The Parkinson’s Disease Questionnaire (PDQ-39): evidence for a method of imputing missing data

Crispin Jenkinson¹,², Catherine Heffernan¹, Helen Doll¹, Ray Fitzpatrick¹,³

¹University of Oxford, Department of Public Health, Old Road Campus, Headington, Oxford OX3 7LF, UK
²Harris Manchester College, Oxford OX1 3TD, UK
³Nuffield College, Oxford OX1 1NF, UK

Address correspondence to: C. Jenkinson, Health Services Research Unit, Department of Public Health, Old Road Campus, Headington, Oxford OX3 7LF, UK. Tel: (+44) 1865 226857. Fax: (+44) 1865 226711. Email: crispin.jenkinson@dphpc.ox.ac.uk

Abstract

Background: the Parkinson’s Disease Questionnaire (PDQ-39) is the most widely used Parkinson’s specific measure of health status. It is increasingly used in treatment trials, sometimes as a primary end-point, where any missing data can potentially cause difficulties in analyses.

Objectives: the purpose of this article is to evaluate the Expectation Maximisation (EM) algorithm for the imputation of missing dimension scores on the 39-item PDQ-39.

Methods: a postal survey of patients diagnosed with Parkinson’s disease (PD). A total of 1,372 patients were surveyed and 839 (61.15%) questionnaires returned completed or partially completed. Of these, complete PDQ data were available in 715 (85.22%) cases. Data were deleted from this complete dataset and a subset of 200 respondents from this dataset and then imputed using the EM algorithm; results were then compared to the dataset before data deletion.

Results: results gained from imputation of data closely mirrored that of the complete dataset in each case. Descriptive statistics, mean scores and spread of scores were almost identical between original and imputed datasets. Furthermore, original and imputed datasets were highly correlated [intra-class correlation coefficient (ICC) = 0.93 or greater], and mean differences were small (±1.00).
Introduction

The Parkinson’s Disease Questionnaire (PDQ-39) [1, 2] is the most widely used disease-specific measure of health status in Parkinson’s disease (PD) and has been recommended as the most comprehensively validated of competing PD-specific outcome measures [3, 4].

The instrument was developed in the United Kingdom but has been translated into over 30 languages [5–10] and has been used in both single-country [11] and cross-cultural trials [12]. Typically, the instrument has relatively low levels of missing data, but inevitably with questionnaire-based outcome instruments, some data will be missing. Rigorous analyses of the data quality of patient-completed questionnaires are becoming more common [10, 13–15], but an area that can cause concern in surveys using health status measures, and especially in clinical trials, is that of missing data [16, 17]. In trials, missing data can potentially cause both loss of power and bias [18]. The purpose of this article is to evaluate a simple missing data algorithm for the imputation of missing dimension scores on the 39-item PDQ-39.

One of the most widely used techniques for data imputation currently used within the social sciences is the Expectation Maximisation (EM) computational algorithm for multiple imputations [19]. The EM approach is an iterative procedure with two discrete steps. First, the ‘expectation’ step computes the expected value of the complete data log likelihood based upon the complete data cases and the algorithm’s ‘best guess’ as to what the sufficient statistical functions are for the missing data based upon the existing data points. The ‘maximisation’ step procedure substitutes the expected values (typically means and covariances) for the missing data obtained from the expectation step and then maximises the likelihood function as if no data were missing to obtain new parameter estimates. The new parameter estimates are substituted back into the expectation step, and a new maximisation step is performed. The procedure iterates through these two steps until convergence is obtained. Convergence occurs when the change of the parameter estimates from iteration to iteration becomes negligible [20]. EM has a number of advantages over rival imputation methods based on multiple regression, not least that it is an established methodology that makes fewer demands of data in terms of statistical assumptions and appears, in general, to be more accurate [18]. Furthermore, EM is implemented in a wide variety of computer packages including popular software such as SPSS.

The purpose of this article is to explore the use of EM for the imputation of missing dimension scores on data gained from the PDQ-39.

Methods

Analysis plan

Data from a survey of patients registered with the Parkinson’s Disease Society were analysed to determine whether EM could accurately estimate mean scores and distributions on the eight dimensions of the PDQ-39. This was achieved in a six-stage process mentioned below:

(i) Any case from the dataset that contained any missing PDQ-39 dimension scores was deleted. This led to a dataset that had a complete set of scores for all respondents on all the eight dimensions of the PDQ.

(ii) Dimension scores were deleted from this newly created sub-set of the original dataset, reflecting the manner in which data had been missing in the original complete dataset (i.e. the pattern of missing data was identical to the original data, although the cases from which deletions were made were selected randomly).

