Doc Ref: EMADOC-1700519818-2810632
Case No.: EMA/SA/0000170378
Human Medicines Division

Initial Qualification Procedure List of Issues and Replies

Virtual Control Groups

General introductory remark by the applicant: the EMA SAWP for Virtual Control Groups request was submitted by Synapse together with a group of five pharmaceutical companies (see Briefing Document chapter 2.3) in the frame of the IMI eTRANSAFE project. This team represents the core of a larger consortium, which submitted a project proposal on VCGs to the Innovative Health Initiative (IHI) in January 2024 as a follow-up of eTRANSAFE. The project proposal was positively evaluated and the new IHI project with the name VICT3R ("Developing and implementing VIrtual Control groups To reducE animal use in toxicology Research", grant agreement number 101172693) started on September 1st, 2024 (VICT3R | IHI Innovative Health Initiative (europa.eu); VICT3R - Developing and implementing Virtual Control Groups to reduce animal use in Toxicology Research). It is the intention of the applicant to further develop and assess the VCG concept in the frame of this IHI project. Therefore, in many replies and comments provided further below reference is made to the new VICT3R project. We intend to share this feedback from EMA and all further interactions with the VICT3R consortium partners.

In this context, we would also like to clarify that the initial VCG database developed before the start of VICT3R for which information was submitted in the Briefing Document was called "VICOG database". This database can be considered as the prototype of the VCG database which will now be further developed in VICT3R. Wherever reference is made to the database submitted in the Briefing Document, we used the term "VICOG db", whereas for comments and statements referring to future developments the term "VICT3R db" was used.

List of issues to be addressed in writing and during the discussion meeting

Based on the coordinators' reports the Scientific Advice Working Party (SAWP) determined that the Applicant should discuss the following points, before advice can be provided:

Issues to be addressed in writing and during the discussion meeting

1. VICOG DB: Clarifications should be provided as to the consideration for missing critical endpoints in the database as endpoints listed in the OECD Test Guideline 407 for rodents

and 409 for non-rodents do not encompass all endpoints recommended in the EMA guideline on repeat dose toxicity (CPMP/SWP/1042/99 Rev 1). This guideline includes ophthalmology in all species (only included in OECD TG 409 for non-rodents), urinalysis in all species (optional under OECD TG 407), electrocardiography in non-rodents (not included in OECD TG 409). These discrepancies should be addressed by the Applicant.

Applicant's Reply: The initial development of the VCG concept and the VICOG database focused indeed on the endpoints listed in the OECD Test Guideline 407 for rodents. During the further development of the VICOG database we expanded the focus to other species including non-rodents. Data for urinalysis (optional in OECD 407) are therefore already available in the latest version of the VICOG database. Also, the applicant concurs that the use of VCGs for non-rodents particularly for CoU II & III and some specific study types will require additional endpoints in the database. Therefore, relevant SEND domains that represent data on electrocardiography will be included in further developments of the database. The CDISC standard for collecting ophthalmology data is available for clinical data but not yet included for preclinical data in the latest version of SEND (SENDIG version 3.1). As soon as this is available, we will expand data collection for these additional domains in VICT3R. Given the fact that most sponsors currently do not use concurrent controls in DRF studies with non-rodents, this momentaneous shortcoming of the database is considered not relevant for CoU I.

2. Matching criteria:

    a. Specific clarifications to be made regarding the matching criterium on route of administration to encompass also specific subcategories. For example, for oral administration there might be gavage or dietary administration and for intravenous administration, one might consider bolus injections or infusions (considering specific infusion time). This is not specifically mentioned in table 10 of the briefing document and variably applied in the qualification studies and as such it is not clear whether this is indeed systematically and rigorously controlled. As this can certainly impact measured parameters, it is considered essential for matching.

    Applicant's Reply: the applicant agrees that subcategories of the RoA should be considered, when selecting VCGs from the HCD database. For the submitted qualification studies, this was indeed the case: for the reanalysis of the legacy studies with oral gavage we only used HCD from oral gavage studies, i.e. no studies with dietary administration were included. This was also the case for the intravenous cynomolgus legacy studies. There, only HCD from intravenous bolus injection studies were used and no HCD from continuous infusion studies was included. (see Briefing Document, chapter 2.7.2 & 2.7.3).

    b. In the presented qualification studies, an additional matching criterium has been listed as compared to table 10 (e.g. housing). The Applicant should include this in a standardized list of matching criteria.

    Applicant's Reply: the applicant agrees to the proposal. It was not explicitly stated for the re-analyzed rodent studies, since the housing conditions were identical for all chosen studies and selected animals. However, for future extension of the qualification procedures, this could become an important criterion. As part of VICT3R, we also intend to assess the influence of the housing conditions on the

control animals in the future VICT3R database to obtain better insight into which parameters might be affected by different housing conditions.

c. The matching or further filtering criterium for non-rodents based on pre-values should be further specified. Whilst for non-rodents, it is mentioned that age and/or bodyweight can be used in combination with pre-values, the process for doing so is not described in Table 10 of the briefing document nor is it unambiguously described elsewhere in the briefing package. This should be further addressed.

