Appendix 1

Harmonised statements between RCTs and SATs

DIFFERENCES marked in blue

RCT adapted/adopted six statements from SAT

WP2/RCT	WP3/SAT	Note
Design2_GEN **Statement:** PRO data collection methods should be aligned with the strategies established at the start of the trial to handle ICEs. **Explanation:** when the strategy is to include PRO measurements even after an ICE in the analysis, all efforts should be made to continue PRO data collection until death or a pre-specified time point after treatment discontinuation. If this is not feasible, it should be justified beforehand in the protocol. During data collection, reasons for missing PRO data should be recorded. A main incentive for patients to complete questionnaires is the principle that all data they provide will lead to a meaningful contribution to the actual results. It is understood that not all data may be equally relevant for all objectives. If the PRO objective limits all analyses to pre-progression data, then PRO data after progression should not be collected. Limitations of the PRO objectives to pre-progression data may be intentional as part of the objective. However, it may also be justified by practical constraints such as when approaching patients post-progression may be logistically difficult.		identical
Design1_GEN **Statement:** it is important to identify beforehand which patient and disease characteristics are expected to be associated with the primary and key secondary endpoints. These should be considered in the analysis, and, therefore, must be recorded and reported. If there is a known core set of variables in the disease domain, efforts should be made to collect, evaluate, and report them all.		identical

WP2/RCT	WP3/SAT	Note
RCT Design1_GEN **Explanation:** the estimation of a causal effect requires appropriate adjustment for possible confounding factors. Therefore, it is important that these variables are available during the analysis to ensure proper and efficient treatment arm comparison. Even in a randomised setting, differences at baseline can occur or a selected analysis set may need to be used. Recording a core set of variables facilitates the comparison of the results of RCTs to external data sources. The core set of variables enables adjusting for corresponding potential confounding in the analysis phase when making such an external comparison. Although the set of variables is likely disease-dependant, a core set of variables typically includes, but is not limited to, age, sex, disease stage, treatment history and comorbidities. Availability of baseline or pre-treatment PRO data may help adjust in part for between-patient variations.	**SAT Design1_GEN** **Explanation:** the estimation of a causal effect requires appropriate adjustment for possible confounding factors. Therefore, it is important that these variables are available both in the single arm study and when the external control data are used, in the external control data to reduce bias in any comparison. Variables needed to perform subgroup analyses, to handle missing data or intercurrent events also need to be collected. A core set of variables is a set of possibly prognostic or predictive variables that is meant to be collected within every study in a particular disease domain. Recording a core set of variables facilitates the comparison of the results of single arm studies to other data sources. For example, the control arm of a previously conducted RCT could be used as historical control data. The core set of variables enables adjusting for corresponding potential confounding in the analysis phase when comparing between the groups. Although the set of key variables is likely disease-dependant, a core set of variables typically includes, but is not limited to, age, sex, disease stage, treatment history and comorbidities. Availability of baseline or pre-treatment PRO data may help adjust in part for between-patient variations.	Different due to setting (RCT vs SAT)
AnalPres1_GEN **Statement:** if the objective of the PRO is to evaluate treatment tolerability and the clinical focus is only on PROs while-on-treatment, PRO scores may be analysed at pre-specified time points using the subset of those still on treatment (while-on-treatment strategy). Alongside this analysis, the percentage of the study population who discontinued treatment, and the reasons for doing so, should be provided. **Explanation:** in certain situations, only the PRO scores of patients still under treatment are of interest, for instance, in PRO scores that assess treatment side effects, such as nausea. To accurately interpret the PRO scores among patients still on treatment, it is important to take into account the percentage of patients who discontinued treatment and their reasons for doing so. However, if there are notable tolerability issues after treatment discontinuation, such as long-term toxicities, the while-on-treatment estimand is not suitable and a different estimand strategy should be chosen.		identical

WP2/RCT	WP3/SAT	Note
ICE1_GEN **Statement:** when a certain ICE can be interpreted as a treatment failure, and a plausible relationship to the PRO domain and time-to-worsening of the PRO is considered, the ICE could be incorporated into the endpoint. This can be achieved through a composite strategy, by using a composite outcome of "time-to-worsening of PRO" and "occurrence of ICE". It is important to provide the rationale for combining the PRO outcome with the ICE, as well as information on the relative frequency of the ICE. **Explanation:** if the PRO objective is tolerability, patients may discontinue treatment because it is no longer tolerable. In those cases, non-response can be defined as having a PRO score reflecting harm, or treatment discontinuation. In a confirmatory study, where improvement in PRO is considered, disease progression may be considered as treatment failure. In those cases, using a composite strategy may be an option if PROs after disease progression are not available. Consequently, non-response at a certain time point may be defined as a PRO score that indicates harm or disease progression. However, the occurrence of disease progression and the worsening of PRO measure can only be linked if there is a plausible association between time-to-worsening of the PRO and the patient-reported symptoms. This approach should, therefore, be used with caution because, in real-life scenarios, interpreting results may be complex when it involves combining different outcomes. For example, a patient may experience growth in lung metastases whilst still reporting a good HRQoL. It is therefore, in general, recommended to measure PROs after the ICE instead of using a composite outcome. When many patients experience the ICE during follow-up, this will dominate the composite outcome. Therefore, information on the relative frequency of ICE should be provided.		identical
ICE2_GEN **Statement:** when a certain ICE can be interpreted as a treatment failure with a plausible relationship to the PRO domain and no further PRO data are available afterwards, the ICE could be considered as part of the responder definition (composite outcome). It is important to provide the rationale for combining the PRO outcome with the ICE, as well as information on the relative frequency of the ICE. **Explanation:** if the PRO objective is tolerability, patients may discontinue treatment because it is no longer tolerable. In those cases, non-response can be defined as having a PRO score reflecting harm, or treatment discontinuation. In a confirmatory study, where improvement in PRO is considered, disease progression may be considered as treatment failure. In those cases, using a composite strategy may be an option if PROs after disease progression are not available. Consequently, non-response at a certain time point may be defined as a PRO score that indicates harm or disease progression. However, the occurrence of disease progression and the worsening of PRO measures can only be linked if there is a plausible association between the symptoms related to disease progression and the patient-reported symptoms. This approach should, therefore, be used with caution because, in real-life scenarios, interpreting results may be complex when it involves combining different outcomes. For example, a patient may experience growth in lung metastases while reporting a good HRQoL. It is therefore, in general, recommended to measure PROs after the ICE instead of using a composite outcome. When many patients experience the ICE during follow-up, this will dominate the combined outcome. Therefore, information on the relative frequency of ICE should be provided.		identical
PROvar1_GEN **Statement:** continuous or ordinal PROs should be analysed as continuous or ordinal outcomes. A motivation should be given if a different approach is used, such as dichotomising PRO values in a responder analysis or a time-until-deterioration/improvement analysis.		identical