(iii) The data that had been deliberately removed were then imputed using EM assuming a normal multivariate distribution, and mean scores, standard deviations, median and 25th and 75th percentiles for each dimension were calculated; these results were compared to the original scores before the deletion of data. Both parametric (t-test) and non-parametric (Wilcoxon) tests were used to determine whether data were statistically significantly different. This procedure was then repeated, but randomly selecting another set of cases for data deletion to check the results of the first analysis.

(iv) Ten per cent of the responses to each PDQ-39 dimension were randomly deleted and then data imputed as before and results compared with the original dataset. Once again this procedure was repeated with another random selection of data, deleted and then imputed.

(v) It has been suggested that the quality-of-life data are most likely to be missing amongst the most severely ill of respondents [21, 22]. Consequently, another analysis was undertaken restricting the 10% of deletions and imputations to those cases in the bottom third of scores on the PDQ-39 Index [23], which is a summary score calculated on the basis of all the eight dimensions. Once again this procedure was repeated.

(vi) The purpose of this sixth step was to ascertain whether EM is an appropriate algorithm for use in smaller datasets. Consequently, a sub-set of 200 cases of the respondents who had complete data in the original dataset was selected. Ten per cent of data for each dimension were deleted and then imputed using EM as before. This step was then repeated.

The original dataset is referred to here as ORIGINAL.
The dataset derived from ORIGINAL, which contains only cases with complete data for all the eight dimensions of the PDQ-39 is called COMPLETE.

The dataset created from COMPLETE, where PDQ-39 values have been deliberately removed is called VALUES_REMOVED.

The dataset created from COMPLETE, where deletions are restricted to the most severely ill respondents is called SEVERELY_ILL.

The dataset which contains a small sub-set of COMPLETE is called SMALL.

Survey design

The data analysed here were gained from a postal survey that was conducted of randomly selected members of 13 local branches of the Parkinson’s Disease Society. Full details of the survey can be found elsewhere [24].

The Questionnaire

The PDQ-39 comprises 39 questions measuring the eight dimensions of health: mobility, activities of daily living, emotional well-being, stigma, social support, cognition, communication and bodily pain. Dimension scores are coded on a scale of 0 (perfect health as assessed by the measure) to 100 (worst health as assessed by the measure). The development of the instrument and studies confirming the reliability and validity of the instrument are documented elsewhere [1, 23, 25].

Results

A total of 1,372 PDS members were surveyed of which 851 (62.03%) questionnaires were returned. Of the 851 questionnaires returned, complete data on the PDQ were available in 715 (84.02%), and it is this sub-set of the data that is analysed here (called COMPLETE). The average age of respondents was 70.48 years (SD 9.57; min = 32.89, max = 90.57, n = 715). A total of 432 (60.42%) of the respondents were male and 283 (39.58%) female. Descriptive statistics for the eight dimensions of the PDQ-39 are summarised in Table 1.

The second stage in testing EM on the PDQ-39 involved removing data from COMPLETE. The distribution of missing data in the subsequent dataset, VALUES_REMOVED, replicated the pattern of missing data in the original dataset (ORIGINAL). In the original dataset, 125 cases were missing a score on at least one dimension because of missing data. Consequently, data in COMPLETE were deleted in 125 cases: the cases for deletion were chosen randomly, but the deletion of data exactly mirrored that in the original dataset. The distribution of dimensions with complete data and descriptive statistics for this dataset before imputation are shown in Appendix 1 in the supplementary data on the journal website (http://www.ageing.oupjournals.org/). As the number of cases with missing data in VALUES_REMOVED is exactly the same as in ORIGINAL, the proportion of cases with missing data is somewhat higher. Consequently, VALUES_REMOVED contains a higher proportion of missing data than is typical of PDQ-39 datasets: this strategy was deliberately chosen to test the data imputation algorithm in datasets with higher than typical amounts of missing data. Descriptive statistics are broadly similar to those gained from the database as a whole.

The third stage in testing EM on the PDQ-39 involved applying the EM algorithm to the dataset. Descriptive statistics are summarised in Table 2. Mean differences between original and imputed scores for all of the eight dimensions were very small (Table 2). Correlations between the two datasets were high [intra-class correlation coefficient (ICC) = 0.95 or higher, P<0.001].

Table 1. Descriptive statistics for all respondents on the eight dimensions of the Parkinson’s Disease Questionnaire (PDQ-39) (COMPLETE dataset, n = 715)

