Analysis, reporting and interpretation of health-related quality of life: a Quality of Life Office position paper

Melanie L Bell[1,2] and Madeleine T King[1,2]

1. The Psycho-Oncology Co-operative Research Group (PoCoG), University of Sydney, Sydney Australia

2. The Cancer Australia Quality of Life Office, University of Sydney, Sydney, Australia

Quality of Life Office
PoCoG, School of Psychology
Transient Building (F12)
The University of Sydney
NSW 2006 Australia

T +61 2 9114 1365
F +61 2 9036 5292
E qol.office@sydney.edu.au
sydney.edu.au

ABN 15 211 513 464
CRICOS 00026A

Glossary

Term	Definition	Example
Auxiliary data	Data collected in addition to the outcome data which may be related to missingness or the outcome.	ECOG performance status
Covariates	Explanatory variables in a statistical model. Also known as predictors, independent variables, or X.	Sex, age, treatment
Domains	Subscales in a questionnaire, each addressing a different phenomenon (or 'construct'). See *questionnaire*.	FACT-G physical domain; Social domain
Dropout	Also known as *monotone missingness*, the data are no longer observed after a certain point (because the participant has dropped out or died).	
Intermittent missing data	An outcome is missing at one assessment but is observed at a following assessment.	
Items	The questions in a questionnaire. See *questionnaire.*	"I have nausea";" I feel close to my friends"
Minimally important difference (MID)	The smallest difference in score in a domain that patients perceive as important, either beneficial or harmful, and which would lead their clinician to consider a change in the patient's management [1].	
Missing data mechanism	The underlying cause of why data are missing.	
Missing completely at random (MCAR)	The probability of the data missing is unrelated to the patient's outcome.	A nurse forgets to administer the questionnaire.
Missing at random (MAR)	The probability of missingness depends only on past observed data (outcomes and possibly covariates). The essential concept of MAR is that *conditional on observed data*, the data are MCAR.	A nurse tells the patient not to complete the HRQoL questionnaire because he was too sick at the last visit.
Missing not at random (MNAR)	The probability of missingness depends on the value of the missing outcome itself, *even when observed data are taken into account* (both outcomes and covariates).	A nurse tells the patient not to complete the HRQoL questionnaire because he was too sick at **this** visit.
Outcome variable	The variable of interest in a study. Also known as the dependent variable or Y.	QoL
Patient Reported Outcomes (PROs)	A measurement based on a report that comes directly from the patient (i.e., study subject) about the status of a patient's health condition without amendment or interpretation of the patient's response by a clinician or anyone else[2].	Pain, depression, quality of life, vomiting, sleeplessness, dyspnea.
Questionnaire	A set of questions (*items*), sometimes grouped into *subscales*, designed to measure an underlying construct. Sometimes called an *instrument* or *measure*.	FACT-G

1. Introduction: aims and scope of this paper

Quality of life measures are widely used in various contexts including cancer clinical trials. We use the term 'quality of life' in a broad but health-related sense to refer to the effects of disease and treatment as perceived and reported by the people affected by cancer. To emphasise the focus on health, we use the term health-related quality of life (HRQoL). HRQoL measures are patient reported outcomes (PROs). They commonly include numerous domains, including various aspects of functioning (e.g. physical, emotional, social, sexual), symptoms (e.g. pain, fatigue, nausea), and global assessments of health and HRQoL.

The objective of this paper is to outline important issues in the statistical analysis of HRQoL as an outcome measure in randomised controlled trials (RCT), and in the reporting and interpretation of HRQoL results. We assume that the questionnaire has been shown to be psychometrically valid and reliable in the target population. For more on establishing psychometric properties, see the US FDA guidance document on PROs [2]. The information in this paper on missing data is a summary of an extensive review on practical and statistical issues in missing data for longitudinal patient reported outcomes.[3]

2. Questionnaire scoring

Most HRQoL questionnaires have scoring instructions, and are available on the authors' website. These instructions should always be used to score patient responses into summary scales for statistical analysis. The main topics in these instructions are standardisation, reverse scoring of items and how to handle missing items.

