

FREQUENT DEFICIENCIES IN BIOEQUIVALENCE STUDY PROTOCOLS

PQT/MED strongly recommends that applicants planning to conduct bioequivalence (BE) studies for submission to PQT/MED submit a final draft of the study protocol to PQT/MED for comment prior to undertaking the study. The purpose of this process is to assist applicants develop studies that will best be able to detect differences in *in vivo* performance between drug products, minimise factors that could introduce variability into the study data, and align with PQT/MED guidelines and requirements.

PQT/MED has noted that there are some protocol issues/deficiencies that are commonly observed during this process. The purpose of this document is to summarize these commonly observed deficiencies such that applicants can address these issues during protocol development, prior to submission of the protocol to PQT/MED for comment.

The following issues should be noted and addressed during protocol development:

COMPARATOR PRODUCT

- The selected comparator product should be identified by the Marketing Authorisation Holder (MAH) and the market of purchase, not the manufacturer and the country of manufacturing.
- The comparator product should be purchased from a well-regulated market with a stringent regulatory authority (SRA), unless otherwise agreed with the PQT/MED (e.g., if not available in SRA countries).
- The comparator product should be selected from those listed on the web page of PQT/MED. The most appropriate strength of the comparator product should be selected to minimise or avoid administration of multiple tablets.
- The comparator product purchase, shipment, and storage should be properly documented and meet the requirements as described in the list of acceptable comparator products found on the PQT/MED website.

TEST PRODUCT

- The proposed product / product strengths should be listed in the Expressions of Interest (EoI).

INCLUSION AND EXCLUSION CRITERIA

- Bioequivalence studies are generally conducted in healthy volunteers. Subjects should be recruited into the study only if their health is fully verified, including verification that serum biochemistry and haematology parameter values are within pre-defined normal ranges. The normal ranges of the study centre should be included as an appendix to the study protocol.
- The enrolment of a subject with measured values for the health verification parameters that fall outside the pre-defined site normal values should not occur, except on a rare, exceptional basis. On the rare occasion when a subject is enrolled in a study despite having a measurement outside the site normal range, the study physician should have a clearly documented and medically rigorous justification for making that exception. Good Clinical Practices (GCP) require that the rights, safety, and well-being of trial subjects be given top priority in every trial conducted. However, the inclusion criteria generally accept that the inclusion of study subjects with values outside of the pre-defined normal range is possible if considered clinically not significant by the principal investigator. This flexibility generally leads to values outside the pre-defined normal ranges in all, or almost all, included subjects. Protocols should be designed to minimise this, particularly with respect to haematological parameters, thus protecting the health and safety of the subjects.
- The exclusion criterion for participation in any clinical study or blood donation should be 90 days.

METHOD OF ADMINISTRATION

- The test and comparator products should be administered under the usual conditions of use. Therefore, it is not acceptable to administer the products under yellow monochromatic light to avoid degradation; normal lighting conditions should be employed.
- The test and comparator products should be administered under the usual conditions of use. Therefore, it is not acceptable to administer a dispersible tablet for paediatric patients with a full glass of water, like a conventional tablet would be administered to adults. The volume used to disperse the tablet and the volume used for rinsing the container (if any) should be those described in the labelling of the product. For example, if the dosing instructions are that the test product should be dispersed in a small amount of water, the test product (e.g., granules) should be dispersed in a small amount of water (e.g., 20 – 40 mL). If the dosing instructions are that the test product will be taken sprinkled on food, the bioequivalence study should follow these dosing instructions. The comparator product should be administered in keeping with its labelling. For example, if the comparator product is an adult strength, conventional-release tablet, it should be taken with a glass of water (e.g., 180 - 240 mL).
- The meal composition employed in fed studies should be consistent with product labelling and/or the PQT/MED specific notes on the design of bioequivalence studies. If the fed study is going to be submitted to regulatory agencies (e.g., US-FDA) where both a fasted and a fed state study are required, a study with the meal composition required by these regulatory authorities may also be acceptable for PQT/MED if both studies are submitted.
- The composition of the meal, including caloric content of carbohydrates, fat and proteins, should be described in the study protocol.

SAMPLING TIMES

- Sampling times should be sufficiently frequent around the expected T_{max} for a proper characterisation of C_{max} .
- Sampling times should cover at least 3 half-lives when calculating AUC_{0-t} .
- Sample collection after 72 hours is not necessary.
- The wash out period should not be excessively large compared with 5 times the largest expected half-life.