<table>
<thead>
<tr>
<th>PDQ dimension</th>
<th>Mean (SD)</th>
<th>Min–Max</th>
<th>25</th>
<th>50</th>
<th>75</th>
<th>Mean difference (SD)</th>
<th>95% CI</th>
<th>Paired sample correlations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mobility</td>
<td>59.42 (30.06)</td>
<td>0–100</td>
<td>37.50</td>
<td>62.50</td>
<td>85.00</td>
<td>0.30 (6.79)</td>
<td>-0.20–0.80</td>
<td>0.97</td>
</tr>
<tr>
<td>Activities of daily living</td>
<td>51.76 (27.69)</td>
<td>0–100</td>
<td>29.17</td>
<td>50.00</td>
<td>75.00</td>
<td>0.10 (3.08)</td>
<td>-0.13–0.32</td>
<td>0.96</td>
</tr>
<tr>
<td>Emotional well-being</td>
<td>37.99 (23.59)</td>
<td>0–100</td>
<td>20.83</td>
<td>37.50</td>
<td>54.17</td>
<td>-0.051 (4.64)</td>
<td>-0.39–0.29</td>
<td>0.96</td>
</tr>
<tr>
<td>Stigma</td>
<td>30.72 (26.25)</td>
<td>0–100</td>
<td>6.25</td>
<td>25.00</td>
<td>43.75</td>
<td>-0.02 (5.23)</td>
<td>-0.41–0.36</td>
<td>0.96</td>
</tr>
<tr>
<td>Social support</td>
<td>25.11 (24.18)</td>
<td>0–100</td>
<td>0</td>
<td>25.00</td>
<td>41.67</td>
<td>0.15 (5.27)</td>
<td>-0.24–0.53</td>
<td>0.95</td>
</tr>
<tr>
<td>Cognitions</td>
<td>43.62 (24.00)</td>
<td>0–100</td>
<td>25.00</td>
<td>43.75</td>
<td>62.50</td>
<td>0.28 (4.03)</td>
<td>-0.02–0.57</td>
<td>0.95</td>
</tr>
<tr>
<td>Communication</td>
<td>33.60 (25.41)</td>
<td>0–100</td>
<td>16.67</td>
<td>33.33</td>
<td>50.00</td>
<td>0.10 (3.34)</td>
<td>-0.14–0.35</td>
<td>0.95</td>
</tr>
<tr>
<td>Bodily discomfort</td>
<td>49.10 (25.27)</td>
<td>0–100</td>
<td>33.33</td>
<td>50.00</td>
<td>66.67</td>
<td>0.17 (4.65)</td>
<td>-0.17–0.51</td>
<td>0.95</td>
</tr>
</tbody>
</table>

PDQ, Parkinson’s Disease Questionnaire.
The procedures outlined above were then repeated, but missing data were created for another random sample from COMPLETE to create a new VALUES_REMOVED. Data reflected the results gained above with very small differences for all PDQ-39 dimensions between the original dataset and the dataset containing imputations (<±0.4). Correlations between the original data and the imputed data for each dimension were high (ICC = 0.97 or higher, P<0.001).

The fourth step in testing the EM procedure on the PDQ-39 involved removing 10% of data per dimension (from COMPLETE) and then imputing values for this missing data. This led to 297 (41.54%) cases with complete data, and the remainder requiring at least one-dimension score to be imputed. Descriptive statistics and mean differences between the original data and the imputed dataset are summarised in Table 3, as well as descriptive statistics for the dataset, including the imputations. Mean differences between original and imputed scores for all of the eight dimensions were very small. Correlations between the two datasets were high (ICC = 0.95 or higher, P<0.001). The procedures outlined above were then repeated, but missing data were created for another 10% random sample for each dimension from COMPLETE: only 307 (42.94%) cases had complete data across all dimensions of the PDQ-39. Data reflected the results gained above with mean differences for all PDQ-39 dimensions between the original dataset and the dataset containing imputations being very small (<±0.35). Similarly, results for all the eight dimensions were highly correlated (ICC = 0.95, or higher, P<0.001).

The fifth step in testing the EM algorithm was similar to that mentioned above, but the 10% of cases were all removed from respondents who scored in the bottom third of scores on the PDQ-39 Index, creating a dataset called SEVERELY_ILL. This led to 492 (68.81%) cases with complete data, and the remainder requiring at least one-dimension score to be imputed. Results in this analysis were similar to those from the random deletions undertaken previously. Mean differences between original and imputed scores for all of the eight dimensions were very small (see Appendix 2 in the supplementary data on the journal website http://www.ageing.oupjournals.org/). Correlations between the two datasets were high (ICC = 0.97 or higher, P<0.001). The procedures outlined above were then repeated, but missing data were created for another 10% random sample for each dimension from the 200 cases of the sample: only 82 (41%) of the respondents had complete data for the PDQ-39. Data reflected the results gained above, with the mean differences for all PDQ-39 dimensions between the original dataset and the dataset containing imputations being small (<±0.70). Similarly, results for all the eight dimensions were highly correlated (ICC = 0.95 or higher, P<0.001).

