Appendix 3

Table 1. SISAQOL non-ratified statements and their considerations

No.	Non-ratified statement (NRS)	Status	Considerations
NRS 1	For evaluating a proportion of patients (with an improvement, stable state or worsening) at time t, we recommend the Cochran Mantel-Haenszel test, logistic mixed model, simple logistic regression model.	POSTPONED	Please refer to Appendix 2 (Table 3.b) to find more details on how the statistical methods were evaluated based on the agreed set of criteria. The logistic mixed model, an extension of the linear mixed model, was proposed as alternative because of the less favorable evaluation of the (Cochran) Mantel-Haenszel test on the set criteria. The mixed model will provide an unbiased estimate of the treatment effect if missing data is dependent on known and observed factors [1], whereas the (Cochran) Mantel-Haenszel test is based on observed cases data [2] and thus only provides valid inference when missing data are missing completely at random. There were reservations for recommending the logistic mixed model due to practical considerations that limit the use of these models [3], including convergence issues. To address this potential limitation, the simpler logistic model was also proposed. The decision whether a logistic mixed model, a (Cochrane)-Mantel Haenszel test or a simple logistic model will be recommended was postponed until these methods are further explored.
NRS 2	PRO assessments are no longer expected from patients who are off the PRO protocol.	POSTPONED	There was variation in calculating the variable denominator rate. To standardize the denominator of this rate, it was agreed to standardize reasons for patients going off PRO assessment (i.e. patients from whom we do not expect PRO assessments anymore). The implication is that these reasons are not seen as missing data, because PRO assessments are not expected from these patients anymore. Off PRO protocol: The protocol describes details on timing and planning of PRO assessments. Under the assumption that the PRO assessment schedule reflects the PRO trial objectives [4] (and thus reflecting what is meaningful for PRO analysis), it was proposed to consider assessments from patients off the PRO protocol as no longer expected because these assessments are not "meaningful for analysis". This means that assessments from patients off PRO protocol do not have a relevant contribution to the PRO estimate.
NRS 3	PRO assessments are no longer expected from patients who explicitly withdraw consent from the PRO study.	POSTPONED	
NRS 4	PRO assessments are no longer expected from patients who are lost to follow-up.	POSTPONED	

No.	Non-ratified statement (NRS)	Status	Considerations
			Withdrawing consent: The distinction was made between (a) a patient refusing to complete one or more PRO assessments (e.g. due to patient being too sick, questionnaire too long, ..) [5] and (b) a patient refusing (to continue) participation in the PRO study, referred to as PRO withdrawal. In the case of PRO refusal (a), the patient refuses one or more PRO assessments, but is still on PRO study. In the latter (b), the patient explicitly and voluntarily terminates informed consent to participate in the PRO study (or the broader clinical trial), for whatever reason [6], entailing that the patient is (no longer) on PRO study. It was proposed to consider assessments from patients withdrawing consent from the PRO study as off PRO study. Assessments from patients off PRO study are no longer to be collected and thus no longer to be expected. Loss to follow-up: Being lost to follow-up was proposed as a possible reason that can lead a patient into being off PRO study and thus off PRO assessment. The definition of loss to follow-up is vaguely defined as the loss of participants during the course of a study [7]. As a consequence, great variability exists concerning the definition of loss to follow-up in the literature [8]. It was decided to postpone the voting on this proposed statement until agreement is reached on a definition for being lost to follow-up. It was difficult to agree whether the above reasons should be considered as missing data or not, depending on the different trial settings. Further discussion on the consequences of categorizing these reasons as being off PRO assessment are needed.
NRS 5	We should establish percentage boundaries for missing data.	REJECTED	There is currently no standard rule of how much missing data is too much [9]. To address this question, the possibility of having percentage boundaries for missing data was proposed (e.g. statistical inference is not recommended with missing data rates above 50% and caution is required with missing data rates are between 10 and 50%). Monte Carlo simulations showed mixed results on bias and power in a typical superiority RCT, depending on a number of factors such as missing data mechanism, choice of analysis method and sample size (reference Gina paper). Based on these results, it was discussed that it is not possible to have one overall cut-off value (e.g. the impact of 40% missing data in a trial with