WP2/RCT	WP3/SAT	Note
RCT PROvar1_GEN **Explanation:** for clinicians and patients, interpreting results from a responder or time-to-deterioration/improvement analysis is often easier. However, from a methodological viewpoint, these analyses have limitations. Categorizing PRO scores may lead to misclassification and reduction in statistical power. Additionally, the choice of threshold for defining responders or events may be hard to justify. Sensitivity analyses are necessary to evaluate the effect of threshold determination and misclassification rates. Additionally, analysing time to deterioration/improvement is complicated by missing data and the fact that the exact moment of "deterioration / improvement" is unknown, since PROs are only measured at specific time points. To address this issue, interval censoring methods can be used. Furthermore, the time-to-deterioration/improvement analysis does not consider the reversibility of the deterioration/improvement status, such as in situations where low PRO values may later improve. Rescaling of PROs, such as transforming an original 4-point response scale into a continuous 0–100 scale, is permissible because a rescaled PRO retains the same statistical properties. However, ordinal scores that have been rescaled should still be analysed as ordinal outcomes.	**SAT PROvar1_GEN** **Explanation:** for clinicians and patients, interpreting results from a responder or time-to-deterioration/improvement analysis is often easier. However, from a methodological viewpoint, these analyses have limitations. Categorizing PRO scores may lead to misclassification and a reduction in statistical power. Additionally, the choice of threshold for defining responders or events may be hard to justify. Sensitivity analyses are necessary to evaluate the effects of threshold determination and misclassification rates. Additionally, analysing time-to-deterioration/improvement is complicated by missing data and the fact that the exact moment of "deterioration / improvement" is unknown, since PROs are only measured at specific time points. To address this issue, interval censoring methods can be used. Furthermore, the time-to-deterioration/improvement analysis does not consider the reversibility of the deterioration/improvement status, such as in situations where low PRO values may later improve. This makes time-to-event endpoints difficult to interpret, especially in single-arm trials without a reference arm. Patient summary measures such as maximum PRO value and area under the curve (AUC) are dependent on the chosen time frame. These measures lack a causal interpretation, and interpretation is especially difficult without a direct comparison group. Different patterns of assessment, each reflecting different clinical scenarios, may result in comparable AUC or maximum PRO value. The maximum PRO value depends on the number of measurements. Furthermore, handling ICEs and missing data is challenging. Rescaling of PROs, such as transforming an original 4-point response scale into a continuous 0–100 scale is permissible because a rescaled PRO retains the same statistical properties. However, ordinal scores that have been rescaled should still be analysed as ordinal outcomes	Additional text in WP3/SAT specific to SATs.

WP3/SAT adapted 19 statements from WP2/RCT

<table>
<tr><th>WP3/SAT</th><th>WP2/RCT</th><th>Note</th></tr>
<tr><td colspan="2">EstFrame1_GEN

Statement: the choice of each estimand should depend on the PRO objective of the clinical trial and on the research question.</td><td>identical</td></tr>
<tr><td>SAT EstFrame1_GEN

Explanation: the PRO objective will determine how to address ICEs, such as treatment discontinuation. If the PRO objective is related to patients' experience while-on-treatment, then PRO measurements while-on-treatment will be more relevant, whereas efficacy objectives measuring long-term benefits on patients would take into account PRO observations even after the treatment is discontinued. Other defining elements may include:

❍ Nature of the disease (disease of interest, stage of disease)
❍ Characteristics of the target population
❍ Whether PROs are measured for benefits and/or tolerability
❍ Treatment regime (e.g., intervention, duration and frequency)
❍ Treatment intent (e.g., palliative or curative)
❍ Treatment blinding
❍ What type of PRO measurements can be used and the availability
❍ PRO time points of interest
❍ The PRO based population summaries of interest (means or medians, changes from baseline, responder, time until deterioration/improvement, etc.)
❍ Relation with other outcomes of interest in the study.

Different stakeholders may prefer different estimands.

Prior to the analysis, an adequate objective should be developed covering all necessary estimand attributes. Clear trial objectives should be translated into key clinical questions of interest by defining suitable estimands.</td><td>RCT EstFrame1_GEN

Explanation: in RCTs, research objectives are often unclear and reported sporadically. The best approach to design, conduct and analyse the PRO part of an RCT depends on the trial context. Several factors may influence the approach chosen. These include the type of treatment and the intent of the study (curative or not) and the type of PRO (symptoms, functional impacts, general HRQoL).

The PRO objective will determine how to address ICEs such as treatment discontinuation. If the PRO objective is related to patients' experience while on treatment, then PRO measurements while on treatment will be more relevant, whereas efficacy objectives measuring long-term benefits on patients would take into account PRO observations even after the treatment is discontinued. Other defining elements may include:

❍ Nature of the disease (disease of interest, stage of disease)
❍ Characteristics of the target population
❍ Whether PROs are measured for benefits and/or tolerability
❍ Treatment regime (e.g., intervention, duration and frequency)
❍ Treatment intent (e.g., palliative or curative)
❍ Treatment blinding
❍ What type of PRO measurements can be used and the availability
❍ PRO time points of interest
❍ The PRO-based population summaries of interest (means or medians, changes from baseline, responder, time until deterioration/improvement, etc.)
❍ Relation with other outcomes of interest in the study.</td><td>Additional text in WP2(RCT specific to RCTs.</td></tr>
</table>

WP3/SAT	WP2/RCT	Note
For example, if the objective is to estimate what proportion of patients on treatment improved their physical functioning at month six, the analysis is limited to patients still on-treatment at month six. However, if the objective is to compare the proportion of patients who improved their physical functioning at month six, the analysis set is not limited by treatment discontinuation.	Different stakeholders may prefer different estimands. Prior to the analysis, an adequate objective should be developed covering all necessary estimand attributes. Clear trial objectives should be translated into key clinical questions of interest by defining suitable estimands. For example, if the objective is to estimate what proportion of patients on treatment improved their physical functioning at month six, the analysis is limited to patients still on treatment at month six. However, if the objective is to compare the proportion of patients who improved their physical functioning at month six, the analysis set is not limited by treatment discontinuation.	
AnalSens1_GEN **Statement:** the overall PRO analysis strategy should include a main PRO analysis supported by sensitivity and/or supplementary analyses (if necessary).		identical
SAT AnalSens1_GEN **Explanation:** an overall PRO analysis strategy should be designed to answer the main research questions and consist of a pre-specified d main analysis/es to reach conclusions. It is important to check the sensitivity of the main conclusions against limitations of the data assumptions and different data analysis approaches. Sensitivity analyses are conducted to ensure that the main findings from PROs are not substantially affected (or remain consistent) when different assumptions or methods are used in the analysis. Supplementary analyses are conducted in addition to the main analysis and sensitivity analysis to better understand the effects of the treatment. **Example:** suppose the PRO objective is to consider levels of physical functioning at month six. However, a number of patients did not have PRO data at month six because they were too ill to fill out the questionnaire. A primary analysis was planned with specific assumptions on how to handle these missing data. As a sensitivity analysis, another analysis method with different missing data assumptions was used to understand the robustness of the findings from the primary analysis.	**RCT AnalSens1_GEN** **Explanation:** an overall PRO analysis strategy should be designed to answer the main research questions and consist of a pre-specified main analysis/es to reach conclusions. It is important to check the sensitivity of the main conclusions against limitations of the data assumptions and different data analysis approaches. Sensitivity analyses are conducted to ensure that the main findings from PROs are not substantially affected (or remain consistent) when different assumptions or methods are used in the analysis. Supplementary analyses are conducted in addition to the main analysis and sensitivity analysis to better understand the effects of the treatment. **Example:** suppose the PRO objective is to consider levels of physical functioning at month six between two treatment arms. However, a number of patients did not have PRO data at month six because they were too ill to fill out the questionnaire. A primary analysis was planned with specific assumptions on how to handle these missing data. As a sensitivity analysis, another analysis method with different missing data assumptions was used to understand the robustness of the findings from the primary analysis.	Additional text in WP2/RCT specific to RCTs.

