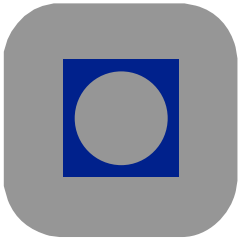
Missing Data in Palliative Care Research

Imputation and Analysis

Peter Fayers

Department of Public Health
University of Aberdeen



NTNU
Det medisinske fakultet

Missing data

- **Missing data is a major problem when reporting QoL, especially …**
  - **in trials with extended follow-up**
  - **in palliative care**

- **The problem:**
  arguably, patients with the worst HRQL are the ones most likely to stop completing questionnaires.

Why does missing data matter?

- **1. Bias**

If the proportion of data missing is not small then:

are the characteristics of patients with missing data different from those for whom complete data are available?

Why does missing data matter?

- **2. Power**

– **A study loses power if data are missing**

  – **a larger sample size is required.**

– **Note that increasing the sample size will compensate for the loss of power, but will *not* reduce the bias.**

- ***Always try to minimise the amount of missing data!***

Compliance

- **Many clinical trials have poor compliance with QoL assessment.**
- **It is common for less than 2/3 patients to return QoL assessments during or after treatment.**
- **Which patients fail to complete the questionnaires? The most ill …?**
- **How can we interpret the results of the study if there is possible bias?**

Compliance

- **Unavoidable reasons – patient attrition including death.**

- **Low compliance – forms which (theoretically) could have been completed but were not;**

  this includes patients who are in pain, very ill, or frail.

Patterns of missing data

- **Missing Completely at Random (MCAR)**
  - **the probability of response at time *t* is independent of both observed data and the unobserved data.**
- **Missing At Random (MAR)**
  - **the probability of response at time *t* depends on the observed values but *not* the unobserved data.**
- **Not Missing At Random (NMAR)**
  - **the probability of response at time *t* depends on the unobserved values.**

Methods for missing forms - imputation

- **Imputation – "best guess" estimate.**

- **Use information from**
  - **other "similar" patients,**
  - **values from previous and/or later assessments by the same patient,**
  - **or a mixture of both.**

- **If items are used only as components of the scale, it may not be necessary to impute values for those items, only for the scale score itself.**

Naïve methods of imputation

- **Last Value Carried Forward (LVCF)**
  - **The values that were recoded by the patient at the last previously completed QoL assessment are carried forward.**
- **Simple Mean Imputation**
  - **The replacement of missing QoL scores by the mean score calculated for patients who *did* complete the assessment.**

Horizontal Mean Imputation

- **"Horizontal" mean imputation uses the mean of each patient's previous scores.**
- **It reduces to the LVCF method if there is only one previous assessment available.**
- **Many other, more general, regression-based methods are available.**

Reduced Standard Deviation (SD)

- **Methods such as Simple Mean Imputation result in a biased estimate of Standard Deviation.**

- **The estimate of the SD will be reduced artificially.**

- **This can lead to distorted significance tests, and falsely narrow CIs.**

(The SD should be corrected, or equivalently we must make adjustment in analyses.)

Markov Chain (MC) Imputation

- In the methods described so far, the imputed values will be the same for any two patients with the same profile of successive non-missing values.
- MC imputation allows these two patients to have different imputed QoL values.
- It assigns, for a patient in a particular QoL state at one assessment, *transition* probabilities of being in each of the possible states.
- E.g. If a female patient aged 60 has a QoL score of 70 (on 0 – 100 scale), what is the probability her next (missing) value would have increased to, say, 80? Or 90? Or decreased to 50? Etc.

Hot Deck (HD) Imputation

- **Suppose in a palliative care RCT a male patient aged 65 with metastatic SCLC has missing data at 1 month.**
- **Identify other patients in the RCT who have same age, same gender, same disease, etc.**
- **Select a QoL score, *at random*, from these patients.**
- **Substitute this as the imputed value for the patient with the missing QoL assessment.**

Multiple imputation

- **Since a random element is included in the selection of values, the augmented dataset will be just a random one out of many potential datasets.**
- **The idea of multiple imputation is that many alternative ”complete” datasets can be created.**
- **The analysis can be repeated for each dataset and then combined into a final summary analysis (Rubin 1987).**

Analytical methods

1. For each of *N* patients with missing data,
2. … identify a set of patients with similar prognosis.
3. Impute a set of values from these similar patients.
4. Analyse the augmented, seemingly-complete, data set.
5. Calculate p-values, estimate treatment effects, etc.
6. Repeat this process *M* times.
7. "Average" the *M* results, using Rubin's methods.
8. The overall results provide the multiple imputation estimates.