No.	Non-ratified statement (NRS)	Status	Considerations
			10 patients is higher than in a trial with 25000 patients or the acceptance threshold might depend on whether the disease stage is early, advanced or chronic). It was therefore agreed NOT to establish percentage boundaries for how much missing data is too much when evaluating PRO outcomes. Sensitivity analyses were suggested as an alternative way to assess the impact of missing data on PRO findings (see CS 32 on the specification of sensitivity analyses in the protocol/statistical analysis plan).
NRS 6	The lower boundary of the missing data rate should be 10/15%, meaning that a missing data rate of 10/15% or less is unlikely to substantially bias a between-arm PRO analysis	CANCELLED	Based on the outcome of NRS 5, the voting on a proposal of actual missing data thresholds was cancelled.
NRS 7	The upper boundary of the missing data rate should be 50%, meaning that we would question the data quality in a between-arm PRO analysis with a missing data rate above 50%	CANCELLED	Based on the outcome of NRS 5, the voting on a proposal of actual missing data thresholds was cancelled.
NRS 8	Agreement with modifications to the proposed case report form (CRF)?	POSTPONED	Results from a simulation study showed that the impact of missing data rates on PRO findings depends on the reasons for missing data (e.g., informative, non-informative or a combination of both; Gina paper). Therefore collecting reasons for missing data is key in assessing the impact of missing data rates on the robustness of PRO findings. Ideally the reason for missing data should be identified to provide more information on the possible impact of missing data and how it should be handled. This way, the level to which results may be biased can be assessed [4] and the most appropriate analysis method can be identified [10]. It was decided to develop a template for capturing these reasons of missingness, to be used in PRO reports. A standard case report form (CRF) to be administered by clinical staff during PRO collection with reasons of missingness was proposed.

No.	Non-ratified statement (NRS)	Status	Considerations
			After expression of concern for staff burden, it was decided that further fine-tuning of the proposed template is needed. Ratification of a final template for collecting reasons of PRO non-completion was postponed.
NRS 9	Agreement with collecting the question 'Is the reason for non-completion related to the patient's health?'	POSTPONED	To assess whether the collected reason for non-completion of the PRO assessment is related to the outcome variable - and thus to determine the underlying missing data mechanism -, the inclusion of the question 'is the reason for non-*completion related to the patient's health*' was proposed. The utility of this item was however questioned, as it was not sure we could ultimately rely on this data. To avoid redundancy and capture of unreliable data [11], it was decided to further assess the utility of this item before inclusion in the standard template for capturing reasons for PRO non-completion. It was decided to postpone the voting on this proposed statement.
NRS 10	Do you agree that the reasons in the proposed CRF for non-completion are easy for research personnel to understand?	POSTPONED	The design of the case report form is key for ensuring the quality of the data collected by the CRF. Guidelines for CRF design state that CRF design should address the needs of all users and the language used should be simple and easy to understand [11]. Based on the outcome of NRS 8, it was decided to await a more developed template before evaluating whether the reasons in the CRF are easy for research personnel.
NRS 11	Do you agree that research personnel can successfully complete this CRF?	POSTPONED	Based on the outcome of NRS 8, it was decided to await a more developed template before evaluating whether the reasons in the CRF are easy for research personnel.

Table 2. Summary of proposed statements and voting results.

Outcome[1]	Proposed statement	Absolute number of votes					Agreement[2] (in %)
		Agree	Dis-agree	Abstain/ no vote	Total incl. abstain	Total excl. abstain	
Taxonomy of Research Objectives							
RATIFIED	1.Two broad PRO research objectives: (1) treatment efficacy/clinical benefit (2) describe patient perspective	30	0	1	31	30	100 %
RATIFIED	2. Clearly state that the PRO domain/item of interest will be used to provide evidence for pre-specifying superiority, equivalence and non-inferiority	30	0	1	31	30	100 %
RATIFIED	3. Taxonomy of PRO objectives: Valid PRO objectives for treatment efficacy/clinical benefit at the within-individual / within-treatment level (for each pre-specified domain) are: - Improvement (time to improvement, proportion of patients with improvement at time t, magnitude of improvement at time t) - Worsening (time to worsening, proportion of patients with worsening at time t, magnitude of worsening at time t) - (End of) stable state (time to end of stable state, proportion of patients with stable state at time t)	30	0	1	31	30	100 %
RATIFIED	4. Taxonomy of PRO objectives: A valid PRO objective for treatment efficacy/clinical benefit at the within-individual/within-treatment level (for each pre-specified domain) is the overall effect: overall PRO score over time.	28	1	2	31	29	97 %
RATIFIED	5. Taxonomy of PRO objectives: A valid PRO objective for treatment efficacy/clinical benefit at the within-individual/within-treatment level (for each pre-specified domain) is the overall	30	0	1	31	30	100 %