2.1 Standardisation

Some, but not all, questionnaires are standardised to a scale range of 0-100 points. A general formula for standardisation is

$$(\text{sum all items} - m*j) \times (100/(m \times (k-j))$$

where m = # of items, j=minimum value an item can take, k = maximum value that an item can take.

2.2 Reverse scoring

When a subscale of a HRQoL questionnaire includes some questions which are worded in a positive way relative to the response options (such as "I am sleeping well") and others that are worded negatively (e.g. "I have pain") then some questions need to be reverse scored. For example, the FACT-G contains a mix of positively and negatively worded questions, each with have responses options of 0 = not at all; 1 = a little bit; 2 = somewhat; 3 = quite a bit; 4 = very much. In order for the subscale and total score to be interpreted as higher scores meaning better outcomes, negatively worded questions need to be reversed. To do this, one just needs to subtract the response from 4:

$$\text{reversed response} = 4 - \text{original response}.$$

Now lower values indicate more pain, and higher values indicate less or no pain, and hence better quality of life.

2.3 Missing items

When items are missing, a common approach is *half-mean imputation*: if half or more of the items are complete, the missing items can be imputed with the mean of the remaining items. If there is more than one subscale, this is done within each subscale. In general, this is a valid approach because psychometrically validated questionnaires measure some underlying construct, so items

within the questionnaire are correlated with one another. Fairclough and Cella[4] investigated various approaches to missing items and found that this fairly simple approach yielded robust results. However, if there is an ordering of difficulty of items within a scale, this may not be appropriate.[5, 6] For example, the physical functioning scale of the QLQ-C30: if patients need help with eating and dressing, then they certainly will have trouble taking a long walk.

3. Missing data: definitions and descriptions

It is quite common for patients in longitudinal studies to miss some planned QOL assessments. This type of missing data can threaten the validity of results.[5, 7, 8] HRQoL in cancer trials can be particularly problematic because of the likelihood that the data are not missing randomly. For example, if patients who are doing worse are less likely to fill out their HRQoL assessments, HRQoL can be overestimated and toxicity can be underestimated.

There are three types of missing data, originally defined by Rubin.[9] Data missing completely at random (MCAR) are those where the probability of missing is unrelated to the patient's outcome. For example, a researcher forgets to administer the questionnaire. Data missing at random (MAR) are those whose probability of being observed depends only on past observed data (outcomes and possibly other covariates). The essential concept of MAR is that *conditional on observed data*, the data are MCAR. Data missing not at random (MNAR) are those whose probability of missingness depends on the value of the missing outcome itself, *even when observed data are taken into account* (both outcomes and covariates). Sometimes a distinction is made between MCAR and covariate dependent MAR. However, when the covariate is conditioned on (by including it in the model) data become MCAR, so we do not distinguish between these types. The *missingness mechanism* is the underlying cause of why data are missing.

It is important to describe the patterns of missingness, both to inform analyses and for transparent reporting. A table, stratified by arm, showing how many participants were assessed at each time point is essential. Graphing the data, stratified by arm and dropout time, can provide insight into the type of missing data. If the trajectories over time are substantially different, data are not MCAR. For example, are patients who have lower baseline values more likely to drop out, or are steeper rates of increase or decrease over time associated with dropout?

4. Imputation

There are two types of imputation: *simple* (sometimes called single) and *multiple*.

4.1 Simple imputation

Simple imputation refers to filling in missing observations with a single value. Common choices are the mean of the sample, the individual's last observation carried forward (LOCF), their mean, best or worst values, or an extrapolated or interpolated value based on a regression on the individual. There are two problems with simple imputation. First, variance in the sample is artificially reduced, leading to underestimated standard errors and increased type I error rates. Second, simple imputation can yield biased results. Although there is folklore that LOCF is conservative, it has been shown that the bias is unpredictable.[10] Under LOCF, patients who drop out would get the same HRQoL scores as their last assessment's HRQoL, which is improbable in many situations (like advanced cancer).