Aberdeen HSRU trials

- ***5 RCTs***

- **Strenuous efforts to recover QoL data that was initially missing,**
  - **By issuing repeated reminders,**
  - **By offering to interview patients .**

- **Unique opportunity to explore the performance of imputation methods.**

1. **Assume data collected by reminders is “missing”, as it would be in most RCTs.**

2. **Apply imputation procedures, and compare results against the data actually retrieved by reminders.**

Five Aberdeen HSRU trials

- **Multiple imputation was shown to be more suitable then simple imputation methods.**

- **MI models the uncertainty in the missing data and is based on the MAR assumption, which is more plausible in QoL data than MCAR.**

Don’t ignore missing data …

- **Many investigators are suspicious about imputation techniques, because of the assumptions overtly involved.**
- **However, *ignoring* missing data in effect makes the assumption that patients who failed to respond are similar to those who did.**
- **Imputation tries to use available information to make better allowance for patients with missing data.**

Sensitivity analysis

- **Check whether choice of imputation method affects the results.**

- **Try extreme cases:**
  1. **Assume missing values correspond to patients with very *poor* QoL.**
  2. **Assume missing values correspond to patients with very *good* QoL.**

- **Are the results essentially unchanged?**

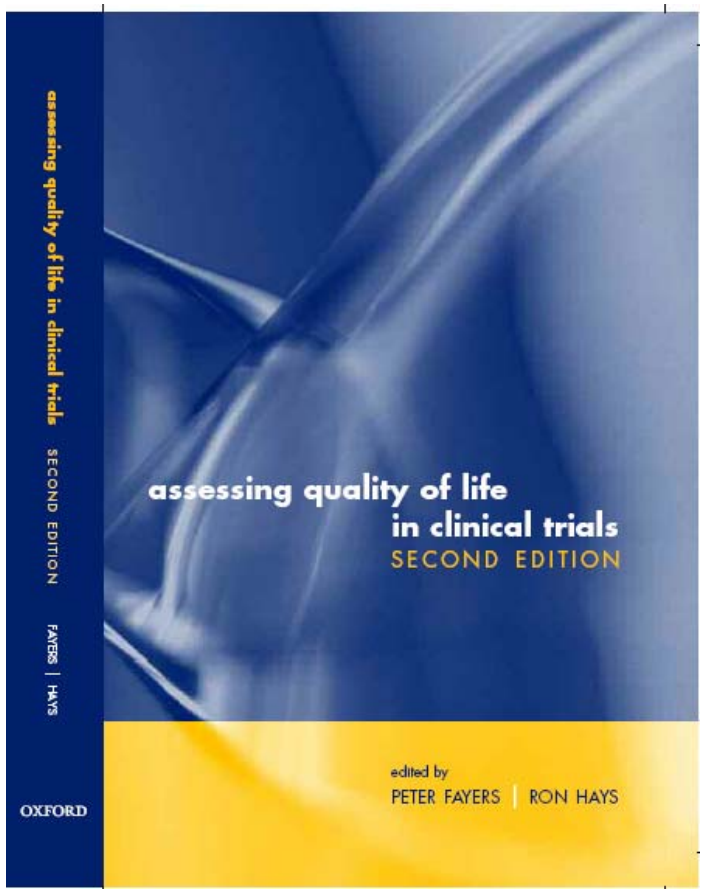
- **Or are the conclusions sensitive to the assumptions made for imputation?**

Conclusions - I

- **Methods such as LVCF are simple; however, naïve use of LVCF cannot be recommended.**
- **Multiple imputation methods are efficient. They take additional patient information into account, and preserve the magnitude of the SDs and CI.**
- **Decide and specify the method of imputation in advance.**
- **The QoL scales that are major endpoints should be the focus for determining the imputation process.**

Conclusions - II

- **Sophisticated imputation methods are no substitute for the real data.**
- **One cannot create data from nothing! Imputation is a salvage job.**
- **The *only* way to be confident of no bias is – ensure good compliance.**
- **Studies with poor compliance remain unconvincing and unpublishable, no matter how carefully the data is analysed.**
- **Always aim for 100% compliance!**



- **Fayers,PM & Hays, R (eds.) 2005**
- **Assessing Quality of Life in Clinical Trials.**
- **Oxford University Press.**
- **ISBN:** 0-19-852769-1

Especially chapters 2.4 (proxies),
3.2 (preventing missing data),
3.3 (missing data: analyses)

Also see:
Fairclough DL (2002) Design and Analysis of Quality of Life Studies in Clinical Trials. Publ: Chapman & Hall ISBN: 1-58488-263-8.