	effect: describing response trajectory over time (response patterns/profiles)						
RATIFIED	6. Definition of Improvement: change from baseline that reaches a pre-defined improvement threshold level (post-baseline improvement). This improvement is maintained if follow-up assessments remain at or are higher than the improvement threshold (definitive improvement). Improvement is discontinued once a follow-up assessment is below the improvement threshold (transient improvement)	30	0	1	31	30	100 %
RATIFIED	7. Definition of Worsening: change from baseline that reaches a pre-defined worsening threshold level (post-baseline worsening). This worsening is maintained if follow-up assessments remain at or are lower than the worsening threshold (definitive worsening). Worsening is discontinued once a follow-up assessment is above the worsening threshold (transient worsening)	30	0	1	31	30	100 %
RATIFIED	8. Definition of Stable State: no change from baseline is observed, or change from baseline is within the pre-defined baseline margin. This stable state is maintained if follow-up assessments remain at the baseline pre-defined margin. The stable state is discontinued once the follow-up assessment leaves the pre-defined baseline margin (and reaches the improvement or worsening threshold)	27	3	1	31	30	90 %
RATIFIED	9. Definition of the broad 'overall effects': summarize all available scores over time for each patient on a specific PRO domain/item	25	2	4	31	27	93 %
Recommending Statistical Methods							
RATIFIED	10. Essential statistical features for analyzing PRO data are:	30	0	1	31	30	100 %

	• ability to perform a statistical test between two samples • ability to produce clinically relevant results Highly desirable statistical features are: • ability to adjust for covariates, including baseline PRO score • ability to handle missing data with the least restrictions • ability to handle clustered data (repeated assessments)						
RATIFIED	11: For evaluating time to event (*improvement, stable state or worsening)* outcomes, the Cox proportional hazards instead of the log rank test is recommended.	23	0	8	31	23	100 %
RATIFIED	12: For evaluating the magnitude of event (*improvement, stable state or worsening)* at time t, the linear mixed model (time as discrete variable) is recommended	26	1	4	31	27	96 %
RATIFIED	13: For evaluating the magnitude of event at time t (simplified case where only 1 FU assessment available by design), linear regression is recommended	28	0	3	31	28	100 %
POSTPONED	14: For evaluating a proportion of patients *(with an improvement, stable state or worsening)* at time t, we recommend the Cochran Mantel-Haenszel test/logistic mixed model?	/	/	/	/	/	/
RATIFIED	15: Summary measures should be part of SISAQOL (as a way to assess overall effects)	16	4	11	31	20	80 %
RATIFIED	16: For describing a response trajectory over time (as a way to assess overall effects), it is recommended to use a linear mixed model (omnibus test; time as discrete variable; time*group interaction) over the repeated measures ANOVA (time*group interaction)	27	0	4	31	27	100 %
Standardizing Statistical Terminology							

RATIFIED	17: Definition of missing data: Missing data are data that would be meaningful for the analysis of a given research objective or estimand, but were not collected	30	0	1	31	30	100 %
RATIFIED	18: ”Meaningful for analysis” refers to the PRO analysis population, which is based on the given research objective or estimand	30	0	1	31	30	100 %
RATIFIED	19: We are not expecting data anymore from patients who have died (although these patients were part of the PRO study population)	29	0	2	31	29	100 %
POSTPONED	20: We are not expecting data anymore from patients who are off the PRO protocol	/	/	/	/	/	/
POSTPONED	21: We are not expecting data anymore from patients who explicitly withdraw consent from the PRO study	/	/	/	/	/	/
POSTPONED	22: We are no longer expecting data from patients who are lost to follow-up	/	/	/	/	/	/
RATIFIED	23: Calculation of the ‘variable’ denominator rate: Numerator as ‘number of patients on PRO assessment submitting the PRO assessment at the designated time point’ and denominator as ‘Number of patients on PRO assessment at the designated time point’.	30	0	1	31	30	100 %
RATIFIED	24: Calculation of the ‘fixed’ denominator rate: Numerator as ‘number of patients on PRO assessment submitting the PRO assessment at the designated time point’ and denominator as ‘number of patients in the PRO study population (all patients who consented and were eligible to participate in the PRO data collection)’.	28	0	3	31	28	100 %
RATIFIED	25: Reporting of completion/compliance rates: In addition to percentages, absolute numbers for both numerator and	30	0	1	31	30	100 %