Applicant's Reply: The applicant considers the pre-values as a potential improvement of the matching procedure for large animals. It is true though, that initially the matching approach using pre-values was not applied to the re-analyzed NHP studies submitted in the Briefing Document. However, during the retrospective re-analyses reported in our Briefing Document, it became clear that pre-values can help detecting and explaining differences in the statistical results with CCG or VCG, particularly in cases where pre-values do not sufficiently match with values of the selected VCGs. Therefore, pre-values turned out to be potential improvements for the matching procedure to identify outliers or extreme values of VCG animals which do not sufficiently match with pre-values of the CCG and to deselect them as potential VCG animals. The applicant will perform an in-depth evaluation of the matching procedure using pre-values for dogs and NHPs as part of VICT3R to qualify VCGs for the CoU II & III and assess whether the use of pre-values improves the matching procedure and if so, which pre-value parameters are the most important.

d. Regarding the overall application of the matching (and filtering) criteria, whilst in table 10 of the briefing document the difference is made between essential parameters and further aspects to be considered, figure 18 of the briefing document describes the approach where the initial filtering of the historical data is based on the application of all criteria (essential and further aspects to be considered) and the obtained dataset is being matched with the range of the initial body weight of the animals in the treatment groups for rodents. For non-rodents, age is used for matching with additional parameters (pre-values or baseline values) that may be considered for matching. The Applicant highlights that further accumulation of data may lead to less stringent criteria if the specific criteria have a neglectable influence on the study outcome. Whilst this approach is acknowledged, it is at present unclear what the disadvantage would be to apply this standardized set of matching criteria systematically as this would clearly limit uncertainty. Therefore, it is recommended that the Applicant – certainly at this point in time – applies a standardized set of matching criteria in a systematic approach (that needs to be clearly defined for rodents and for non-rodents separately).

Applicant's Reply: the applicant acknowledges that there is a perceived mismatch between Table 10 and the matching procedure as described in the Briefing Document, chapter 2.7.2, chapter 2.7.3 and Figure 8 (the applicant assumes that Figure 8 was meant by the EMA instead of Figure 18 mentioned in issue 3d). The more stringent criteria were used for the re-analyses of the rodent studies, since the applicant considered that the data collected so far was insufficient to explore

and assess the effect of less stringent criteria. Using this standardized set for matching resulted in a high reproducibility of the outcome of the re-analyzed legacy studies. However, the disadvantage of using such a stringent standardized set of criteria in a canonical way for all further VCG matching approaches will most probably limit the applicability of the historical data in situations where data is scarce for certain criteria. For example, the vehicle composition used in some studies may require additional ingredients for preparing solutions or stable suspensions. The effect of these additional ingredients on control animals will necessitate further statistical analyses to assess whether it influences clinical pathology or histology parameters or whether certain small variations of vehicle composition can be grouped together.

However, the applicant concurs that the application of the more stringent standardized approach as described in chapter 2.7.2 should be used until there is evidence that leaving out certain matching criteria will not increase uncertainty of the analysis. The applicant also agrees that a clearer definition is necessary for the matching procedure of non-rodents.
For the re-analyzed study in NHPs, less stringent selection criteria were used for the matching of vehicles (see chapter 2.7.3, p. 39) and pre-values were not used for matching as described further above. Nevertheless, the results of both re-analyzed studies in NHP point towards a reproducibility of findings even with less stringent matching procedures. Notably, the threshold doses (NOAEL, MTD) were reproduced. Therefore, the applicant still considers a less stringent matching procedure appropriate for CoU I for NHPs.

e. While a relatively simple matching procedure has been proposed, many others can be considered – with multiple matches within each VCG study: Arguments should be provided to support why the current approach was chosen as the `best'.

Applicant's Reply: the applicant does not claim that the current matching procedure applied for the reanalysis of legacy studies is the "best". Rather, we described a stringent and conservative approach within the limits of data availability. This was also partially based on the results from the proof-of-concept study described in chapter 2.6 of the Briefing Document indicating that, by choosing only a subset of the criteria considered in the matching procedure, it was possible to replicate a significant portion of the original statistical results obtained with CCG. However, to further approximate our matching procedure for VCGs to the procedure applied to animals in the CCG of the legacy studies, we decided to be even more conservative than in the proof-of-concept study by selecting only animals from the HCD which fulfilled all criteria of the original legacy study protocol.
The matching procedure is, therefore, essentially based on the criteria set forth in the original study protocol, i.e. it replicates the strain, study duration, age of animals (NHPs)/initial body weight (rats), route of administration, and the other listed parameters as much as possible. Particularly for the rat studies, matching of initial body weight proved to be an important parameter, as recently shown in Gurjanov et al. 2024.
Admittedly, matching based on pre-values for larger animals has so far not been systematically explored by reanalysis of legacy studies. Using pre-values for

matching would allow for "multiple matching", e.g. the selection criteria could consider key clinical chemistry or haematology parameters for matching. However, regarding rodent studies, where pre-values are usually not available, the applicant is not sure how "multiple matching" could be applied, since the only initial parameter determined at day 0 of a study (i.e., before the first treatment with the test item) is the body weight.

