

## Appendix 7

Diverging views among stakeholders emerged during the consensus process; however, most considerations were addressed in the explanatory text accompanying the final recommendations.

Statement	% agreement*	Considerations	Implementation
EstFrame1_GEN	98% (range 50-100%)	It should be noted that different stakeholders may prefer different estimands.	This issue was discussed and has been included in the explanatory text accompanying the statement.
Pop2_RCT	88% (range 50-100%)	It is important to differentiate between comparative tolerability vs descriptive tolerability objectives, especially when handling intercurrent events (e.g., only looking at patients who are still on treatment).	In SISAQOL framework, tolerability is mostly considered a descriptive objective. However, the concept of 'comparative tolerability' is increasingly being explored as a confirmatory objective, which requires defining an appropriate endpoint, application of appropriate intercurrent event strategies and the use of suitable analysis sets.
ICEdisprog1_RCT	90% (range 50-100%)	<p>Collecting post-progression data is an aspirational goal that may not always be feasible in practice. If this is not feasible for a specific context, this could be a deviation that need to be justified.</p> <p>If there is a need to collect post-progression data, using magnitude of change at time t may be the least burdensome way to implement this since one would only have to collect this information once at that pre-defined time point.</p> <p>The estimation of the treatment effect including post-progression data will be influenced by subsequent treatment.</p> <p>It would be important to support the main analysis by performing supplementary/sensitivity analysis excluding the post-progression assessments.</p>	These issues were discussed and included in the explanatory text accompanying the statement.
ICEdisprog3_RCT	80% (range 33-100%)	<p>Collecting post-progression data is an aspirational goal that may not always be feasible in practice. If this is not feasible for a specific context, this could be a deviation that need to be justified. The feasibility of collecting post-progression data for time-to event outcomes need to be further discussed.</p> <p>The estimation of the treatment effect including post-progression data will be influenced by subsequent treatment.</p> <p>It is relevant to assess PROs post progression even when the patients' HRQOL may not be related to the main treatment since patients may still experience on-going side effects from the treatment.</p>	These issues were discussed and included in the explanatory text accompanying the statement.

ICEdisprog2_RCT	75% (range 0-100%)	<p>Collecting post-progression data is an aspirational goal that may not always be feasible in practice.</p> <p>The estimation of the treatment effect including post-progression data will be influenced by subsequent treatment.</p>	These issues were discussed and included in the explanatory text accompanying the statement.
ICEdisprog4_RCT	90% (range 60-100%)	<p>Collecting post-progression data is an aspirational goal that may not always be feasible in practice.</p> <p>The estimation of the treatment effect including post-progression data will be influenced by subsequent treatment.</p>	These issues were discussed and included in the explanatory text accompanying the statement.
ICEdisc1_GEN	84% (range 57-100%)	<p>Following disease progression, patients typically receive additional treatments outside the scope of the study, leading to challenges in accounting for subsequent therapies. These subsequent treatments are not randomly assigned, posing risks related to the treatment policy estimand, which is not considered the optimal strategy for all cases.</p> <p>The opportunities for successfully collecting QoL data after progression, treatment discontinuation, or subsequent therapy are limited. Efforts to do so often result in low compliance and poor data quality.</p>	These issues were discussed and included in the explanatory text accompanying the statement.
ICE3_RCT	76% (range 50-100%)	<p>Some regulators did not agree with using non-inferiority (NI) objective for PRO-based endpoints due to risks of bias from insensitive measures, and poor data quality and/or variability that will tend to pull the result toward NI. They emphasized clarifying analytic approaches, aligning with guidelines like ICH E9, and tailoring strategies to specific intercurrent events. Ambiguities in defining analysis populations and assumptions must be addressed, as poor study quality can compromise the interpretation of NI results.</p>	This issue was discussed and included in the explanatory text accompanying the statement.
CMCimp8_GEN	87% (range 50-100%)	<p>It was proposed to use "may" instead of "should," as the use of "should" implies a requirement to justify not performing these analyses. Sensitivity analyses shouldn't be done if not needed.</p>	<p>This issue was discussed and included in the explanatory text accompanying the statement:</p> <p><i>"Depending on the available evidence supporting the validity of a given threshold value, the sensitivity analysis for the trial results may be more or less extensive."</i></p>
CMCrep3_GEN	87% (range 50-100%)	<p>Valuable elements, but their reporting should be encouraged, not be mandatory; they can be seen as part of the rationale for choosing a particular threshold.</p>	This issue was discussed and the explanatory text further clarified the rationale for the recommendation.
CMCrep4_GEN	95% (range 50-100%)	<p>Valuable elements, but their reporting should be encouraged, not be mandatory; they can be seen as part of the rationale for choosing a particular threshold.</p>	This issue was discussed and the explanatory text further clarified the rationale for the recommendation.
CMCselc6_GEN	82% (range 50-100%)	<p>Concerns were raised about the distinction between minimum thresholds and thresholds of larger magnitude, particularly when different methodologies provide varying sizes of thresholds.</p> <p>The inconsistent use of the term "minimal" in the field was highlighted.</p>	These issues were discussed and included in the explanatory text accompanying the statement.

