Appendix 3: Definition of the types of missingness and terminology used by each paper to describe the different reasons for missing participant data, with specific terms underlined.

Definitions

Missing completely at random: missingness of an outcome is not related to any observed or unobserved variables; the missingness probability is independent of all previous, current and future assessments.

Missing at random: missingness of an outcome may be related to observed or unobserved variables, but is not related to the actual value of the outcome, conditional on the observed variables; the missingness probability does not depend on the missing values.

Informatively missing: missingness of an outcome is related to the value of that outcome, even conditional on other observed variables; the missingness probability may depend on the unobserved values.

References:

Higgins J, White I, Wood A. Imputation methods for missing outcome data in meta-analysis of clinical trials. Clinical Trials 2008; 5: 225–239

Troxel AB, Fairclough DL, Curran D et al. Statistical analysis of quality of life with missing data in cancer clinical trials. Stat Med. 1998,17(5-7):653-666

Terminology used to describe the different reasons for missing participant data

Staquet 1996 [13]

- Missing data due to non-completion of the questionnaire resulting from the death of the patient;
- Missing data due to non-completion of the questionnaire for reasons other than death
- Missing data due to non-response to items

It should be specified what missing data are due to informative (non-random) censoring (i.e. missing data due to the patient's health state or to his particular treatment) or to non-informative (random) censoring.

Bernhard 1998 [12]

In some circumstances, data are unavoidably missing, as for example when patients have died or are too ill to complete forms.

Obviously, certain sources of missing QOL data, such as attrition due to death and withdrawal of some patients from QOL assessment due to progressive disease or treatment-related toxicity, are unavoidable.

A number of other factors can account for missing data: questionnaires are inadvertently not

given to the patient, the wrong questionnaires are given, or questionnaires are administered at the wrong time; the patient is either not given instructions or not appropriately instructed in how to complete the questionnaire; questionnaires are not routinely inspected for missing responses after collection.

Deteriorating health status is one of the patient factors that both generates missing QOL data and is not very amenable to intervention.

In case the *patient refused* to complete the questionnaire, give reason:

- Felt too ill
- Lack of interest or time
- Complained of burden
- Bothered by questions
- Other (specify):

Troxel 1998 [20]

QOL assessments are often not obtained because of negative events experienced by patients such as treatment toxicities, disease progression, or even death.

Missing data are often described as either 'dropout' or 'intermittent'. Dropout occurs when a subject, once missing an assessment, is never observed again. Intermittent or non-monotone missingness occurs when a subject misses an assessment but is later observed.

Paper discusses: missing at random (MAR), missing not at random (MNAR), and missing completely at random (MCAR)

Liu 2006 [21]

A selection model is a natural choice, where patients withdrew from the study because of worsening in treatment responses in FPG and HbA1C

In the pattern mixture model, parameter f, the parameter conditional on the missing data pattern, is the primary concern, i.e., probability distribution of the response depends on dropout status and different response patterns can be modeled for patients who drop out or continue for different reasons

Discussed MCAR, MAR and NMAR

- Examples of MCAR include patients who moved away due to relocation (not reasons related to study outcomes), closing of the study and late entry of patients administratively 'censored'
- Examples of MAR include the predefined dropout process based on recorded biomarker values. For example, in a diabetes study, the protocol pre-defined a withdrawal criterion as when a patient's recorded fasting plasma glucose (FPG) level exceeds 270 mg/dl.
- Dropouts due to drug-related adverse events (AEs) such as weight gain, GI intolerance or hypoglycemic events could also be considered as MAR. An example of NMAR is missing FEV1 in asthma trials when patients withdraw due to acute worsening

conditions, and the reason for the missing FEV1 measurements are likely related to the unobserved FEV1 values, and hence not missing at random

Patients are likely to withdraw from the treatment due to lack of efficacy.

Missing data/dropouts

In diabetes clinical trials, it is usually anticipated that patients would withdraw if their glycemic conditions worsen because they are not on any anti-hyperglycemic drug (in placebo-controlled trials) or they develop resistance or do not respond to a certain class of anti-hyperglycemic drugs.

The second type of missing data includes those withdrawals due to treatment-related AEs that are likely to be MAR provided that treatment is included in the analysis model. The third type of missing data includes those due to administrative reasons, such as patients' moving away, which would be MCAR. Other types of missing data are caused by patients withdrawing consent (refuse to be on Statin), protocol violations (e.g., did not use double barrier), etc., which could be MAR or NMAR.

Amico 2008 [22]

Further classifications for attrition in terms of reasons (eg, requested withdrawal vs lost) or other categorization systems were used in only 18 (35%) of the studies, and partially used in 10 (20%).

Sterne 2009 [23]

Reasons for missing data are commonly classified as: MAR, MCAR and NMAR.

Polit 2010 [24]

Yet, the advantages of an experimental design can be undermined by removals, withdrawals, and subject losses, which can nullify the initial equivalence of the experimental and control group.

Two categories:

1- Researcher-initiated: removal of subjects who were not really eligible, non-compliers/ dropouts/ protocol violators, Removal of subjects who actually received

the alternative treatment (crossovers)

2- Subjects-related: death, severe illness/disability, severe pain, lack of interest or time, preference for other treatment, transportation problems, Unable to locate subject to obtain post-randomization outcome data.

Discuss the mechanism of missingness (MAR, MCAR and MNAR)

- For example, an accidental death would typically result in outcomes that are MCAR

If the intervention condition is especially attractive, however, there is a risk that control group

participants will drop out of the study because of their disappointment at not receiving it, or they may seek similar alternative treatments.

In many studies, outcome data are missing for people who drop out of the study because of time constraints or other barriers.

Alshurafa 2012 [25]

Missing outcome data (MOD) mainly due to patients lost to follow-up for the primary analysis.