In case larger sets of HCD are available, one might be able to choose different matching criteria for different VCGs in the same study. Within the VICT3R consortium we will explore and evaluate other matching procedures. We expect a relaxed matching procedure might yield an overly heterogeneous VCG group, as compared to the homogeneity of the animals selected for the dose groups, and thus affect the sensitivity of the analysis. In contrast, more stringent matching criteria might yield too few animals to consider for the VCGs creation. Additionally, there may be cases where data does not allow for a good matching within the available VCG data, and thus using the VCG study design may need to be rejected all together. For instance, when a new breed of animals or an entirely new type of vehicle is intended to be used. Thorough investigation of the data and the data quality is required to address these questions. This is one of the goals within the VICT3R consortium.

3. Specific remarks with regard to the cynomolgus monkey Sanofi qualification studies:

    a. It should be noted that monkeys between 3 to 4 years are categorized as pubertal whilst 2 years is prepubertal with varying maturation of several organ systems and hence there is a potential impact on measured parameters (see ICH S11). It is not clear what the proportion of ages in the CCGs is compared to the VCGs. This should be clarified, and the impact assessed in the light of the above stated differences in maturity.

    Applicant's Reply: The NHPs selected as VCGs were all control animals previously used in standard 4-week toxicity studies. It is acknowledged that some of the animals were still very young, with an age of approximately 2 years. However, according to their body weight, it was concluded during the execution of the original study that they already reached puberty to be qualified for use in typical 4-week toxicity studies. With the availability of a bigger data set in VICT3R the effect of confounding factors such as age will be explored in more detail.

    b. There is no information on the possible matching or filtering based upon pre-values. This should be clarified.

    Applicant's Reply: it is correct that pre-values were not used as matching criteria. However, after obtaining results of the reanalysis with VCGs, it became clear that pre-values could help explain statistical result differences between the CCG and VCG analysis. Therefore, pre-values could improve the matching procedure, for example by identifying outliers in the pool of animals used for VCG selection whose pre-values do not sufficiently match with pre-values of the dosing groups. Such an application of pre-values as additional filtering and matching criteria will be explored in VICT3R.

4. The use of VCGs within CoU1 i.e. for non-GLP DRF could be considered potentially acceptable. However, the impact of the use of VCGs on the actual purpose of DRF studies, namely, to define dose levels for later studies, is not directly addressed and could benefit from some further detail.

Applicant's Reply: the effect of using VCGs for defining dose levels for later studies was directly addressed in the respective results section for each study in the Briefing Document (e.g. for study A: "The NOAEL was still considered to be the highest dose of the study and non-adverse effects were seen in the liver for this test item." or for study B: "Therefore, the conclusion of this study regarding STD10 remained unchanged after replacing the CCG with VCGs"). In the legacy DRF study D, no dose limiting toxicity was observed. This observation did not change after replacing HCDs with VCGs.
The upper dose level for later studies derived from results of DRF studies is usually determined by severe toxicities such as mortality in the highest dose group(s) or severe macroscopic or microscopic pathologies. It is highly unlikely that the use of VCG will have an influence in these findings. This is particularly true for non-rodents, where many sponsors anyway use no concurrent controls in DRF studies.

5. At present, historical data should not be older than 5 years.
How can one be sufficiently confident to obtain a trustworthy result in the future based on the study of the historical control data in case of full replacement of concurrent controls (CoU III)? What can be done to quantify the level of confidence and to put appropriate sensitivity analysis in place?

Applicant's Reply: the HCD used for the re-analysis of legacy studies fulfilled the criterion of the 5 years' limit. While it is not debated that the data will underly certain shifts and drifting over time for various reasons, the scientific rationale for a limit of exactly 5 years is so far limited. The amount of HCD datasets we have collected so far in the VICOG database was not sufficient to allow us to perform the requested sensitivity analysis. Larger datasets representing longer periods will be needed to identify potential trends over time. Such analyses will be performed in the context of the VICT3R project. Using the future VICT3R DB different time windows will be tested when creating VCGs and sensitivity analyses will be performed to quantify the level of confidence (see also reply to issue 7b).

6. Not only the median, or the mean results, should be rendered to understand the practical properties and way forward with the VCG results, but also the spread and tails of that distribution to understand how far off one can get with what probability.