WP3/SAT	WP2/RCT	Note
AnalSens2_GEN **Statement:** when performing a confirmatory analysis, supplementary analyses could be conducted to provide additional insights into the understanding of the treatment effect. **Explanation:** supplementary analyses are conducted in addition to the main analysis with the intent to provide additional insights into the understanding of the treatment effect. A distinction should be made between sensitivity analyses and supplementary analyses. Sensitivity analyses are used to assess the robustness of the estimator by varying the assumptions about, for instance, the missing data mechanism. On the other hand, supplementary analyses provide additional results by considering different estimands through alternative ICE strategies, changes in population-level summaries definition or subgroup analyses. Relevant supplementary analyses should be pre-specified in the statistical analysis plan as this is important for strengthening the analysis. Post-hoc analyses may be considered to augment (not replace) the pre-specified analyses.		identical
ICEdisc1_GEN **Statement:** when the goal of the PRO objective is to draw conclusions about clinical benefit (confirmatory objective), the treatment policy strategy (i.e. collect and use PRO data after treatment discontinuation) is the preferred strategy for incorporating treatment discontinuation (for reasons other than treatment completion) or the start of subsequent therapy as ICEs in the analysis. If justified, a hypothetical strategy, a composite strategy, or a while-on-treatment strategy can be considered.		identical
SAT ICEdisc1_GEN **Explanation:** in cancer trials, some patients may discontinue treatment during the trial. There are different strategies on how to incorporate study treatment discontinuation (for reasons other than treatment completion) or the start of subsequent therapy in the analysis, and each strategy corresponds to a different interpretation of the treatment effect. Different stakeholders may be interested in different estimands and this should be discussed with the relevant stakeholder. In the treatment policy strategy, the ICE is considered to be part of the compared treatment effects under the intent-to-treat (ITT) principle. It foresees the collection of PRO data after discontinuation. The schedule of data collection in all treatment groups should be consistent prior to and after discontinuation. In cases where the impact of study treatment discontinuation and/or subsequent anti-cancer therapy is not of clinical interest (e.g. cross-over to the experimental therapy) or gathering PRO data after discontinuation is deemed unfeasible (e.g., in the case of advanced disease), the hypothetical strategy, composite strategy, or while-on-treatment strategy can be considered.	**RCT ICEdisc1_GEN** **Explanation:** in cancer trials, some patients may discontinue treatment during the trial. There are different strategies on how to incorporate study treatment discontinuation (for reasons other than treatment completion) or the start of subsequent therapy in the analysis, and each strategy corresponds to a different interpretation of the treatment effect. Different stakeholders may be interested in different estimands and this should be discussed with the relevant stakeholder. Note that in a randomised cancer clinical trial, the study treatment refers to treatment that defines both the experimental and control arm(s). In the treatment policy strategy, the ICE is considered to be part of the compared treatment effects under the intent-to-treat (ITT) principle. It foresees the collection of PRO data after discontinuation. The schedule of data collection in all treatment groups should be consistent prior to and after discontinuation. In cases where the impact of study treatment discontinuation and/or subsequent anti-cancer therapy is not of clinical interest (e.g., cross-over to the experimental therapy) or gathering PRO data after discontinuation is deemed unfeasible (e.g., in the case of advanced disease), the hypothetical strategy, composite strategy, or while-on-treatment strategy can be considered.	Additional text in WP2/RCT specific to RCTs.

WP3/SAT	WP2/RCT	Note
The strategy selection should be based on the trial's objectives, with consideration given to possible scenarios after treatment discontinuation or start of subsequent anti-cancer therapy. If understanding the treatment effect is important, irrespective of whether the treatment is discontinued or followed by subsequent anti-cancer therapy, a treatment policy strategy is advised. When a hypothetical strategy is applied in the context of study treatment discontinuation, the treatment effect is estimated in the hypothetical scenario where no treatment discontinuation and/or starting of a subsequent anti-cancer therapy occurs. This strategy could be applied if the impact of the treatment discontinuation/and or subsequent anti-cancer therapy is not of clinical interest. Such a scenario may be of interest to stakeholders, for instance to patients who wish to understand what would have happened had they kept adhering to treatment. According to this strategy, it is assumed that there is no significant difference between patients who discontinued treatment and patients who remained on treatment. By using a composite strategy, study treatment discontinuation and/or start of a subsequent anti-cancer therapy is included as a component of the PRO variable of interest. This approach should only be used when combining the PRO values with study treatment discontinuation is deemed feasible. Furthermore, an endpoint that is driven by study treatment discontinuation is not recommended. This is because a large proportion of patients who discontinued study treatment will severely impact the accuracy of the estimated treatment effect and make the results harder to interpret. With a while-on-treatment strategy, the approach involves descriptively summarising the data by reporting the magnitude of PRO (change) score or proportion of responders in the group of patients who are receiving the study treatment at each assessment time.	The strategy selection should be based on the trial's objectives, with consideration given to possible scenarios after treatment discontinuation or start of subsequent anti-cancer therapy. If understanding the treatment effect is important, irrespective of whether the treatment is discontinued or followed by subsequent anti-cancer therapy, a treatment policy strategy is advised. When a hypothetical strategy is applied in the context of study treatment discontinuation, the treatment effect is estimated in the hypothetical scenario where no treatment discontinuation and/or starting of a subsequent anti-cancer therapy occurs. This strategy could be applied if the impact of the treatment discontinuation and/or subsequent anti-cancer therapy is not of clinical interest. Such a scenario may be of interest to stakeholders, for instance, to patients who wish to understand what would have happened had they kept adhering to treatment. According to this strategy, it is assumed that there is no significant difference between patients who discontinued treatment and patients who remained on treatment. By using a composite strategy, study treatment discontinuation and/or start of a subsequent anti-cancer therapy is included as a component of the PRO variable of interest. This approach should only be used when combining PRO values with study treatment discontinuation is deemed feasible. Furthermore, an endpoint that is driven by study treatment discontinuation is not recommended. This is because a large proportion of patients who discontinued study treatment will severely impact the accuracy of the estimated treatment effect and make the results harder to interpret. With a while-on-treatment strategy, the approach involves descriptively summarising the data by reporting the magnitude of PRO (change) score or proportion of responders in both arms in the group of patients who are receiving the study treatment at each assessment time.	