**Discussion**

This article has evaluated a widely used missing data algorithm (EM) for use with the PDQ-39. Missing data can lead to problems in analyses of data and, in treatment trials, can cause loss of power and are potentially a source of bias. Furthermore, common use of list-wise deletion (i.e. completely omitting cases with missing data) can not only lead to bias but could be viewed as unethical as patients have

<table>
<thead>
<tr>
<th>Mean (SD)</th>
<th>Min–Max</th>
<th>Mean difference (SD) from original dataset</th>
<th>Lower</th>
<th>Higher</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mobility</td>
<td>59.0 (28.71)</td>
<td>0–100</td>
<td>-0.1 (8.72)</td>
<td>-0.75</td>
</tr>
<tr>
<td>Activities of daily living</td>
<td>51.4 (26.18)</td>
<td>0–100</td>
<td>-0.2 (8.05)</td>
<td>-0.81</td>
</tr>
<tr>
<td>Emotional well-being</td>
<td>38.4 (22.72)</td>
<td>0–100</td>
<td>0.3 (6.18)</td>
<td>-0.12</td>
</tr>
<tr>
<td>Stigma</td>
<td>30.7 (24.84)</td>
<td>0–100</td>
<td>-0.1 (7.92)</td>
<td>-0.63</td>
</tr>
<tr>
<td>Social support</td>
<td>24.9 (23.05)</td>
<td>0–100</td>
<td>-0.1 (7.47)</td>
<td>-0.64</td>
</tr>
<tr>
<td>Cognitions</td>
<td>43.2 (22.91)</td>
<td>0–100</td>
<td>-0.1 (7.00)</td>
<td>-0.63</td>
</tr>
<tr>
<td>Communications</td>
<td>33.4 (24.19)</td>
<td>0–100</td>
<td>-0.1 (7.35)</td>
<td>-0.61</td>
</tr>
<tr>
<td>Bodily discomfort</td>
<td>49.0 (24.40)</td>
<td>0–100</td>
<td>0.0 (8.10)</td>
<td>-0.55</td>
</tr>
</tbody>
</table>
been given time to complete the survey. However, it is possible that imputation of values may severely skew results and hence be inappropriate and that simply reporting results based on the data available is the most appropriate solution [26]. However, the latter procedure can lead to reduced sample size and reduced power and hence the decision to undertake the analyses that have been reported in this article.

The procedure used here imputes dimension scores and not individual item responses. This approach was adopted, because previous analyses have shown that the eight dimensions of the PDQ-39 are tapping aspects of a broad general underlying phenomenon (overall health). Indeed, the concept of unidimensionality underlies the PDQ-39 Single Index Score that is calculated by summing scores for the eight dimensions [23]. Consequently, one would expect a pattern to dimension scores, and the results gained here would support this hypothesis. Indeed, differences between the original full dataset and the consequent datasets in which data have been deleted and then imputed are very small and are unlikely to be meaningful [22].

We would not advise using this, or any other data imputation algorithm, on smaller \( n \) (< 200) datasets than those presented here. Indeed, the use of EM on small datasets can produce misleading variance to the distribution of data [27]. Furthermore, datasets with very large amounts of missing data (in excess of 10%) must be viewed with caution as this is a high rate of missing data atypical of the PDQ-39 and is likely to reflect some aspects of the sample surveyed or the method of data collection. Indeed, the PDQ-39 exhibits low levels of missing data compared with generic instruments that have been used in elderly samples [28]. However, in the analyses undertaken here, we have adopted conservative estimates (i.e. large) for missing data, and the success in imputing the data scores suggests that the imputation method will work in more typical datasets where less data are likely to be missing. Results reported here provide encouraging evidence of a method of imputation for data on health-related quality of life. Furthermore, we have explored the possibility that missing data are systematically related to the overall health status of individuals, and the algorithm performs well.

In conclusion, we would recommend EM as a method for imputing PDQ data in surveys and trials. The technique is implemented in a wide variety of software packages including, for example, SPSS. However, use of this technique should not be at the expense of trying to determine what the reasons for such missing data may be and whether strategies may exist for improving data collection. Researchers would be advised to undertake a sensitivity analysis by comparing results gained from their original dataset, which includes the missing data, to those gained after implementing the EM algorithm. In any instance where meaningful differences were to be found between the two datasets, this may raise questions about the appropriateness of this technique and would suggest that non-responders may differ to responders in an important respect.

### Key points

- Missing data on health status questionnaires can reduce sample size and power in trials which include such measures as outcomes.
- The PDQ-39 is one of the most widely used measures of outcome for studies evaluating the impact of PD.
- Reports of missing data on the PDQ-39 are generally low, but ideally such data could be imputed.
- EM is a widely implemented method for imputing data, found in software such as SPSS.
- In this study, EM was found to be an accurate method of imputing missing data on the PDQ-39 in datasets with sample sizes >200 cases.

### Sources of Funding

We thank the Parkinson’s Disease Society for funding the original studies on which the data used in this article are based.

### Conflict of interest

None.

### References


Received 3 June 2005; accepted in revised form 25 April 2006