177/014920639502100511

Rubin, D.B., 1987. Multiple Imputation for Nonresponse in Surveys. 1st Edn., Wiley, New York, ISBN-10: 047108705X, pp: 288.

Sheehan, D.V., 1986. The Anxiety Disease. 1st Edn., Bantum Books, New York, ISBN-10: 0553255681, pp: 190.

Snaith, R.P., S.J. Baugh, A.D. Clayden, A. Husain and M.A. Sipple, 1982. The clinical anxiety scale: An instrument derived from the Hamilton Anxiety Scale. Br. J. Psychiatry. 141: 518-523. DOI: 10.1192/bjp.141.5.518

Spielberger, CD., Gorsuch, RL., Lushene, RE., 1970. Manual for the State-Trait Anxiety Inventory. Consulting Psychologists Press, Palo Alto, California.

Taylor, J.A., 1953. A personality scale of manifest anxiety. Abnormal Soc. Psychol., 48: 285-290. DOI: 10.1037/h0056264

Wilks, S.S., 1938. Weighting systems for linear functions of correlated variables when there is no dependent variable. Psychometrika, 3: 23-40. DOI: 10.1007/BF02287917

Zung, W.W.K., 1971. A self-rating depression scale. Arch Gen. Psychiatry, 12: 371-379. DOI: 10.1001/archpsyc.1965.01720310065008