WP3/SAT	WP2/RCT	Note
When assessing trends over time, consideration should be given to missing data and ICEs. The group of patients included in the summary measure may vary at each assessment time. These summary measures should be supplemented with numbers on treatment discontinuation. Therefore, consideration should be given to missing data and ICEs to assess the validity of the analysis. In principle, the study protocol and analysis plan should outline in advance the strategy to handle ICEs. One primary strategy should be determined at least if the PRO objective is confirmatory. Supplementary analyses can consider alternative strategies based on the objectives of the study, for instance, adopting a treatment policy strategy as the primary approach and using a hypothetical strategy for supplementary analysis. An overview of the relevant reasons for treatment discontinuation should be reported and discussed to assess the extent to which the ICE could have impacted the PRO results. For descriptive purposes, for instance, to explore patient experience on tolerability other strategies than treatment policy could be relevant, such as while-on-treatment strategy.	When assessing trends over time, consideration should be given to missing data and ICEs. The group of patients included in the summary measure may vary at each assessment time. These summary measures should be supplemented with numbers on treatment discontinuation. Therefore, consideration should be given to missing data and ICEs to assess the validity of the analysis. Treatment groups might no longer be comparable if only reporting data on patients who are receiving treatment. Therefore, consideration should be given to missing data and ICE to assess the validity of the comparative analysis. In principle, the study protocol and analysis plan should outline in advance the strategy to handle ICEs. One primary strategy should be determined at least if the PRO objective is confirmatory. Supplementary analyses can consider alternative strategies based on the objectives of the study, for instance, adopting a treatment policy strategy as the primary approach and using a hypothetical strategy for supplementary analysis. An overview of the relevant reasons for treatment discontinuation should be reported and discussed to assess the extent to which the ICE could have impacted the PRO results. For descriptive purposes, for instance, to explore patient experience on tolerability other strategies than treatment policy could be relevant, such as while-on-treatment strategy.	
SAT Design3_GEN **Statement:** the timing of the PRO assessments and the duration of their associated time windows should be pre-specified and not depend on post-baseline events related to the disease. Furthermore, when there is an external control group, the timing of the PRO assessments should be comparable or accounted for between groups. During the trial, it is advised to adhere as much as possible to these assessment time points.	**RCT Design3_GEN** **Statement:** the timing of the PRO assessments and the duration of their associated time windows should be pre-specified and not depend on post-baseline events related to the disease. Furthermore, the timing of the PRO assessments should be comparable or accounted for between treatment arms. During the trial, it is advised to adhere as much as possible to these assessment time points.	Different due to setting (RCT vs SAT)

WP3/SAT	WP2/RCT	Note
SAT Design3_GEN **Explanation:** the schedule for PRO assessments should be established in advance. When selecting appropriate time points for assessment, it is, however, important to consider the natural history of the disease/progression, the hypothesised impact of therapy over time, the expected treatment side effects, the mode and schedule of treatment administration, and practical considerations such as alignment of assessments with clinic visits and recall period of the PRO measures (so that PROs and clinical data can be collected at the same time). In cancer trials, PRO assessment times should remain as structured as possible and not depend on disease-related events. Asking patients to respond to a questionnaire only when they experience a disease- or treatment-related event, such as diarrhoea, is not recommended, as doing so biases the comparison with external control data. This approach also complicates issues with missing data, as absences may result from either the questionnaire not being completed or the trigger event not occurring. The PRO assessments should be completed according to the protocol at that specific time point. Visits may be delayed or brought forward for practical reasons or not occurred on the planned date. It is therefore recommended to provide time windows for each PRO assessment. These windows are pre-specified time intervals around the planned date where a PRO assessment would still be considered valid and accurately represent the intended assessment date. Additionally, when comparing timing of the PRO measure with external control data, the recall period of the PRO assessment should be taken into account to ensure comparability.	**RCT Design3_GEN** **Explanation:** the schedule for PRO assessments should be established in advance. When selecting appropriate time points for assessment, it is, however, important to consider the natural history of the disease/progression, the hypothesised impact of therapy over time, the expected treatment side effects, the mode and schedule of treatment administration, and practical considerations such as alignment of assessments with clinic visits and recall period of the PRO measures (so that PROs and clinical data can be collected at the same time). In cancer trials, PRO assessment times should remain as structured as possible and not depend on disease-related events. Asking patients to respond to a questionnaire only when they experience a disease- or treatment-related event, such as diarrhoea, is not recommended, as doing so biases the comparison between treatment arms. This approach also complicates issues with missing data, as absences may result from either the questionnaire not being completed or the trigger event not occurring. The PRO assessments should be completed according to the protocol at that specific time point. Visits may be delayed or brought forward for practical reasons or not occurred on the planned date. It is therefore recommended, to provide time windows for each PRO assessment. These windows are pre-specified time intervals around the planned date where a PRO assessment would still be considered valid and accurately represent the intended assessment date. Additionally, the recall period of the PRO assessment should be taken into account to ensure comparability between treatment arms.	Different due to setting (RCT vs SAT)

WP3/SAT	WP2/RCT	Note
SAT Anal1_GEN **Statement:** both completion rates and available data rates should be reported for each assessment time point. The completion rate is calculated by setting the denominator to the expected number of assessments at that time point, defined as the number of patients scheduled for a PRO measurement at that time point according to the protocol. The available data rate is calculated by setting the denominator to the number of patients included in the trial. For both the completion rate and available data rate, the numerator is set to the number of patients who completed the PRO assessment at that specific time point. Any deviations should be justified.	**RCT Anal1_GEN** **Statement:** both completion rates and available data rates should be reported for each assessment time point, in both the confirmatory and descriptive setting. The completion rate is calculated by setting the denominator to the expected number of assessments at that time point, defined as the number of patients scheduled for a PRO measurement at that time point according to the protocol. The available data rate is calculated by setting the denominator to the number of patients randomised in the trial. For both the completion rate and available data rate, the numerator is set to the number of patients who completed the PRO assessment at that specific time point. Any deviations should be justified.	Different due to setting (RCT vs SAT)
SAT Anal1_GEN **Explanation:** in order to clearly show the amount of missing data, both the completion rates and available data rates should be reported for each assessment time point, in both confirmatory and descriptive settings. The completion rate can be interpreted as the proportion of patients with a scheduled PRO assessment at time *t* who complete the PRO assessment. The available data rate can be interpreted as the proportion of all patients included in the trial completing the PRO assessment at time *t*. Any deviations should be justified. To accurately calculate completion rates, a distinction should be made between instances where relevant data could not be collected (resulting in missing data) and cases where data was deliberately not collected or used due to an ICE. When a PRO observation is missing, the corresponding patient is removed from the numerator of the completion rate. When a PRO observation is not collected because of the chosen ICE strategy, the corresponding patient is removed from both the numerator and denominator of the completion rate. The denominator would exclude death since patients who died cannot be expected to provide PRO assessments. Death is included in the denominator when calculating available data rates.	**RCT Anal1_GEN** **Explanation:** in order to clearly show the amount of missing data, both the completion rates and available data rates should be reported for each assessment time point and per treatment arm, in both confirmatory and descriptive settings. The completion rate can be interpreted as the proportion of patients with a scheduled PRO assessment at time *t* who complete the PRO assessment. The available data rate can be interpreted as the proportion of all patients randomised in the trial completing the PRO assessment at time *t*. Any deviations should be justified. To accurately calculate completion rates, a distinction should be made between instances where relevant data could not be collected (resulting in missing data) and cases where data was deliberately not collected or used due to an ICE. When a PRO observation is missing, the corresponding patient is removed from the numerator of the completion rate. When a PRO observation is not collected because of the chosen ICE strategy, the corresponding patient is removed from both the numerator and denominator of the completion rate. The denominator would exclude death since patients who died cannot be expected to provide PRO assessments. Death is included in the denominator when calculating available data rates.	Different due to setting (RCT vs SAT)

<table>
<tr><th>WP3/SAT</th><th>WP2/RCT</th><th>Note</th></tr>
<tr><td>Example: in a cancer single arm trial where PROs are to be collected at the start of the 6th cycle of chemotherapy, the scenario is as follows:
<table><tr><th>Criterion</th><th>Number of patients:</th></tr><tr><td>Included in the trial</td><td>200</td></tr><tr><td>Deceased before 6th cycle</td><td>20</td></tr><tr><td>Lost-to-follow up before 6th cycle</td><td>5</td></tr><tr><td>Discontinued treatment before 6th cycle</td><td>75</td></tr><tr><td>Started the 6th cycle</td><td>100</td></tr><tr><td>Provided PRO data</td><td>80</td></tr></table>Completion rate = number of patients who provided PRO data/expected number of assessments at that time point = 80/100 = 80%