4.1 Multiple imputation (MI)

MI is based on filling in missing data by drawing from a distribution of likely values, and does not suffer from the variance underestimation of simple imputation.[11] MI can be a useful way to incorporate auxiliary data when one does not want to use an adjusted model, and can be used for

either outcomes or covariates (particularly in observational studies when important covariates are missing). This is achieved by using auxiliary data as covariates in the imputation model (see below). There are three steps to MI: 1) Impute multiple (*M*) times, using a regression model called the imputation model, so that there are *M* complete sets of data; 2) Analyse each of the *M* data sets; 3) Combine the results using Rubin's rules.[12] These rules take into account the within and between imputation variability, so that the uncertainty associated with imputation is accounted for. These steps are implemented in many statistical packages. For an excellent review of software see Horton and Kleinman.[13]

5. Analysing longitudinal data

The following briefly outlines the important points covered in detail in Bell & Fairclough [3].

5.1 Approaches to avoid

5.1.1 Complete case analysis

When only those participants with complete data for all assessments are analysed, it is called a complete case analysis. Multivariate analysis of variance (MANOVA) and repeated measures ANOVA (depending on software) are both forms of a complete case analysis. Estimates will be unbiased only if data are MCAR, but even then, a lot of data is thrown away, which is unethical and statistically inefficient.

5.1.2 Repeated univariate analyses

Repeatedly testing at each time-point is common but suboptimal. In addition to producing biased results (if data are not MCAR) by ignoring observed data at other points in time, it also does not take advantage of the longitudinal nature of the data, by comparing different groups of patients at each time-point.

5.1.3 Unweighted Generalised Estimating Equations

GEE is a frequentist method which can be used to analyse normal and non-normal longitudinal data.[14-16] If data are not MCAR results can be biased.[17] However, in this case weighted methods are often recommended (see below).[18, 19]

5.2 Approaches for unbiased estimation

5.2.1 Maximum likelihood estimation

Maximum likelihood refers to an estimation method which is used in various longitudinal models including mixed models, latent variable modelling, item response theory and structural equation models. Estimates obtained through maximum likelihood are unbiased if data are MAR and the model has been specified correctly.[8, 9]

Mixed models are commonly used for longitudinal data because they allow for non-independent observations and a variable number of observations per patient. Information from the observed data is used to implicitly impute unobserved data. Time can be continuous or categorical. When time is included categorically the model is sometimes referred to as a mixed model for repeated measures (MMRM), means model, response profile analysis or saturated model. The MMRM is a special case of a means model which models all the covariance within subject using an unstructured covariance.[20] The generality of this model, both in the mean and covariance structure, protects against misspecification to all but the missing data mechanism and because of this has been recommended for use in the regulatory environment.[20] When time is continuous these models

have been referred to as growth curve models (GC), linear mixed models and mixed effect regression models. An excellent and readable text is [21].

5.2.2 GEE Extensions: MI-GEE, Weighted GEEs, Doubly Robust GEEs

While estimates obtained by unweighted GEEs are biased for data which are not MCAR, it is still possible to use GEE either with weights, multiple imputation (MI), or both: a procedure called doubly robust estimation.[19, 22, 23] Modelling with weighted GEEs (WGEE) is a two-step process. The first step is to model the probability of being observed to obtain predicted probabilities for each patient. The second step is to fit a GEE, using the inverse of these probabilities as weights. Only observed data are used, but are weighted to account for those who drop out. Estimates from weighted GEEs are unbiased for data which are MAR, provided the weight model and the longitudinal outcome model are specified correctly. It is necessary to assume an independent "working covariance" in these models to ensure that the weights are correctly incorporated.[21]

Doubly robust GEEs (DR-GEE) combine inverse probability weighting with MI. They require correct specification of the weight model or the imputation model, but not necessarily both, and will yield unbiased estimation for MAR as long as there are no unmeasured confounders.[19, 23, 24]