Applicant's Reply: the applicant agrees that the statistical analysis needs to be extended. We plan to systematically assess how the clinical pathology parameters distributions vary across studies, similarly to what is displayed in Figure 3 in the Briefing Document, and to identify the factors influencing those distributions. This will allow us to further inform the selection of the relevant factors to consider when building VCGs. The VICT3R consortium aims at addressing those statistical aspects to refine the curation of the VICT3R database by identifying possible outliers, unexpected trends or patterns, correlations between the various measured parameters and to compute prediction intervals.

7. It may be considered adding a reduced current control group to new studies rather than no current control data at all, to help assess the assumptions the VCG relies on.

Applicant's Reply: the applicant agrees that a partial replacement of concurrent control in new studies (hybrid design) is most probably the best way to proceed particularly for CoU II and III.

a. Could data collected in this way become part of the 'learning curve' of the VICOG database?

Applicant's Reply: yes, data from control animals will need to be collected regularly to feed the database used to create VCGs. The applicant sees the necessity that the future VICT3R database will require continuous updates in order to avoid data shifts away from current data caused by genetic drift, changes in procedures or assays used etc.

b. What are the plans regarding sensitivity analyses?

Applicant's Reply: In the context of VICT3R, we plan sensitivity analysis to tackle three angles of analysis. First, sensitivity analyses will address factors for which there is no stringent matching procedure. For example, species and strain can be matched exactly between VCG and treated animals and no sensitivity analysis will be necessary. For other factors, such as the body weight of the animals at study start or the year when the study was conducted, sensitivity analysis will determine the maximal body weight range and the maximal time window for which there is little or no influence or impact on the study conclusions. Second, sensitivity analyses will also be performed to assess the influence of outliers in VCG and CCG. Finally, sensitivity analysis will also be performed to evaluate the impact of missing data, in case measurements collected in treated animals are not present in the corresponding VCG.

c. How often would one require to go back to a full concurrent control to make sure the assumptions stay in line with the evolving reality?

Applicant's Reply: this point will be addressed in-depth in the VICT3R consortium. Although we cannot answer how frequently a full CCG will be required, we already foresee that it will be occasionally necessary to conduct studies with a full CCG in various facilities to continuously feed the database of control animals. Besides, studies with a full CCG will still be necessary if the matching criteria for VCGs cannot be fulfilled due to the lack of matching animals from the future VICT3R database, e.g. if an entirely new vehicle is planned to be used. Moreover, the evolution of the database could be followed up closely for time-dependent shifts or trends in key parameters away from the animals in the treatment groups.

d. Would a new trial with concurrent controls still be feasible after the VCG trial has failed?

Applicant's Reply: it would depend on the type of failure in the study. If VCGs cannot be created because matching has failed, for example if one important characteristic of the planned study (e.g. vehicle type) is not represented in the pool of controls used to build VCGs, a CCG would need to be incorporated in the study. Since matching occurs before study start, it should be possible to still include CCGs in time before the start of the n-life phase of a study. Therefore, we do not consider this as a failure. Other types of failure would need to be assessed case by case. In

general, any repetition of an animal study has high legal hurdles, i.e. it needs to be assured that the risk of failure is reduced to a minimum.

e. How does one incorporate the historical data into the analysis of a new study?

Applicant's Reply: the incorporation of historical data into the analysis of a new data will be performed in the same way it is currently done in the evaluation of a study, including the calculation of reference ranges and assessment of biological significance of data within it. This is described in chapter 2.7.1, p. 24 of the Briefing Document.
However, it is worthwhile stating that VICT3R has the potential of influencing the way HCD are used in future toxicity studies independent of its use for VCGs. Currently, HCD are used without any further preparatory characterisation or analysis, e.g. many sponsors group HCD together over a period of 3-5 years without reflecting animals' initial body weights, vehicle composition, etc. VICT3R's statistical analyses might identify criteria, where grouping HCD across certain parameters might introduce a bias in the data and should thus be avoided.

f. Are the historical data downgraded?

Applicant's Reply: the historical data are by no means downgraded. The process of study evaluation will still require historical reference values as described in the referenced publication and the briefing book (see chapter 2.7.1). See also the importance of HCD for assessing a dual VCG study design in the reply to issue 7g. In the CoU I it should, however, be mentioned, that HCD and the resulting reference values (ULN, LLN) are rarely used to evaluate these studies.

g. What to do in case of conflict with the new experiment?

Applicant's Reply: a "case of conflict with the new experiment" could particularly occur in a dual VCG study design (using the definition of Golden et al. 2024). In this scenario, VCGs are used in parallel with a CCG, potentially leading to conflicting statistical analysis when treatment groups are compared to either control groups. If such a conflicting result is considered of biological relevance it could be subsequently analysed in light of the HCD to discern whether the VCG or the CCG better represents the general distribution of data. In addition, comparing measurements from animals in the treatment groups (particularly for parameters showing an off-set of the normal distribution which is not dose-dependent) will elucidate whether specific environmental conditions had influenced the study.