Available data rate = number of patients who provided PRO data/patients included in the trial = 80/200 = 40%.</td><td>Example: in a cancer RCT where PROs are to be collected at the start of the 6th cycle of chemotherapy, the scenario for the control arm is as follows:
<table><tr><th>Criterion (control arm):</th><th>Number of patients:</th></tr><tr><td>Randomised</td><td>200</td></tr><tr><td>Deceased before 6th cycle</td><td>20</td></tr><tr><td>Lost-to-follow up before 6th cycle</td><td>5</td></tr><tr><td>Discontinued treatment before 6th cycle</td><td>75</td></tr><tr><td>Started the 6th cycle</td><td>100</td></tr><tr><td>Provided PRO data</td><td>80</td></tr></table>Completion rate = number of patients who provided PRO data/expected number of assessments at that time point = 80/100 = 80%

Available data rate = number of patients who provided PRO data/patients randomised in the trial = 80/200 = 40%.</td><td></td></tr>
<tr><td colspan="2">AnalMain7_GEN

Statement: using single imputation techniques, complete-case analysis or available case analysis to handle missing data is generally not recommended. A justification should be given if these approaches are used.</td><td>Identical</td></tr>
<tr><td>SAT AnalMain7_GEN

Explanation: analyses could be heavily impacted by the assumptions regarding the type of missing data, i.e., whether or not missingness can be considered to occur randomly. In longitudinal studies attrition bias could occur i.e., some participants are more likely to drop out than others. Reasons for missingness should be explored. Several statistical techniques exist to deal with missing data.

When missing data are non-negligible, simple techniques such as single imputation, complete case analysis (only including patients with no missing data), or available case analysis (only including patients with no missing data at the time point of interest) are generally not recommended due to potential bias and loss of information. In the case of a binary endpoint (success/failure), single imputation is often implemented by considering patients with missing data as non-responders. For a continuous endpoint, single imputation is often implemented by imputing the last observed PRO score (Last Observation Carried Forward, [LOCF]) or the mean of the observed PRO scores of the patient (unconditional mean imputation).</td><td>RCT AnalMain7_GEN

Explanation: analyses could be heavily impacted by the assumptions regarding the type of missing data, i.e., whether or not missingness can be considered to occur randomly. In longitudinal studies, attrition bias could occur, i.e., some participants are more likely to drop out than others. Reasons for missingness should be explored. Several statistical techniques exist to deal with missing data.

When missing data are non-negligible, simple techniques such as single imputation, complete-case analysis (only including patients with no missing data), or available case analysis (only including patients with no missing data at the time point of interest) are generally not recommended due to potential bias and loss of information. In the case of a binary endpoint (success/failure), single imputation is often implemented by considering patients with missing data as non-responders. For a continuous endpoint, single imputation is often implemented by imputing the last observed PRO score (Last Observation Carried Forward, [LOCF]) or the mean of the observed PRO scores of the patient (unconditional mean imputation).</td><td>Additional text in WP2/RCT specific to RCTs.</td></tr>
</table>

WP3/SAT	WP2/RCT	Note
Although these approaches are straightforward, the results from most simple imputation techniques are only valid under assumptions which are often untenable and too restrictive. For example, LOCF assumes patients' PRO score will no longer change from last observed score; considering patients with missing data as non-responders assumes the missingness is due to adverse PRO scores. These methods may lead to biased estimates. Secondly, the use of single imputation values ignores the uncertainty associated with the missingness, resulting in an overestimation of the precision of the treatment effect. However, it is important to distinguish this recommendation from the composite strategy used to address ICEs when defining the PRO variable of interest. Within this estimand framework, the imputation of a pre-specified value is based on the relevant PRO objective. Single imputation of missing data could be reasonable in specific cases where there is a clear justification. For example, imputation of extreme PRO scores or (non-) responses after a certain event, when there is a clear justification, could be one component of the sensitivity analysis. This may be relevant in the context of tolerability analyses if the missingness is assumed to be related to a patient's health status. This recommendation does not concern imputation as part of the instrument scoring instructions. As an alternative, multiple imputation techniques can be considered to address missing data. These techniques generate several completed datasets and combine the results from these datasets, thereby better accounting for the uncertainty associated with missing data. When the PRO objective is to inform safety and tolerability, an available case analysis may be considered when summarising PRO scores descriptively at each assessment time point. This approach is only justifiable with high PRO completion rates and no evidence of selective missingness to avoid misinterpretation of PRO score results. The number of patients expected to complete the PRO measure and the number of patients who did not complete the PRO measure at each assessment time point should be reported along with reasons for non-completion.	Although these approaches are straightforward, the results from most simple imputation techniques are only valid under assumptions which are often untenable and too restrictive. For example, LOCF assumes patients' PRO score will no longer change from last observed score; considering patients with missing data as non-responders assumes the missingness is due to adverse PRO scores. These methods may lead to biased estimates. Secondly, the use of single imputation values ignores the uncertainty associated with the missingness, resulting in an overestimation of the precision of the treatment effect. However, it is important to distinguish this recommendation from the composite strategy used to address ICEs when defining the PRO variable of interest. Within this estimand framework, the imputation of a pre-specified value is based on the relevant PRO objective. Single imputation of missing data could be reasonable in specific cases where there is a clear justification. For example, imputation of extreme PRO scores or (non-) responses after a certain event, when there is a clear justification, could be one component of the sensitivity analysis. This may be relevant in the context of tolerability analyses if the missingness is assumed to be related to a patient's health status. This recommendation does not concern imputation as part of the instrument scoring instructions. As an alternative, multiple imputation techniques can be considered to address missing data. These techniques generate several completed datasets and combine the results from these datasets, thereby better accounting for the uncertainty associated with missing data. When the PRO objective is to inform safety and tolerability, an available case analysis may be considered when summarising PRO scores descriptively at each assessment time point by treatment arm. This approach is only justifiable with high PRO completion rates and no evidence of selective missingness to avoid misinterpretation of PRO score results. The number of patients expected to complete the PRO measure and the number of patients who did not complete the PRO measure at each assessment time point and by treatment arm should be reported along with reasons for non-completion.	