PHARMACOKINETIC ANALYSIS

- The protocol should indicate the software to be used for pharmacokinetic calculations as well as the trapezoidal method employed for AUC calculation.
- Any below LLOQ value(s), including those between two valid concentration values, should be reported as zero.
- Plots should be provided in original scale and semi-logarithmic scale for each subject in addition to mean data plots.
- T_{max} is considered a primary PK parameter only in the exceptional cases where onset of action is clinically relevant.
- AUC_{0-inf} is not considered to be a primary PK parameter in single dose studies of immediate release products. AUC_{0-t} is the primary parameter for assessment of extent of exposure.
- For drugs with long half-life, AUC truncated at 72 h could be used. However, in the event of a missing sample at 72 hours, that profile should be excluded. In the case of a 2x2 design, this implies the exclusion of all AUC data for that subject.

SAMPLE SIZE CALCULATION

- The sample size of replicate cross-over designs where C_{max} acceptance range is widened based on intra-subject coefficient of variation (CV) of the comparator product but AUC acceptance range is the classical 80-125%, should be calculated for C_{max} with widening and AUC without widening. Sometimes AUC without widening requires a larger sample size than C_{max} with widening.
- The sample size calculation for a replicate design with widening of the acceptance range should be calculated as described by Tothfalusi and Endrenyi in “Sample Sizes for Designing Bioequivalence Studies for Highly Variable Drugs”. J Pharm Pharmaceut Sci 15(1) 73-84, 2012. The conventional methods for replicate designs do not take into account the impact of acceptance range widening on the sample size calculation.

- The sample size calculation for 2x2 cross-over designs or parallel designs are often not justified adequately or presented with sufficient detail. For more information on sample size calculation see Julious S.A. "Sample sizes for clinical trials with Normal data." *Statistics in Medicines* 2004; 23: 1921-1986. The sample size should be based on the desired power (e.g., 80 or 90%), consumer risk (5%), pre-defined equivalence margins (usually 80 – 125%), expected difference between formulations (e.g., 5%), and the expected inter-subject variability for parallel designs or intra-subject variability for the cross-over designs.
- The number of subjects that are added to the sample size calculation to compensate for potential dropouts or withdrawals should be realistic and consistent with the study design (e.g., more dropouts expected in longer studies) and tolerability profile of the drug.
- The sample size calculation is sometimes based on assumptions that are not consistent with differences observed in previous pilot studies and the expected variability based on previous pilot studies or the variability described in the literature.
- Sample size is sometimes calculated to detect a difference between treatments instead of being based on a calculation aimed to show equivalence.
- In case of a parallel design, it is extremely important to obtain balanced groups in all demographic characteristics that might impact the pharmacokinetics of the drugs. The methods employed to ensure balanced groups are generally not described in the protocols.

STATISTICAL ANALYSIS

- The protocol should indicate the software to be used for the statistical calculations.
- The factors to be included in the ANOVA should be defined in the protocol.
- The calculation of the 90% confidence interval (CI) of the mean test/comparator ratio for the primary PK parameters should not be confused with the two one-sided t-tests employed to reject the null hypothesis of non-equivalence. The end result is the same, but these are not the same calculations.
- In cases of two-stage designs, the consumer risk should be adjusted to preserve the type I error at or below 5%. It is not acceptable to use 90% confidence interval in both the interim and the final analyses. Due to the statistical complexity of the alpha level expenditure in two-stage bioequivalence cross-over trials, two stage designs are not encouraged and, if used, the design should be as simple as possible e.g., with equal sizes in both stages. The Applicant should demonstrate that the consumer risk is not inflated above 5% with the proposed design and alpha expenditure rule, taking into account that simulations are not considered sufficiently robust and analytical solutions are preferred.
- While in cases of 2x2 cross-over or parallel trials only the data of subjects who complete the study can be considered for the pharmacokinetic and statistical comparison, in cases of replicate designs, data from subjects who did not complete all periods of treatment may still be considered. All subjects with two observations of the comparator product should be considered for the calculation of the intra-subject CV of the comparator. Those subjects with at least one observation

for the test and one observation for the comparator should be considered for the average bioequivalence assessment.