Based on the results of these described analyses, it is the decision of the study director in discussion with SMEs to decide which control group (VCG or CCG) will be used for the final analysis and the study conclusions.

h. Please discuss considerations regarding the type I error rate (control) in the case of VCG?

Applicant's Reply: When testing the null hypothesis (no difference between group means) type 1 error can occur due to chance during random sampling of VCG. For certain parameters measured the VCG could then be different from the treatment groups. This could lead to the false conclusion that a treatment effect exists.

Because many factors (matching procedure selected, confounders, presence of outliers ...) could impact the conclusions drawn, we propose that the final interpretation of the comparisons involving VCGs should not only rely on null hypothesis significance testing and corresponding p-values but also on the effect size (which quantifies the magnitude of the difference between the groups means) and its precision. This approach will downweigh the importance of the Type I error. It is worth mentioning that the same approach can be considered when comparing treatment groups to CCG.

8. Harmonisation of methodology

   a. Please discuss strategies to harmonize matching criteria between experimental sub-categories (e.g. oral gavage and bolus IV) and perhaps more importantly, rodent and non-rodent species. From a methodological/qualification perspective, a uniform approach is more appropriate to select animals independent of species as this has the lowest chance of variance, even if this would result in matching variables that are not, or not often used for rodents. At minimum, this would result in transparent reporting of used matching criteria and selection of VCG animals. If additional non-rodent criteria are essential and cannot be used for rodents, this distinction should be made clear in a separate standard operating procedure.

Applicant's Reply: the applicant concurs that it will be extremely important to harmonize the different matching criteria. Within the VICOG database presented in the Briefing Document, this harmonization work has already been started. For example, harmonization and categorization of vehicles into more general categories of oral (aqueous solutions, oils and emulsions) or intravenous vehicles (harmonising different description for physiological saline) was undertaken. This work will continue in the VICT3R consortium and will also include a qualification procedure and methodology, resulting in standard operating procedures which eventually may differ between species.

   b. Approaches between the two companies submitting the data are – in appearance – different. The Applicant should discuss whether exactly the same protocol was followed by both companies and/or highlight and discuss where inter-operator variance was introduced. This variance should either be resolved to arrive at a single harmonized SOP or should be described in detail as part of the method. This will be particularly important when pursuing CoU II and III.

Applicant's Reply: the applicant concurs that there were slight differences between protocols used by the two companies performing reanalysis of legacy studies. However, the different protocols also provided the opportunity to apply different methodologies regarding the matching criteria. The overall goal of both protocols was the same: to assess whether the main study conclusions were affected by replacing the CCG with VCGs, which was confirmed by the different methodologies. That said, as part of VICT3R, harmonized SOPs need to be developed so multiple companies follow exactly the same procedure, even though probably individually for each species, or at least for rodents and non-rodents.
Interoperator variance was avoided whenever possible. Rodent studies from Bayer were re-analysed by the same study director who conducted the original study

analysis, effectively removing interoperator variance. Studies from Sanofi were evaluated by different study directors. The newly reanalysed mice study added in the Appendix was also evaluated by different study directors, due to retirement of the original study director.

9. Context of use

    a. While it is intuitively assumed that mouse VCGs would yield similar results as rats, this was not confirmed. Please discuss whether the rat dataset is representative for all rodent species and justify or provide strategies on how to demonstrate VCG for all rodent species.

    Applicant's Reply: the applicant concurs that from a biophysiological perspective results from mice would most probably yield similar results as shown for different strains of rats. The applicant has in the meantime extended the analysis to three further DRF studies with mice. Results of the three reanalysed mouse DRF studies are added to this document as Appendix. In addition, the applicant intends to extend the analysis within the VICT3R consortium to further species.

    b. The CoU for non-rodents has been demonstrated by using a limited sample of NHP data. The Applicant is requested to discuss the validity of VCG use under CoU I for all non-rodents and to provide approaches to support other non-rodent species if necessary.

    Applicant's Reply: see reply to issue 9c below.

    c. CoU I is not defined optimally. The Applicant is requested to comment on the suggestions to amend CoU I and to propose a definitive text for discussion.

    Applicant's Reply: the applicant agrees that based on the currently available re-analysed study data, the CoU I definition should be modified. The applicant proposes to limit the CoU I to rats, mice and NHPs by redefining the CoU I as "Application in dose-range finding (DRF) non-GLP studies with rats, mice and NHPs (for NHPs in cases where sponsors do not use concurrent controls)". Despite the small number of NHPs studies, the applicant still considers that CoU I is applicable to NHPs, since the risk of uncertainty introduced using VCGs in DRF with NHPs will be lower than the risk caused by using no control group at all.
    Within the VICT3R project we plan to reanalyse the effect of VCG on studies using other non-rodent species, primarily focusing on dog and NHP studies.