WP3/SAT	WP2/RCT	Note
AnalSens3_GEN **Statement:** it is recommended to perform sensitivity analyses to assess the robustness of results with varying plausible assumptions and methods for handling missing data and its impact on the conclusions for the primary and key secondary PRO study objectives. Each sensitivity analysis should be designed to assess the effect on the results of the particular assumptions and methods used to account for the missing data.		Identical
SAT AnalSens3_GEN **Explanation:** sensitivity analysis should be used to assess how varying the assumptions about the missing data mechanism impact conclusions of interest. When the missingness mechanism is assumed to be MAR (missing at random), standard estimation methods can be used, but the assumption of MAR against MNAR (missing not at random) itself cannot be tested on the available data. As PRO data collection depends on the voluntary participation of the patient, MNAR is more plausible as patients' health status may directly or indirectly influence the completion of PRO assessments. Therefore, different scenarios in the direction of MNAR should be considered. All MNAR models are based on assumptions that cannot be verified based on the observed data, yet the plausibility of the models can be formulated in terms of clinical arguments. A large number of different methods exist to perform a sensitivity analysis, such as multiple imputation in the pattern-mixture framework, local influence approaches, shared-parameter models, etc. Each sensitivity analysis should be designed to assess the effect on the results of the particular assumptions made to account for the missing data. Key sensitivity analyses should be pre-specified in the statistical analysis plan. In sensitivity analyses, the information on deviation from the assumptions should be clearly described. The deviations should be plausible and support interpretability.	**RCT AnalSens3_GEN** **Explanation:** sensitivity analysis should be used to assess how varying the assumptions about the missing data mechanism impacts conclusions of interest. When the missingness mechanism is assumed to be MAR (missing at random), standard estimation methods can be used, but the assumption of MAR against MNAR (missing not at random) itself cannot be tested on the available data. As PRO data collection depends on the voluntary participation of the patient, MNAR is more plausible as patients' health status may directly or indirectly influence the completion of PRO assessments. Therefore, different scenarios in the direction of MNAR should be considered. All MNAR models are based on assumptions that cannot be verified based on the observed data, yet the plausibility of the models can be formulated in terms of clinical arguments. A large number of different methods exist to perform a sensitivity analysis, such as multiple imputation in the pattern-mixture framework, local influence approaches, shared-parameter models, etc. Each sensitivity analysis should be designed to assess the effect on the results of the particular assumptions made to account for the missing data. If the primary analysis, for example, assumes the same missingness mechanism to hold across all randomised treatment arms, then multiple imputation using an imputation model that differs across treatment arms can be used as a sensitivity analysis. Key sensitivity analyses should be pre-specified in the statistical analysis plan. In sensitivity analyses, the information on deviation from the assumptions should be clearly described. The deviations should be plausible and support interpretability.	Additional text in WP2/RCT specific to RCTs.

<table>
<tr><th>WP3/SAT</th><th>WP2/RCT</th><th>Note</th></tr>
<tr><td colspan="2">Design4_GEN

Statement: it is recommended to collect and report the reasons and frequencies for missingness and incorporate these in the sensitivity analyses.

Explanation: it is rarely appropriate to assume that the association between missingness and the PRO score is the same, whatever the cause for the missingness. This is especially relevant in cancer trials, where the disease's aggressive nature and the toxicity of the treatment can lead to missing PRO data for different reasons. Each reason for dropout may be related to the PRO score in a different way.

For example, if PRO assessments are missing due to a lack of a translated questionnaire, this can reasonably be assumed to be non-informative of the PRO endpoint, whilst missing PRO assessments due to the patient being too ill to fill in the questionnaire could reasonably be assumed to be informative of the PRO endpoint. These two different reasons for missing data could, therefore, be accounted for in the sensitivity analysis in a different way.

Collecting information about the reason for missing data may, however, not be feasible in all cases, for instance when using remote data collection of at-home diaries. During the design phase, it is important to implement strategies that reduce the occurrence of missing data and to evaluate how missing data might correlate with PRO scores.

Methods that could be used are, for example, multiple imputation and tipping point analysis. Missingness at a given time point (yes, no) can be a function of key baseline demographic and clinical characteristics, which can be analysed with a multiple logistic regression model.</td><td>Identical</td></tr>
<tr><td colspan="2">AnalMain5_GEN

Statement: when performing a time-to-event analysis on PRO data, it is recommended to apply statistical methods that incorporate the interval-censored nature of the data.</td><td>Identical</td></tr>
<tr><td>SAT AnalMain5_GEN

Explanation: in practice, when PROs are used to define an event, the exact time of the event is unknown, as the PRO assessments are performed at pre-specified periodic assessment time points. PRO data, therefore, are interval-censored, as the event time is only known to fall within a particular interval (i.e., between two pre-specified assessment time points). Right-censoring occurs when a participant's survival time is known to exceed a certain value (equivalent to single right-point imputation). This could lead to bias in estimating the effect of the treatment and may result in reduced statistical power. The extent of this bias depends on the frequency of measurements, as it relates to the ratio of the length of time between assessments and the anticipated duration until an event occurs.

Secondly, different assessment schemes in the single arm study and external control data can create substantial bias.</td><td>RCT AnalMain5_GEN

Explanation: in practice, when PROs are used to define an event, the exact time of the event is unknown, as the PRO assessments are performed at pre-specified periodic assessment time points. PRO data, therefore, are interval-censored, as the event time is only known to fall within a particular interval (i.e., between two pre-specified assessment time points). Right-censoring occurs when a participant's survival time is known to exceed a certain value (equivalent to single right-point imputation). This could lead to bias in estimating the effect of the treatment and may result in reduced statistical power. The extent of this bias depends on the frequency of measurements, as it relates to the ratio of the length of time between assessments and the anticipated duration until an event occurs.

Secondly, different assessment schemes between treatment arms can create substantial bias.</td><td></td></tr>
</table>

WP3/SAT	WP2/RCT	Note
The standard time-to-event analysis assumes that censoring is independent of the event. With this in mind, reasons for censoring should be explored and reported. Overall, it is therefore advised to use methods for interval-censored data when PRO data are obtained at pre-specified assessment time points.	The standard time-to-event analysis assumes that censoring is independent of the event. With this in mind, reasons for censoring should be explored and reported. Whilst most non-PRO time-to-event endpoints in RCTs (e.g. overall survival, progression-free survival) have been analysed using right-censoring techniques, the underlying assumptions are rarely verified. PRO data collection, which is dependent on patient participation, is even more susceptible to deviations from right-censoring assumptions due to higher rates of missing data and deviations from a fixed assessment schedule. Moreover, this is an issue that is not corrected by the randomisation process as mechanisms causing deviations to the assessment schedule may differ across the different arms. Overall, it is therefore advised to use methods for interval-censored data when PRO data are obtained at pre-specified assessment time points.	Different due to setting (RCT vs SAT)
AnalMain2_GEN **Statement:** when the goal of the PRO objective is to draw conclusions about clinical benefit (confirmatory objective), it is not recommended to analyse data at each time point separately using multiple cross-sectional analyses of the magnitude of PRO (change) score or proportion of responders.		Identical
SAT AnalMain2_GEN **Explanation:** when the PRO assessments take place at pre-specified time points with repeated assessments expected per patient over the course of the study, the data can be analysed using a multiple cross-sectional univariate analysis performed at each time point separately. This is a valid approach if the focus is on assessing the treatment effect at specific time points. However, several caveats apply. First, there is a considerable loss of information by considering cross-sectional analyses instead of modelling the full longitudinal profiles, which results in reduced statistical power at each time point. Second, when analysing time points one at a time, it becomes challenging to fully take into account missing data.	**RCT AnalMain2_GEN** **Explanation:** when the PRO assessments take place at pre-specified time points with repeated assessments expected per patient over the course of the study, the data can be analysed using a multiple cross-sectional univariate analysis performed at each time point separately. This is a valid approach if the focus is on assessing the treatment effect at specific time points. However, several caveats apply. First, there is a considerable loss of information by considering cross-sectional analyses instead of modelling the full longitudinal profiles, which results in reduced statistical power at each time point. Second, when analysing time points one at a time, it becomes challenging to fully take into account missing data. Third, simple univariate tests like the two-sample t-test are based on the assumption that the scores are identically and independently distributed. This assumption is often untenable and too restrictive.	The additional text in WP2/ RCT is not needed for WP3/SAT