5.3 The Intention to treat principle

The intention to treat (ITT) principle states that all patients are analysed, and analysed in the group to which they were randomised. ITT maintains the benefits of randomisation, i.e., that the trial arms are similar except for intervention. Randomisation is why the RCT is the gold standard of study designs. It also accounts for lack of adherence by patients as well as protocol deviations by staff, thereby enhancing trial external validity. [25] Efficacy analyses, sometimes called per-protocol or on-treatment analyses, can introduce substantial bias.[26]

When data are missing sometimes researchers are at a loss as to how to follow the ITT principle [27], and use methods that can introduce bias or inflate type I errors, such as simple imputation or the responder analysis (see below). However, maximum likelihood mixed model analysis, multiple imputation and the GEE extensions discussed above are consistent with the intention to treat principle, as discussed in references [10, 28].

6. Responder analysis

Responder analysis is an approach that deals with missing data by translating the PRO into a binary outcome defined by whether the participant achieved a clinically important response, such as an improvement of at least the minimally important difference in HRQoL. Participants who dropout or die are coded as a treatment failure, and missing data is thus reduced or eliminated. In some cases, it may be a clinically meaning way to define the trial endpoint; e.g. for a trial comparing alternative palliative interventions for advanced cancer, it may be reasonable to define and endpoint achieving a pre-defined degree of change (sometimes within a predefined period, or required to persist for a pre-defined duration). It has been recommended when missing data exceed 20% or when the best method for imputing PRO data is uncertain [29].

There are considerable drawbacks to responder analysis, including substantial loss of power, the use of arbitrary cutoff values which define a response, biased estimation, and the potential to mislead. A response may be due to regression to the mean, measurement error, the natural history of disease, or other concurrent therapies[30-33]. Simulation studies have shown misclassification of individuals of approximately 30% [31]. Snappin et al [33] compared the sample size required for adequate power to detect meaningful differences between a test in mean differences versus the proportion of patients who responded. The responder analysis had lower power, and in one case required 8

times as many patients. While this is an extreme example, Senn [34] states that an increase of at least 40% in sample size is required when dichotomising, as compared to keeping a measurement continuous. Loss of power may not be an issue when HRQoL is a secondary outcome to a dichotomous primary outcome such as death or disease progression, because sample sizes for the latter are typically larger than those required for analysis of HRQoL as a continuous measure.

7. Multiple comparisons & multiplicity

Multiplicity arises in HRQoL analysis because 1) there are often multiple outcomes (such as overall HRQoL, physical, social, emotional, and functional well-being, plus symptoms) and 2) HRQoL is often measured at more than one time-point. Multiple testing increases the likelihood of a type I error (a false positive result). There are three main strategies towards reducing this risk[5]: 1) limiting the number of hypothesis tests by focussing on key domains and time points; 2) using summary statistics, such as area under the time-curve; and 3) multiple comparison procedures, which includes adjustment to the alpha level of the tests. Whichever approach is taken it should be pre-specified. More information can be found in many texts, including Fairclough [5].

8. Sensitivity analysis

Sensitivity analysis is a broad term for seeing how results change when assumptions are varied. In the HRQoL analysis context, this often refers to changing the assumptions about missing data. All models for missing data, whether MCAR, MAR or MNAR, rest upon strong assumptions and can lead the researcher astray if their assumptions or model(s) are incorrect. Thus it is important to try different, clinically plausible models which use different assumptions and see how the estimates change. A thorough and sensible sensitivity analysis is an important step in producing and reporting robust estimates.[7, 8, 28]

If there is a substantial amount of missing data (say, greater than 10%, although there are no hard and fast rules), and one suspects that data may be MNAR, a MNAR model may be fit. MNAR models include pattern mixture models, selection models, shared parameter models, and joint multivariate models. [3, 5, 8]. However, each of these models make strong, untestable assumptions, and lack of fit for a particular MNAR models does not indicate that the data are MNAR.