10. Fit for purpose

    a. The Applicant has presented data primarily for selection of rat VCGs. The sample size, while sufficient for a non-GLP compliant study, is nevertheless small. The Applicant should discuss what strategies will be employed to increase this sample, particularly since only 2 companies participated in this exercise. To be fit for purpose, a method should be robust, uniform, reproducible and usable for all users.

    Applicant's Reply: the applicant is not entirely sure, what is considered as small sample size. If the number of re-analysed studies is meant by sample size, the applicant agrees that more studies are needed for a qualification particularly for CoU II and III. Within the VICT3R project the number of reanalysed studies will be

increased, both in number and by test facilities, companies and species.
If, however, the interpretation of sample size is related to the pool of animals from which the VCGs were constructed, the applicant disagrees that the sample size is small. For example, in the reanalysis of the three rat studies (study A, B &C), the pool consisted of 119 female and 126 male animals (see Briefing Document chapter 2.7.2, p. 27). Given that initial randomisation in a conventional 4-week rat studies is typically performed with a pool of only about 40 males and females, the pool of animals used for VCG generation exceeds that of conventional studies significantly and is, therefore, not considered to be small.

b. While there are no regulatory barriers to using VCG in DRF studies (and this is indeed encouraged), including DRF studies in NHP, the sample size to demonstrate that the method is fit for purpose in NHP was limited. The Applicant is requested to propose strategies to address this, considering the need for very robust data sets in order to qualify the VCG approach for CoU II and III.

Applicant's Reply: The applicant acknowledges that the current dataset for non-rodents, particularly the sample data for NHPs from Sanofi, is limited. To address this limitation and ensure the robustness of the VCG approach for CoU II and III, the applicant proposes several strategies. Firstly, the VICT3R consortium will be instrumental in expanding the data set. The aim is to aggregate data from multiple partners (including CROs having access to large amount of data from control animals), enhancing the diversity and volume of information available for NHP studies. Secondly, leveraging advanced statistical modelling and machine learning techniques can help analyse existing data more effectively demonstrating their fit for purpose. Finally, ongoing dialogue with regulatory bodies will be essential to ensure that the evolving dataset meets the rigorous standards required for qualification.

c. No tandem experiment was used to corroborate that VCG data would confirm results of the DRF in terms of estimating NOAEL, dose levels for the GLP compliant repeat dose toxicity study or enrich studies by identifying toxicity parameters of interest. The Applicant is requested to discuss why this approach was not chosen or to comment on whether this is subject of a future study.

Applicant's Reply: the applicant assumes that "tandem design" describes either a dual control or a hybrid design according to the definitions provided by Golden et al. 2024. It is true that such design approaches have yet not been used for GLP compliant repeat dose toxicity studies for the following reasons:

- The VICOG database as well as the applied algorithms for matching have not undergone computer system validation according to OECD 17 and can, therefore, not yet be used in a GLP context.
- Inclusion of non-GLP parts in a GLP study always carries the risk that the whole study will be impaired with regard to its GLP compliance, particularly when the general outcome of the study might be affected by the non-GLP parts.
- Study directors first need evidence for the feasibility of the VCG approach

before they are willing to introduce it into prospective studies. In a first step, the applicant strived to provide this evidence through the reanalysis of legacy study.

Given these reasons the applicant decided to first request scientific advice for the CoU I based on the retrospective analyses. However, it is clearly the intention of the applicant to explore the feasibility of the VCG approach also in prospective GLP studies, probably, first through a dual control design followed by a hybrid design (see also reply to issue 11).

11. The Applicant is recommended to engage further with EMA to start a prospective phase of data collection and submission preferably for a modified CoU II-approach involving a dual control design (see Golden et al., 2024) to further build confidence. This is considered possible as recommended in the EMA Guideline on the principles of regulatory acceptance of 3Rs testing approaches (EMA/CHMP/CVMP/JEG-3Rs/450091/2012) section 5.4.3 which describes the possibility of a voluntary submission of data or 'safe harbour' period that allows for evidence building in support of a qualification package. The specifics of setting up this process can be further discussed. The Applicant should comment.

Applicant's Reply: the applicant welcomes the proposal for further engagement with EMA. In the frame of setting up the IHI VICT3R project aiming to further explore and assess the VCG concept, EMA was contacted with a request to participate as project partner or in the Scientific Advisory Board. In its reply, EMA has clarified that it rather intends to add value to the VICT3R project through existing mechanisms of interaction (e.g. ITF or qualification).

Therefore, the applicant (now also representing the IHI VICT3R consortium) highly welcomes the proposal to explore the possibility of a voluntary submission of data under safe harbour conditions and to further engage with EMA on the development of the VCG concept. The applicant is looking forward to further discuss the details of the process and the potential design of future dual or hybrid studies with EMA during our meeting.

In this context the applicant would also like to discuss with EMA, whether there is a possibility to extend the safe harbour approach to FDA through the Parallel Scientific Advice Program between EMA and FDA, since most toxicity studies are also performed for future submission to FDA.