WP3/SAT	WP2/RCT	Note
Additionally, multiple statistical testing will inflate the type I error rate. Finally, performing multiple cross-sectional univariate analysis for each time point does not enable researchers to study evolutions in PRO values over time, drawing longitudinal interpretation, nor does it allow them to consider the correlation between different observations of the same patient.	Additionally, multiple statistical testing will inflate the type I error rate. Finally, performing multiple cross-sectional univariate analysis for each time point does not enable researchers to study evolutions in PRO values over time, drawing longitudinal interpretation, nor does it allow them to consider the correlation between different observations of the same patient.	
AnalMain3_GEN **Statement:** when performing a responder analysis or an analysis of continuous PROs at a pre-specified time point of interest, a longitudinal analysis that considers the data's repeated measurement structure should be conducted, in line with the defined estimand of interest. **Explanation:** when the endpoint is magnitude of PRO (change) score, with time included as a discrete variable, and repeated assessments have been made per patient over the course of the study, a linear regression model can be applied. This model accounts for the correlation between a patient's measurements through a residual covariance matrix. When the endpoint is binary (success/failure), the probability of being a responder over time can be modelled using a longitudinal model with a logit link-function. One can choose, for example, to use a marginal model using generalised estimating equations (GEE), or to use a hierarchical model such as a generalised linear mixed model (GLMM) using maximum likelihood. However, when using linear mixed models (LMMs) or GLMMs, one should be aware that these methods implicitly impute values after death, while in reality PRO values after death do not exist.		Identical
AnalMain4_GEN **Statement:** when performing a responder analysis or an analysis of the magnitude of PRO (change) score, it is recommended to include time as a categorical variable as this requires fewer assumptions compared to a model using time as a continuous variable. If applying models using time as a continuous variable, their underlying assumptions need to be justified.		Identical
SAT AnalMain4_GEN **Explanation:** in analyses focused on the proportion of responders or on the magnitude of PRO (change) score, when patients have been assessed multiple times throughout the study, it is important to consider the timing of these assessments in relation to the baseline anchor time, such as time of the start of treatment. Assessment time should be included as a discrete variable. This approach ensures that no assumptions are made on the relationship between time and the outcome variable, even though it requires the use of more parameters. The requirements are that the assessment schedule is pre-specified and similar between treatment groups (in frequency and timing) for all included patients and that the sample size is sufficiently large compared to the number of assessments.	**RCT AnalMain4_GEN** **Explanation:** in analyses focused on the proportion of responders or on the magnitude of PRO (change) score, when patients have been assessed multiple times throughout the study, it is important to consider the timing of these assessments in relation to the baseline anchor time, such as time of randomisation. Assessment time should be included as a discrete variable. This approach ensures that no assumptions are made on the relationship between time and the outcome variable, even though it requires the use of more parameters. The requirements are that the assessment schedule is pre-specified and similar between treatment arms (in frequency and timing) for all included patients and that the sample size is sufficiently large compared to the number of assessments.	Different due to setting (RCT vs SAT)

WP3/SAT	WP2/RCT	Note
When the requirements are not met (e.g., in a rare cancer trial with a small sample size and frequent PRO assessments), time could be included as a continuous variable. In this case, the model should be sufficiently flexible in order to accurately represent the relationship between the outcome and time. Any assumption made on the functional relationship between time and the PRO variable of interest should be justified. If the data shows a linear relationship between the PRO variable of interest and time, time could be included as a continuous variable.	When the requirements are not met (e.g., in a rare cancer trial with a small sample size and frequent PRO assessments), time could be included as a continuous variable. In this case, the model should be sufficiently flexible in order to accurately represent the relationship between the outcome and time. Any assumption made on the functional relationship between time and the PRO variable of interest should be justified. If the data shows a linear relationship between the PRO variable of interest and time, time could be included as a continuous variable.	
Assump1_GEN **Statement:** model diagnostics should be used to check the model assumptions, where feasible, and investigate the presence of outlying and influential observations. **Explanation:** when analysing PRO data, the assumptions underlying both the statistical model and, if relevant, the test for the significance of the treatment effect should be clearly stated and verified. Incorrect assumptions can render the results invalid. Additionally, the diagnostics should be used to identify the presence of any outlying observations (i.e., an observation for which the predicted value based on the model is significantly different from the observed value) and influential observations (i.e., data points that greatly affect the model's fit). Assumptions that are often encountered are independence of the observations, normality of the residual errors of a general linear mixed model for a continuous outcome, and proportional hazards assumption for a proportional hazards model, amongst others. Some assumptions, such as the independence of the observations, are related to the design of the study and cannot be checked using model diagnostics. Other assumptions, such as the normality of the residual errors, are related to the statistical model. Whilst not all assumptions can be easily checked, reasonable efforts should be made to assess their plausibility, where feasible. In case no formal test exists for a particular assumption, indirect evidence (such as graphical methods) can be explored.		Identical
AnalMain6_GEN **Statement:** when a longitudinal model is fitted, appropriate correction for the baseline value of the PRO variable should be considered. **Explanation:** in a longitudinal study, there are two different approaches depending on the parameter of interest. Approach A Under the first approach, the interest is to estimate a single average (or contrasts between such averages from different groups). The single average can be expressed as the original outcome or as a change from the baseline. The estimate can be obtained through a cross-sectional model (i.e., ANOVA or ANCOVA) or, more appropriately, a longitudinal model to estimate the average at the time point of interest. If these models are used, correcting for the baseline outcome is recommended, as this may account for unwanted variability between subjects in terms of their change from baseline or at the time point of interest, thereby reducing standard errors.		

<table>
<tr><th>WP3/SAT</th><th>WP2/RCT</th><th>Note</th></tr>
<tr><td colspan="2">Approach B

Under the second approach, the parameter of interest is the longitudinal trend (i.e., slope) or the contrast of trends between randomised groups. When the change from the baseline is modelled, correcting for the baseline will not affect the slope's estimate or standard error. Therefore, correcting for baseline is not required. However, when the original outcome is being modelled, adjusting for baseline would imply that the outcome at time $t = 0$ is not included in the response vector, as it cannot be a response and a covariate simultaneously. Given the interest in estimating slopes, and since correction for the baseline outcome would not affect the results, such a correction for the baseline outcome would imply that baseline is no longer used in the analysis hence resulting in a loss of efficiency.

Example: in a randomised longitudinal study where the main objective is to compare the improvement in a specific domain (e.g., pain) over time between two treatment arms, PROs are collected at baseline (i.e., before treatment administration) and then at months one, two, three and four. The analysis of interest could be to compare the pain scores between the two treatment arms at month four. This can be achieved by fitting a linear mixed model to the study data and calculating the difference in average pain scores at month four between the two treatment arms. In this case, it is recommended to use the baseline pain score as a covariate in the model regardless of whether the endpoint is the pain score itself or the change from the baseline pain score to account for the baseline variability.

If the research interest is to compare the time trends (i.e., slopes) in pain between the two treatment arms, a linear mixed model can be fitted. In this scenario:
❍ If the model uses changes from baseline values, there is no advantage in also correcting for the baseline value.
❍ If the model uses the actual pain scores at each time point, the baseline score should be included as part of the response vector, not as a covariate in the model to make maximum use of the data. In this case, it is not recommended to use the baseline pain score as a covariate in the model.