- The statistical procedure should be conducted without imputing values to the missing observations.
- In those cases where the subjects are recruited and treated in groups, it is appropriate to investigate the statistical significance of the group-by-formulation interaction e.g., with the following ANOVA model: Group, Sequence, Formulation, Period (nested within Group), Group-by-Sequence interaction, Subject (nested within Group*Sequence) and Group-by-Formulation interaction. If this interaction is significant, the study results are not interpretable. However, it is not considered to be correct to report the results of the 90% confidence interval of the ratio test/comparator based on the standard error derived from this ANOVA. If the group-by-formulation interaction is not significant, the 90% confidence interval should be calculated based on the ANOVA model defined in the protocol. This model may or may not include the group effect as pre-defined in the protocol. This depends on whether the group effect is believed to explain the variability observed in the data.
- The *a posteriori* power of the study does not need to be calculated. The power of interest is that calculated before the study is conducted to ensure that the adequate sample size has been selected. Furthermore, the calculated power is often the power to detect differences, which is not relevant. The relevant power is the power to show equivalence within the pre-defined acceptance range.
- In case of multiple comparisons, the consumer risk should be adjusted. Another option is a hierarchical approach, where it is required first to show BE in the easiest or most desired comparison (e.g., when a dispersible tablet is administered with a glass of water) and if bioequivalence is shown, then the second comparison is performed (e.g., when the dispersible tablet is administered dispersed in the glass of water). However, if the first comparison is not able to show equivalence, the second is not conducted.
- In case of dose-normalization when the same dose is not administered in both treatments, it is not necessary to normalise all PK values individually. It is possible to dose-normalise only the point estimate of the ratio T/R used for the calculation of the 90% CI. Dose-normalisation of the point estimate of the test/comparator ratio in log scale is done by adding the $\ln(\text{dose of comparator/dose of test})$.
- It is not necessary to calculate the non-parametric 90% CI of T_{\max} . A numerical comparison of the median values and its range is considered sufficient.

EXCLUSION OF DATA

- In order to exclude the pharmacokinetic results of those subjects who vomit during the study, the protocol should define in hours the value of two times median T_{max} (as documented in the literature) since the decision to exclude or include the subject should be made before analysing the study samples and as soon as possible, when emesis occurs. This time may differ for different drugs of a fixed dose combination.
- Statistical tests to identify “outlier observations” are not acceptable in case of parallel or 2x2 cross-over designs. Therefore, these tests should not be conducted. In case of replicate cross-over designs where the acceptance range is widened based on the intra-subject CV of the comparator product, outliers may inflate the estimation of intra-subject variability of the comparator and outliers should be investigated. The outliers of interest in case of a replicate design are not those subjects that are discordant with the rest of subjects, because if the subject behaves in a similarly discordant way in both periods, the behaviour is confirmed. The outliers of interest in a replicate design are those that behave differently in the periods where the comparator product is replicated and, consequently, inflate the intra-subject variability of the comparator product. Then, the application of a conventional statistical test is not sufficient to detect the outliers and the type of outlier should be discussed. In any case, a sensitivity analysis with and without the detected outliers is considered essential to assess the impact of the outliers.
- Re-dosing in case of suspected outliers is not considered acceptable as a method to confirm the outlier behaviour of the subject.
- The clinical and analytical investigations conducted for suspected outliers are usually conducted only in the subjects that are considered as outliers or at least with more emphasis on those subjects, and this is not considered correct because all subjects should be treated equally. The findings that may justify an outlier behaviour might be found also in subjects that do not behave as outliers if these subjects are investigated in a similar fashion.
- The exclusion from statistical analysis due to very low concentrations observed following administration of the comparator product requires the pre-definition of what is considered to be ‘very low’ in the protocol (e.g., in accordance with EMA guideline, <5% of the geometric mean of the other subjects).
- A single missed blood sample should not be considered as a reason for the exclusion of a subject by the principal investigator. If such a decision is to be made, it must be made prior to the initiation of the bioanalytical portion of the study.
- Subjects with C_{max} in the first sampling point should not necessarily be excluded. The study should be designed so C_{max} does not occur at the first sampling time, but if that happens only in a small number of profiles the data can still be considered.
- The samples of subjects that do not finish the study because of an adverse event should be measured for safety reasons in order to investigate whether the withdrawal was related to unexpectedly high drug concentrations causing the adverse events.
- Subjects should not be excluded simply because three consecutive sampling points are missing.

BIOANALYTICAL METHOD

- The list of parameters for the validation of the bioanalytical method that are included in the clinical protocol is often not complete.
- Incurred sample reanalysis (ISR) does not need to be conducted in 7% of the study samples exceeding 1,000 samples, 5% of these samples is sufficient.
- ISR samples should not be selected based on concentrations above the low QC.
- Manual integration should be performed only when the instrument is not able to correctly integrate the peak with the default settings and it cannot be solved by refining the settings such as the expected retention time.
- Plasma samples from subjects that dropout or are withdrawn due to an adverse event should be analysed for a complete safety analysis of the data, in order to assess if the dropout or withdrawal is due to an adverse event that might be related to high concentrations. These data should not be considered for efficacy or bioequivalence analysis but, they are supportive for the safety analysis.

OTHER ISSUES

- The protocol should include the centres where the study is going to be conducted.
- All studies should be monitored. It is not acceptable to state that the study may be monitored at the discretion of the sponsor by any of its representatives.
- Monitoring and auditing activities should not be confused because they are different activities. Monitoring is conducted by the sponsor or a CRO contracted by the sponsor and the audits are performed by the centre where the study is conducted.