EstFrame4_SAT	70% (range 0-100%)	Concerns regarding making comparisons in single-arm setting.	This issue was discussed and included in the explanatory text accompanying the statement.
Pop3_SAT	85% (range 50-100%)	It was recommended to exercise caution in using the term 'efficacy' and making causal claims for single-arm studies.	The issue was discussed; no changes implemented.
AnalPres3_SAT	88% (range 50-100%)	Concerns regarding making comparisons in single-arm setting, feasibility of the recommendation, and the need for more detailed statistical guidance.	This issue was discussed and included in the explanatory text accompanying the statement.
AnalMain9_SAT	95% (range 50-100%)	Concerns regarding making comparisons in single-arm setting, feasibility of the recommendation, and the need for more detailed statistical guidance.	This issue was discussed and included in the explanatory text accompanying the statement.
ICEdeath3_SAT	70% (range 50-100%)	This necessitates identifying the specific situations or contexts in which this approach would be valid.	This issue was discussed and included in the explanatory text accompanying the statement.
ICEdeath4_SAT	71% (range 50-100%)	This necessitates identifying the specific situations or contexts in which this approach would be valid.	The issue was discussed; no changes implemented.
Design2_GEN	95% (range 50-100%)	There were concerns about the alignment between the recommendations for RCTs and single-arm studies.	This was resolved through harmonization between the two work packages.
ICE5_SAT	83% (range 60-100%)	Concerns about the feasibility of collecting data post-treatment discontinuation.  More guidance needed on what would be the alternative strategy when collecting post-treatment discontinuation data is not feasible.	The issue was discussed; no changes implemented.
ICEdeath2_SAT	76% (range 50-100%)	Concerns about the difficulty in assessing PRO results that incorporate death without a valid reference arm, or the relationship between death and the PRO.  In a regulatory context, the applicability of this statement is debatable as these endpoints in single arm trials would be challenging to include.	The issue was discussed; no changes implemented.
Assump2_SAT	84% (range 50-100%)	Clarifications in the explanation section are needed including alternative methods to use.	The issue was discussed; no changes implemented.
ICE1_GEN	86% (range 50-100%)	Caution on incorporating intercurrent event in the PRO outcome e.g., progression of disease is not regularly to be seen as a deterioration or equivalent to minimum PRO scores.  Time to event PRO endpoint for single-arm study have limited interpretability in regulatory context.  More emphasis was needed in the explanation to caution that this kind of combined outcomes can lead to biased interpretation especially when the intercurrent event is very prevalent.	These issues were discussed and included in the explanatory text accompanying the statement.
ICE2_GEN	79% (range 50-100%)	There should be a cautionary note here on the possible bias in interpreting the results if the intercurrent event happens frequently and dominates the endpoint.	This issue was discussed and included in the explanatory text accompanying the statement.
ICE4_SAT	74% (range 43-100%)	General concerns about the validity of the hypothetical strategy.	These issues were discussed and included in the explanatory text accompanying the statement.

		Hypothetical strategy has limited regulatory acceptance because of underlying modelling assumptions.	
VizSciType1_GEN	89% (range 50-100%)	Consistency is important in how data are shown to familiarize people with the data. It is important that people are comfortable with the data. Therefore, a limited number of recommendations is useful rather than many recommendations.	The issue was discussed; no changes implemented.
VizSciType2_GEN	84% (range 50-100%)	Line graphs should only be used when there is no relevant attrition.  If estimated data is plotted, confidence intervals should be included.  As line graphs might give the impression of inappropriate linearity between data points shown, data points should not be connected if any misconception is likely to happen (e.g., long time intervals between measurement, high attrition).	These issues were discussed and some of them has been included in the explanatory text accompanying the statement and some are included in other statements e.g., including confidence intervals.
VizSciType4_GEN	89% (range 50-100%)	Forest plots are not only used for visualizing effects for multiple PRO domains but can also be used to present sensitivity analysis for a single domain e.g., different effect sizes for one PRO domain.  Cautionary note about overinterpreting results based on forest plots. There is a need to ensure that multiplicity is taken into account if it will be used to make inferential conclusions.	The final statement refers to multiple effect sizes and does not specify whether it is for sensitivity analysis or multiple PRO domains.  Caution about multiplicity correction was addressed in other general statements for visualization.
VizSci12_GEN	88% (range 63-100%)	Why not other options (e.g., histograms, violin plots, cumulative distribution function, probability density function)? Flexibility should be allowed.  The intention of the recommendation should be to provide a possibility to show distributions, but this is not a strict recommendation.  The choice of the methods should be informed by the properties of the distribution, and distribution graphs are expected to be part of all standard statistical analysis. And this statement is about giving due consideration when to include any distribution graphs in a publication.	This is addressed by not mentioning a specific graph type, but referring only to "visualizations of distributions"
VizPlain4_GEN	95% (range 50-100%)	If properly explained, p-values are no problem.  Purposely avoiding reporting p-values limits the information conveyed to the reader.	The issue was discussed; no changes implemented.
VizPlainType1_GEN	92% (range 50-100%)	Consistency is important in how data are shown to familiarize people with the data. It is important that people are comfortable with the data. Therefore, a limited number of recommendations is useful rather than many recommendations.	The issue was discussed; no changes implemented.
VizPlainType3_GEN	92% (range 0-100%)	Unclear whether picking one time point can communicate the entire information.	The issue was discussed; no changes implemented.

		All items of the graph should be clearly explained and reviewed by the target audience. Unclear which option to choose for proportion for plain graphs given multiple suitable representations suggested.	
--	--	--	--

\* Level of agreement among all stakeholder groups combined with minimum and maximum in parenthesis. Details in the % of agreement by stakeholder group is given in *Appendix 8 - Table of diverging views*.