A simpler approach than MNAR models for sensitivity analyses may be to carry out the following imputations: 1) the worst score imputed for all; 2) the best score imputed for all; 3) the worst score imputed for the active group and the best score imputed for the control group; and 4) the best score imputed for the active group and the worst score imputed for the control group.[35] A sensitivity analysis that might be used with a responder analysis is to vary the cut-off slightly, using the standard error of measurement ($SEM = \sqrt{s(1-r)}$, where r = reliability) on an individual as guidance.

9. Reporting

PRO outcomes from clinical trials should be reported as for other trial outcomes, as described in the CONSORT 2010 statement[36]. A PRO extension of the CONSORT 2010 is currently under review [37]. This extension provides guidance regarding the application of the existing CONSORT 2010 checklist items to the reporting of PROs in RCTs. The PRO specific checklist items are: the PRO should be identified as an outcome in the abstract, the PRO hypothesis and identification of relevant domains (when a multidimensional PRO has been used) should be described, evidence of PRO instrument validity and reliability should be provided or cited if available, the statistical approaches

for dealing with missing data should be explicitly stated and PRO specific limitations and implications for generalisability and clinical practice should be discussed.

As for other outcomes, confidence intervals should be reported, as an informative adjunct to p values for both reporting and interpreting PRO results [38].

10. Interpretation

Interpreting HRQoL results can be challenging for various reasons. They are often multidimensional, including disease symptoms, treatment side-effects and aspects of functioning. Different domains may have different trajectories relative to treatment duration; some PROs may reflect benefits of treatment (e.g. due to disease palliation) while others may reflect harms (e.g. due to treatment toxicity). Some PROs may not be affected by treatment at all, or may be affected similarly in both/all arms of the trial, and thus will not contribute anything to treatment comparisons. Therefore, when interpreting PRO results, it is useful to consider the direction (improvement or deterioration) and timing of treatment effects across all domains and timepoints.

As noted above, multiplicity of domain, timepoints and hypothesis tests should be considered when interpreting p values (to reject or accept the null hypothesis). It is best to focus on the domains and timepoints that are most important for the interventions and patient population in the trial. Ideally these will have been specified *a priori* in the trial protocol, and be the focus of analysis.

The other issue is the difficulty of interpreting the size of effects on HRQoL scales. That is, distinguishing clinical importance from statistical significance. The minimally important difference (MID) is an important concept in the PRO literature, because it provides a threshold above which a clinically important difference is deemed to have been demonstrated. There are various methods for estimating MIDs [39], and there are an increasing number of papers reporting MIDs. There is no global MID, although an effect size of between 0.2 and 0.5 may provide a useful ballpark guideline [1]. For a particular HRQoL instrument or scale, the MID is not an immutable characteristic, but may vary by population and context. Therefore, when interpreting your PRO results, search for MID estimates that match the instrument and patient population in your trial. Use confidence intervals (CI) in conjunction with the MID; if your lower CI includes the MID, you have a definitive trial. Responder graphs can be a means of presenting PRO results in a clinically accessible and meaningful way [40] but the primary analysis should be based on the more sensitive original scale[40].As noted above, responder analysis may not be the optimal approach to statistical analysis of PRO data. As Lewis puts it, “Responder analysis is a means of estimating the practical clinical consequences of effects shown to be statistically significant by other means, but no more than this.”[41]

11. Summary

QoL data can play a significant part in cancer trials. They can inform researchers about choosing the best treatment, e.g., where two treatment regimes may be expected to give similar survival outcomes, but have differing side effect profiles; for understanding the patient experience, e.g., for offering counselling; and for improving clinical trials, e.g. when PROs are used in prognostic modelling [42]. Although it may be tempting to keep HRQoL analysis as simple as possible, simple approaches such as complete case analysis, simple imputation and responder analysis have the potential to mislead.