	denominator should be reported at every time point (for both rates)						
RATIFIED	26: The term 'completion rate' should be used to express the rate with the variable denominator rate.	30	0	1	31	30	100 %
RATIFIED	27: The term 'available data rate' should be used to express the rate with the fixed denominator rate.	25	1	5	31	26	96 %
Missing Data							
RATIFIED	28: When conducting clinical trials, exploring the reasons for missing PROs is important.	30	0	1	31	30	100 %
REJECTED	29: We should establish percentage boundaries for missing data.	5	17	9	31	22	23 %
CANCELLED	30: The lower boundary of the missing data rate should be 10/15%, meaning that a missing data rate of 10/15% or less is unlikely to substantially bias a between-arm PRO analysis	/	/	/	/	/	/
CANCELLED	31: The upper boundary of the missing data rate should be 50%, meaning that we would question the data quality in a between-arm PRO analysis with a missing data rate above 50%	/	/	/	/	/	/
POSTPONED	32: Agreement with modifications to the proposed CRF?	/	/	/	/	/	/
POSTPONED	33: Agreement with collecting the question 'Is the reason for non-completion related to the patient's health?'	/	/	/	/	/	/
POSTPONED	34: Do you agree that the reasons in the proposed CRF for non-completion are easy for research personnel to understand?	/	/	/	/	/	/
POSTPONED	35: Do you agree that research personnel can successfully complete this CRF?	/	/	/	/	/	/
RATIFIED	36: Minimize missing data prospectively through clinical trial and PRO design strategies and by training/monitoring approaches	29	0	2	31	29	100 %
RATIFIED	37: We recommend capturing data that will be needed for handling missing PRO data prospectively in the statistical analysis plan (i.e. reasons for missing data and auxiliary data for interpretation/imputation)	29	0	2	31	29	100 %

RATIFIED	38: Primary statistical analysis approach: Missing data approach at the item- and scale-level should be specified a priori within the protocol/statistical analysis plan	29	0	2	31	29	100 %
RATIFIED	39: Primary statistical analysis approach: Critical assessment of missing data reasons and rates (by arm and time point) should be undertaken	29	0	2	31	29	100 %
RATIFIED	40: Primary statistical analysis approach: Item-level missing data within a scale should be handled according to the scoring algorithm developed during the scale's development (when available)	28	0	3	31	28	100 %
RATIFIED	41: Primary statistical analysis approach: Use all available data, using the specified method from Statistical Methods WG Recommendations	29	0	2	31	29	100 %
RATIFIED	42: Primary statistical analysis approach: Explicit imputation is not recommended unless justified within the context of the clinical trial	29	0	2	31	29	100 %
RATIFIED	43: Sensitivity analyses should be specified a priori within the protocol/statistical analysis plan. Use of at least two different approaches to handle missing data is recommended to assess impact of missing data across various assumptions	26	1	4	31	27	96 %

[1]Four possible outcomes for the proposed statements: ratified, rejected, cancelled or postponed.

RATIFIED: At least two third agreed with the proposed statement.

REJECTED: More than half disagreed with the proposed statement.

CANCELLED: Voting for the proposed statement was cancelled because the statement was made obsolete due to the preceding votes or discussions.

POSTPONED: Voting for the proposed statement was postponed because the statement has to be further explored /discussed first.

[2]Agreement (in %) is calculated as the number of green votes divided by the total number of green and red votes (abstain excluded).

1 Xu S, Blozis SA. Sensitivity Analysis of Mixed Models for Incomplete Longitudinal Data. J Educ Behav Stat 2011; **36**: 237–56.

2 Ali MW, Talukder E. Analysis of Longitudinal Binary Data with Missing Data Due to Dropouts. J Biopharm Stat 2005; **15**: 993–1007.

3 Booth JG, Hobert JP. Maximizing generalized linear mixed model likelihoods with an automated Monte Carlo EM algorithm. J R Stat Soc Ser B Stat Methodol 1999; **61**: 265–85.

4 Calvert MJ, Freemantle N. Use of health-related quality of life in prescribing research. Part 2: methodological considerations for the assessment of health-related quality of life in clinical trials. J Clin Pharm Ther 2004; **29**: 85–94.

5 Atherton PJ, Burger KN, Pederson LD, Kaggal S, Sloan JA. Patient-reported outcomes questionnaire compliance in Cancer Cooperative Group Trials (Alliance N0992). Clin Trials 2016; **13**: 612–20.

6 Gabriel AP, Mercado CP. Data retention after a patient withdraws consent in clinical trials. Open Access J Clin Trials 2011; **3**: 15.

7 Higgins JPT, Green S. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. 2011.

8 Vervölgyi E, Kromp M, Skipka G, Bender R, Kaiser T. Reporting of loss to follow-up information in randomised controlled trials with time-to-event outcomes: a literature survey. BMC Med Res Methodol 2011; **11**: 130.

9 Papageorgiou G, Grant SW, Takkenberg JJM, Mokhles MM. Statistical primer: how to deal with missing data in scientific research?†. Interact Cardiovasc Thorac Surg 2018; **27**: 153–8.

10 Fielding S, Fayers PM, Ramsay CR. Investigating the missing data mechanism in quality of life outcomes: a comparison of approaches. Health Qual Life Outcomes 2009; **7**: 57.

11 Bellary S, Krishnankutty B, Latha MS. Basics of case report form designing in clinical research. Perspect Clin Res 2014; **5**: 159–66.