A Appendix

Summary of reanalysed mouse DRF studies - Study Overview

Study	Species	Duration	Dose Groups[1]	RoA	GLP (Y/N)	CCG (Y/N)	HCD period	VCG Matching Procedure	CCG & VCG group size	VCG Repetition
H	Mouse (CD1)	2-week	LD_M, MD_M, HD_M LD_F, MD_F, HD_F	Gavage (once daily)	N	Y	≤ 5 years	Based on initial body weight of treatment group animals	5 M/5 F	100x
I	Mouse (CD1)	2-week	LD_M, MD_M, HD_M LD_F, MD_F, HD_F	Gavage (once daily)	N	Y	≤ 5 years	Based on initial body weight of treatment group animals	5 M/5 F	100x
J	Mouse (CD1)	2-week	LD_M, MD_M, HD_M LD_F, MD_F, (HD F)[2]	Gavage (once daily)	N	Y	≤ 5 years	Based on initial body weight of treatment group animals	5 M/5 F	100x

Method:

The VCG selection procedure for the three mouse DRF studies essentially recapitulated the procedure described in the Briefing Document for the rat studies (chapter 2.7.2).

The three selected studies H, I and J were performed between 2018 and 2019. The HCD was selected from the period of 2015-2019. From the pool of HCD, 26 male mice and 22 female mice were eligible for VCG creation, based on the criteria of the original study protocols of the three studies. Matching was performed based on initial body weight. Despite the small number of animals, background incidences for macroscopic and microscopic findings were calculated and compared to the findings reported for the treatment groups.

Results:

Study H

This repeated dose systemic toxicity study in CD1-Mice with a 2-week daily administration by gavage was conducted to generate initial data on tolerability and toxicokinetic of the test item.

The use of VCGs resulted in the identification of a significant decrease in red cell hematology parameters and a significant decrease in creatinine. However, these decreases had no effect on the NOEL/NOAEL setting. The microscopic finding for the kidney was not reported in control animals. Therefore, a background incidence could not be calculated.

NOEL and NOAEL were determined to be above the HD of the original study. These results were not changed by using VCGs, i.e. dose setting for subsequent pivotal systemic studies was not influenced by using VCGs.

Test item related mortality							
None							
Clinical findings							
None							
Quantitative parameters							
Original noteworthy findings with CCG				**Noteworthy findings after replacing CCG with VCGs**			
Parameter name	**Increase (+) decrease (-)**	**Sex (M/F)**	**Starting dose**	**Compared to CCG**	**Increase (+) decrease (-)**	**Sex (M/F)**	**Starting dose**
Body weight	-/-	-/-	-/-	consistent			
Water intake	-/-	-/-	-/-	consistent			
Food intake	-/-	-/-	-/-	consistent			
Hematology							
- **HGB**	-/-	-/-	-/-	- new	(-)	M	MD
- **MCHC**	-/-	-/-	-/-	- new	(-)	F	HD
- **MCV**	-/-	-/-	-/-	- new	(-)	M	HD
- **RBC**	-/-	-/-	-/-	- new	(-)	M	HD
Clin. Pathol.							
- **CREAT**	-/-	-/-	-/-	- new	(-)	M	LD
Organ weights	-/-	-/-	-/-	consistent			
Noteworthy necropsy & histopathological findings							
Kidney: tubules, cytoplasmic alteration (up to grade 2) HD_F							

Study I

This repeated dose systemic toxicity study in CD1-Mice with a 2-week daily administration by gavage was conducted to generate initial data on tolerability and toxicokinetic of the test item.

NOEL and NOAEL were determined to be above the HD of the original study. These results were not changed by using VCGs. Therefore, dose setting for subsequent pivotal systemic studies was not influenced by using VCGs.

Test item related mortality							
None							
Clinical findings							
None							
Quantitative parameters							
Original noteworthy findings with CCG				**Noteworthy findings after replacing CCG with VCGs**			
Parameter name	**Increase (+) decrease (-)**	**Sex (M/F)**	**Starting dose**	**Compared to CCG**	**Increase (+) decrease (-)**	**Sex (M/F)**	**Starting dose**
Body weight	-/-	-/-	-/-	consistent			
Water intake	-/-	-/-	-/-	consistent			
Food intake	-/-	-/-	-/-	consistent			
Hematology - **LEUCO** - **LYMP** - **MCV**	 (-) (-) -/-	 M M -/-	 HD HD -/-	 - not reprod.[1] - not reprod.[1] - new	 (-)	 M	 HD
Clin. Pathol. - **Urea**	 (+)	 M	 HD	 - not reprod.[2]			
Organ weights - **Spleen (abs.)**	 -/-	 -/-	 -/-	 new	 (+)	 M	 HD
Noteworthy necropsy & histopathological findings							
None							

[1] Both parameters (LEUCO, LYMP) were within the limit of normal in the original study. The trend of an increase of LEUCO and LYMPH was also visible with VCGs in both sexes.
[2] The increase of urea was above the ULN in the original study.