Our view, along with others in the missing data field [8, 10, 20, 21, 43], is that a sensible approach is to base the primary analysis on MAR assumption, and then to use MNAR models, if possible, for sensitivity analyses. Researchers and journal editors may be uncomfortable with more than one set of results, and results from the sensitivity analysis may have to go in the discussion and electronic

appendices, however, one can't be certain about results when there are substantial amounts of missing data without these analyses. Caution must be exercised in interpretation of results, and the inherent uncertainty in analysing patient reported data where some are missing needs to be acknowledged.

REFERENCES

1. King, M., *A point of minimal important difference (MID): A critique of terminology and methods.* Expert Review of Pharmacoeconomics & Outcomes Research,, 2011. **11**(2): p. 171-184.
2. Administration., F.a.D., *Guidance for industry on patient reported outcome measures: Use in medical product development to support labeling claims.* Federal Register, 2009. **74**: p. 65132-3.
3. Bell, M.L. and D. Fairclough, *Practical and statistical issues in missing data for longitudinal patient reported outcomes.* Statistical Methods in Medical Research, 2012. **In Press**.
4. Fairclough, D.L. and D.F. Cella, *Functional assessment of cancer therapy (FACT-G): Non-response to individual questions.* Quality of Life Research, 1996. **5**(3): p. 321-329.
5. Fairclough, D.F., *Design and analysis of quality of life studies in clincial trials*. 2nd ed. 2010, Boca Raton, Fl: Chapman & Hall/CRC.
6. Fayers, P.M., D. Curran, and D. Machin, *Incomplete quality of life data in randomized trials: Missing items.* Statistics in Medicine, 1998. **17**(5-7): p. 679-696.
7. National Research Council, *The Prevention and Treatment of Missing Data in Clinical Trials*, in *Committee on National Statistics, Division of Behavioral and Social Sciences and Education*, Panel on Handling Missing Data in Clinical Trials, Editor 2010, National Academies Press Washington DC.
8. Carpenter, J. and M. Kenward, *Missing data in randomised controlled trials - a practical guide*, 2008, National Institute for Health Research: Birmingham.
9. Rubin, D.B., *Inference and missing data.* Biometrika, 1976. **63**(3): p. 581-592.
10. Molenberghs, G., et al., *Analyzing incomplete longitudinal clinical trial data.* Biostatistics, 2004. **5**(3): p. 445-464.
11. Schafer, J.L., *Multiple imputation: A primer.* Statistical Methods in Medical Research, 1999. **8**(1): p. 3-15.
12. Rubin, D.B., *Multiple Imputation for Nonresponse in Surveys* 1987, New York: J. Wiley & Sons.
13. Horton, N. and K. Kleinman, *Much ado about nothing: A comparison of missing data methods and software to fit incomplete data regression models.* American Statistician, 2007. **61**: p. 79-90.
14. Liang, K.Y. and S.L. Zeger, *Longitudinal data analysis using generalized linear models.* Biometrika, 1986. **73**(1): p. 13-22.
15. Burton, P., L. Gurrin, and P. Sly, *Extending the simple linear regression model to account for correlated responses: An introduction to generalized estimating equations and multi-level mixed modelling.* Statistics in Medicine, 1998. **17**(11): p. 1261-1291.
16. Hanley, J.A., et al., *Statistical analysis of correlated data using generalized estimating equations: An orientation.* American Journal of Epidemiology, 2003. **157**(4): p. 364-375.
17. Lipsitz, S.R., et al., *GEE with Gaussian estimation of the correlations when data are incomplete.* Biometrics, 2000. **56**(2): p. 528-536.
18. Robins, J.M., A. Rotnitzky, and L.P. Zhao, *Analysis of semiparametric regression models for repeated outcomes in the presence of missing data.* Journal of the American Statistical Association, 1995. **90**: p. 106-121.