Therefore, the choice to correct for baseline variables in the statistical model depends on the nature of the research question, as illustrated in this hypothetical clinical trial scenario.</td><td>Identical</td></tr>
<tr><td>SAT AnalMain1_GEN

Statement: when performing a descriptive analysis, summaries of the observed data should be reported by assessment time with, ideally, a variability measure (e.g., standard deviation, variance or interquartile range) and all estimates should have a measure of error (standard error or confidence interval). Inferential statistics are not part of such an analysis.</td><td>RCT AnalMain1_GEN

Statement: when performing a descriptive analysis, summaries of the observed data should be reported for each treatment arm by assessment time, with, ideally, a variability measure (e.g., standard deviation, variance or interquartile range) and all estimates should have a measure of error (standard error or confidence interval). A formal comparison between treatment arms is not recommended. Inferential statistics are not part of such an analysis.</td><td>Additional text in WP2/RCT specific to RCTs.</td></tr>
</table>

WP3/SAT	WP2/RCT	Note
SAT AnalMain1_GEN **Explanation:** the goal of a descriptive analysis is to summarise the observed data. In most trials the tolerability profiles are best addressed by applying descriptive statistical methods to the data. When interpreting summary measures from later time points that are based on data from a smaller number of patients, caution is crucial due to the potential limitations. If the data are sparse and affected by selection bias, it is recommended not to rely on the summary measures. The ICE strategy should be clearly specified, such as strategies where summaries are limited to specific patients (alive, on treatment and/or progression-free) at the different assessment time points, according to the study objective. When appropriate, a distribution of the occurrence of relevant ICEs for each time point could be provided.	**RCT AnalMain1_GEN** **Explanation:** the goal of a descriptive analysis is to summarise the observed data. In most trials, the tolerability profiles are best addressed by applying descriptive statistical methods to the data. When interpreting summary measures from later time points that are based on data from a smaller number of patients, caution is crucial due to the potential limitations. If the data are sparse and affected by selection bias, it is recommended not to rely on the summary measures. The ICE strategy should be clearly specified, such as strategies where summaries are limited to specific patients (alive, on treatment and/or progression-free) at the different assessment time points, according to the study objective. When appropriate, a distribution of the occurrence of relevant ICE for each treatment arm and for each time point could be provided.	Additional text in WP2/RCT specific to RCTs.
Psum3_GEN **Statement:** when performing a descriptive time-to-event analysis, the data can be summarised by the median or by another relevant time-to-event percentile, and by the probability of experiencing an event at a specific time point. Adding a measure of variability is recommended. **Explanation:** the goal of a descriptive analysis is to summarise the observed data. When a time-to-event analysis is performed, the data can be summarised by the median (or other relevant percentile) or the event probability at a specific time point. Measures of variation such as standard deviations, percentiles, interquartile ranges and interval estimates such as confidence intervals should be reported. For descriptive purposes, other population-level summaries might be of interest such as restricted mean survival times. It is advised to calculate the summary measures using a method that takes into account the interval-censored nature of the data. However, the median event time is not always directly available in cases of interval-censored data. To estimate the median event time, one can either take the upper bound of the interval or make a reasonable assumption regarding the distribution of the mass in the region of support (e.g., linear or exponential). If fewer than 50% of the patients had an event during the trial, it is not possible to calculate the median time-to-event based on the data. In this case, another percentile or the event-free rate at a specific time point can be used to summarise the data. Providing the Kaplan-Meier curve and interval estimates can make the results easier to interpret.		Identical

WP3/SAT	WP2/RCT	Note
SAT Psum1_GEN **Statement:** when performing an analysis of the descriptive magnitude of (PRO) change score, the outcomes can be reported as the magnitude of change at a pre-specified assessment point(s) along with a measure of variability. When evaluating trends over time, consideration should be given to underlying assumptions about missing data and ICEs strategies.	**RCT Psum1_GEN** **Statement:** when performing an analysis of the descriptive magnitude of (PRO) change score, the outcomes can be reported as the magnitude of change for each arm at a pre-specified assessment point(s) along with a measure of variability. When evaluating trends over time, consideration should be given to underlying assumptions about missing data and ICEs strategies.	Additional text in WP2/RCT specific to RCTs.
SAT Psum1_GEN **Explanation:** the goal of a descriptive analysis is to summarise the observed data. The data can be summarised by means, medians or the magnitude of change at each assessment time point. Measures of variation such as standard deviations, percentiles, interquartile ranges and interval estimates such as confidence intervals should also be reported. The summary measure reported for each assessment time can cover a different subset of patients; for example, it might only include data from patients who are alive. Therefore, apparent time trends in repeated cross-sectional outcomes must be interpreted with care. The pattern of differences over time, based on these summary statistics, is subject to selection bias due to attrition. This bias increases over time as patients drop out of the analysis due to lost-to-follow-up. Furthermore, ICEs or varying missing data over time (for instance, lower completion rates with increased follow-up) may further complicate the interpretation.	**RCT Psum1_GEN** **Explanation:** the goal of a descriptive analysis is to summarize the observed data. The data can be summarized by means, medians or the magnitude of change in both arms at each assessment time point. Measures of variation such as standard deviations, percentiles, interquartile ranges and interval estimates such as confidence intervals should also be reported. The summary measure reported for each assessment time can cover a different subset of patients; for example, it might only include data from patients who are alive. Therefore apparent time trends in repeated cross-sectional outcomes must be interpreted with care. The pattern of differences over time, based on these summary statistics, is subject to selection bias due to attrition. This bias increases over time as patients drop out of the analysis due to lost-to-follow-up. Furthermore, ICEs or varying missing data over time (for instance, lower completion rates with increased follow-up) may further complicate the interpretation. Furthermore, because treatment arms might no longer be comparable, missing data and ICEs must be taken into account before comparing treatment arms (for example, by reporting the difference in magnitude of change between arms).	Additional text in WP2/RCT specific to RCTs.
SAT Psum2_GEN **Statement:** when performing a descriptive responder analysis, the outcome can be reported as the proportion of responders at predefined assessment point(s), along with a measure of variability. When evaluating trends over time, consideration should be given to underlying assumptions about missing data and ICEs strategies.	**RCT Psum2_GEN** **Statement:** when performing a descriptive responder analysis, the outcome can be reported as the proportion of responders for each arm at predefined assessment point(s), along with a measure of variability. When evaluating trends over time, consideration should be given to underlying assumptions about missing data and ICEs strategies.	Additional text in WP2/RCT specific to RCTs.

WP3/SAT	WP2/RCT	Note
SAT Psum2_GEN	**RCT Psum2_GEN explanation**	
the goal of a descriptive analysis is to summarise the observed data. When performing a responder analysis, the data can be summarised by the proportion of responders. Measures of variation such as standard deviations, percentiles, interquartile ranges and interval estimates such as confidence intervals should be reported.	the goal of a descriptive analysis is to summarise the observed data. When performing a responder analysis, the data can be summarised by the proportion of responders in both arms. Measures of variation such as standard deviations, percentiles, interquartile ranges and interval estimates such as confidence intervals should be reported.	
The proportion of responders reported for each assessment time can cover a different subset of patients– for example, it might only include data from patients who are alive. Therefore, apparent time trends in repeated cross-sectional outcomes must be interpreted with care.	The proportion of responders reported for each assessment time can cover a different subset of patients– for example, it might only include data from patients who are alive. Therefore, apparent time trends in repeated cross-sectional outcomes must be interpreted with care.	Additional text in WP2/RCT specific to RCTs.
The pattern of differences over time, based on these summary statistics, is subject to selection bias due to attrition. This bias increases over time as patients drop out of the analysis due to lost-to-follow-up. Furthermore, ICEs or varying missing data over time (for instance, lower completion rates with increased follow-up) may further complicate the interpretation.	The pattern of differences over time, based on these summary statistics, is subject to selection bias due to attrition. This bias increases over time as patients drop out of the analysis due to lost-to-follow-up. Furthermore, ICEs or varying missing data over time (for instance, lower completion rates with increased follow-up) may further complicate the interpretation.	