Study J

This repeated dose systemic toxicity study in CD1-Mice with a 2-week daily administration by gavage was conducted to generate initial data on tolerability and toxicokinetic of the test item.

Key results of the study are listed below together with the results after replacing the CCG with VCGs.

Evidence for potential target organs identified in the original study by histopathology was strengthened for liver, where an increase of relative and absolute weight was determined with VCGs (missing with CCG), whereas a significant increase of ALAT in the dosed groups in relationship to the CCG was not reproduced with VCGs.

The significant increase of thrombocytes potentially related to the histopathology findings in pancreas was not reproduced with VCGs, whereas new significant effects on hematology were identified with VCGs. Notably, the significant decrease of eosinophils can be interpreted as pharmacodynamic effect of the test item. In the original study with CCG this decrease was not significant.

Microscopic findings were reported for control animals for:

- liver (inflammatory cell infiltrate, grade 1) with a background incidence of 4% and 14% for males and females, respectively
- pancreas (degeneration, grade 1) with a background incidence of 4%

Given the higher grades observed for these findings in the treatment groups and additional findings for these organs, it is concluded that the use of VCGs did not change noteworthy findings in this study.

NOEL and NOAEL were determined to be above the HD of the original study. These results were not changed by using VCGs, i.e. dose setting for subsequent pivotal systemic studies was not influenced by using VCGs.

Test item related mortality							
None							
Clinical findings							
None							
Quantitative parameters							
Original noteworthy findings with CCG				**Noteworthy findings after replacing CCG with VCGs**			
Parameter name	**Increase (+) decrease (-)**	**Sex (M/F)**	**Starting dose**	**Compared to CCG**	**Increase (+) decrease (-)**	**Sex (M/F)**	**Starting dose**
Body weight	-/-	-/-	-/-	new	(+)	F	HD
Water intake	-/-	-/-	-/-	consistent			
Food intake	-/-	-/-	-/-	consistent			
Hematology - **THRO** - **EOS** - **MCHC**	(+)	F	HD	- not reprod.[1] - new - new	 (-) (+)	 M M	 MD HD
Clin. Pathol. - **ALAT** - **TRIGL**	 (+) (-)	 M(F) M	 HD HD	 - not reprod.[2] - consistent	 - (-)	 - M	 - HD
Org. weights - **Kidneys (abs.)** - **Liver (abs.)** - **Liver (rel.)**	(+)	F	HD	- consistent - new - new	 (+) (+)	 M/F M/F	 HD HD
Noteworthy necropsy & histopathological findings							
Lymph nodes	increase in size					F	HD
Liver	hepatocellular hypertrophy (grade 2)					M	HD
	cytoplasmic alteration (grade 1-3)					(M)F	HD
	necrosis/inflammatory infiltrate (grade 1-2)					M	MD
	increased mitosis (grade 2)					M/F	HD
Pancreas	focal acinar degeneration (grade 1-2)					M/F	MD
	diffuse acinar vacuolation (grade 1-2)					M/F	HD
	diffuse acinar atrophy (grade 2)					M/F	HD/MD
Spleen	increased megakaryocytes (grade1-2)					F	HD

[1] The significant THRO increase with CCG was above the ULN of HCD. The trend of an increase in HD of both sexes is also visible with VCGs.

[2] The significant ALAT increase with CCG was below the ULN of HCD. The trend of an increase in both sexes is also visible with VCGs.

Conclusion

A limited number of DRF studies in mice was reanalyzed replacing the CCGs with VCGs. In summary, the results of these re-analyses regarding the conclusions of the studies were in line with the results reported for rats and NHPs in the Briefing Document. Replacing the CCGs with VCGs did not result in changed threshold doses or target organs.

Differences in the significance of parameter changes (hematology, clinical pathology, organ or body weights) did not influence the overall conclusions of the studies. A trend of higher sensitivity was

observed using the VCGs. This observation is in line with a recent VCG publication re-analyzing 14 dog and 8 NHP 4-week toxicity studies (Li et al. 2024).

B References (not yet provided in the Briefing Document)

1. Gurjanov, A., Vaas, L. A. I. and Steger-Hartmann, T. (2024) The road to virtual control groups and the importance of proper body weight selection, ALTEX - Alternatives to animal experimentation, 41(4), pp. 660–665. doi: 10.14573/altex.2403141.
2. Li, D., Garren, J., Mangipudy, R., Martin, M., Tomlinson, L., Vansell, N.R. (2024) Statistical Applications of Virtual Control Groups to Nonrodent Animal Toxicity Studies: an Initial Evaluation, Regulatory Toxicology and Pharmacology, 2024, 105733, ISSN 0273-2300, https://doi.org/10.1016/j.yrtph.2024.105733.