19. Birhanu, T., et al., *Doubly robust and multiple-imputation-based generalized estimating equations.* Journal of Biopharmaceutical Statistics, 2011. **21**(2): p. 202-225.
20. Mallinckrodt, C.H., et al., *Assessing response profiles from incomplete longitudinal clinical trial data under regulatory considerations.* Journal of Biopharmaceutical Statistics, 2003. **13**(2): p. 179-190.
21. Fitzmaurice, G.M., N.M. Laird, and J.H. Ware, *Applied Longitudinal Analysis*. 2nd ed. 2011, Hoboken NJ: Wiley.
22. Bang, H. and J.M. Robins, *Doubly robust estimation in missing data and causal inference models.* Biometrics, 2005. **61**(4): p. 962-972.
23. Carpenter, J.R., M.G. Kenward, and S. Vansteelandt, *A comparison of multiple imputation and doubly robust estimation for analyses with missing data.* Journal of the Royal Statistical Society. Series A: Statistics in Society, 2006. **169**(3): p. 571-584.
24. Seaman, S. and A. Copas, *Doubly robust generalized estimating equations for longitudinal data.* Statistics in Medicine, 2009. **28**(6): p. 937-955.
25. Hollis, S. and F. Campbell, *What is meant by intention to treat analysis? Survey of published randomised controlled trials.* British Medical Journal, 1999. **319**(7211): p. 670-674.
26. Lachin, J.M., *Statistical considerations in the intent-to-treat principle.* Controlled Clinical Trials, 2000. **21**(3): p. 167-189.
27. Altman, D.G., *Missing outcomes in randomized trials: Addressing the dilemma.* Open Medicine, 2009. **3**(2).
28. White, I.R., et al., *Strategy for intention to treat analysis in randomised trials with missing outcome data.* BMJ, 2011. **342**.
29. Sloan, J.A., et al., *Analysis and interpretation of results based on patient-reported outcomes.* Value in Health, 2007. **10**(SUPPL. 2): p. S106-S115.
30. Kelley, J.M. and T.J. Kaptchuk, *Group analysis versus individual response: The inferential limits of randomized controlled trials.* Contemporary Clinical Trials, 2010. **31**(5): p. 423-428.
31. Kunz, M., *On responder analyses when a continuous variable is dichotomized and measurement error is present.* Biometrical Journal, 2011. **53**(1): p. 137-155.
32. Senn, S., *Individual response to treatment: Is it a valid assumption?* British Medical Journal, 2004. **329**(7472): p. 966-968.
33. Snapinn, S.M. and Q. Jiang, *Responder analyses and the assessment of a clinically relevant treatment effect.* Trials, 2007. **8**.
34. Senn, S., *Disappointing dichotomies.* Pharmaceutical Statistics, 2003. **2**(4): p. 239-240.
35. Berger, V.W., *Conservative handling of missing data.* Contemporary Clinical Trials, 2012. **33**(3): p. 460.
36. Moher, D., et al., *CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials.* BMJ (Clinical research ed.), 2010. **340**.
37. Calvert, M., et al., *Improved Reporting of Patient Reported Outcomes in Randomised Trials: the CONSORT PRO 2012 Statement*, 2012.
38. Guyatt, G., et al., *Basic statistics for clinicians: 2. Interpreting study results: Confidence intervals.* CMAJ, 1995. **152**(2): p. 169-173.
39. Revicki, D., et al., *Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes.* Journal of Clinical Epidemiology, 2008. **61**(2): p. 102-109.
40. Guyatt, G. and H. Schunemann, *How can quality of life researchers make their work more useful to health workers and their patients?* Quality of Life Research, 2007. **16**(7): p. 1097-1105.
41. Lewis, J.A., *In defence of the dichotomy.* Pharmaceutical Statistics, 2004. **3**(2): p. 77-79.
42. Au HJ, R.J., Brundage M, Palmer M, Richardson H, Meyer RM; NCIC CTG Quality of Life Committee, *Added value of health-related quality of life measurement in cancer clinical trials: the experience of the NCIC CTG.* Expert Review of Pharmacoeconomics and Outcomes Research, 2010. **10**(2): p. 119-128.
43. Molenberghs, G. and M.G. Kenward, *Missing Data in Clinical Studies.* Missing Data in Clinical Studies, 2007.