Standard Operating Procedure for the Implementation of Virtual Control Groups in Preclinical Studies

Authors: this SOP was jointly developed by experts from the following members of the VICT3R consortium: Bayer, Merck, Roche, Sanofi, UPF

Version: 1.0

Table of Content

1. Introduction....1
2. Scope....2
3. Responsibilities....2
4. Definitions....2
5. Requirements for selecting HCD to be used for VCG generation....3
   5.1. Data Collection....3
   5.2. Data Standard and Curation....3
6. Procedure for Generation of VCGs....4
   6.1. List of matching criteria....4
   6.2. Potential modifications of the matching criteria....5
   6.3. Required animal number in VCGs....5
   6.4. Matching based on pre-values for larger species....5
   6.5. Assessing the relevance of the matching criteria....6
   6.6. Documentation and Reporting of VCGs....7
7. Conclusions....7
8. References....7

1. Introduction

This Standard Operating Procedure (SOP) describes the recommended processes for generating and applying Virtual Control Groups (VCGs) in regulatory toxicity studies. Historical control data (HCD) from animals used in legacy studies are selected by study-specific criteria, termed “matching criteria”, to replace or augment concurrent control groups (CCGs) with VCGs. The replacement of CCG animals may lead to a reduction of animal use while maintaining scientific rigor and integrity (4), ((17).

The SOP details the processes of HCD collection, data curation, selection based on study-specific matching criteria, and study evaluation. The SOP was developed by members of the VICT3R consortium and reflects the current knowledge and experience in the application of VCGs. It will likely evolve as the VICT3R consortium gathers further evidence on how to construct and implement VCGs.

Current European and international regulations require animal studies for the nonclinical safety assessment prior to conducting clinical trials and market authorization for pharmaceuticals (8). The conventional setting of a regulatory toxicology study uses 25 % of the animals as controls (e.g. OECD TG 407 “Repeated Dose 28-Day Oral Toxicity Study in Rodents”, (12)). The relevant guidelines for chronic studies or carcinogenicity studies follow a similar scheme (11)),(13), with

the exception that the animal numbers per group increase with the duration of the study. Applying this SOP to other or less conventional study designs may require additional consideration not addressed here.

The use of HCD from previous studies is advised in some regulatory documents mainly for the purpose of performance control of the study and the assessment of outliers (12). The use of HCD to create VCGs, also called synthetic control arms in the clinical context, is an established procedure for randomized clinical trials and their use is described in the International Council of Harmonisation (ICH) guideline E10 (9). However, there are currently no guidelines for a similar approach in preclinical animal studies. A recent reflection paper limited to single-arm clinical trials further outlines considerations for non-randomized designs lacking a control arm (2).

2. Scope

This SOP has been developed by members of the VICT3R consortium. Since the VICT3R consortium does not have and cannot acquire the status of a Test Facility according to GLP regulations, this SOP needs to be adapted by each Test Facility to fulfill the requirements of internal quality assurance frameworks. Therefore, this SOP should be considered as a master document.

The SOP applies to all animal studies performed under GLP and is applicable to all personnel involved in the design and execution of regulatory toxicity studies utilizing VCGs within the organization. Deviations of the described processes should be documented according to the requirements set forth in the individual Test Facilities.

The processes described in this SOP can also be applied to non-GLP animal studies, such as dose-range finding studies. However, due to the less stringent design requirements for these non-GLP studies, certain steps of the VCG generation may be disregarded or adapted.

3. Responsibilities

It is envisioned that at a minimum, the following personnel will be involved:

Study Director: Overall responsibility for the implementation of VCGs in animal studies.

Data Manager: Responsible for data collection, curation, and maintenance of the historical control database.

Statistical Analyst: Responsible for statistical evaluation and analysis of the VCG data.

Subject Matter Experts: Responsible for the evaluation of specific endpoints, like clinical pathology or histopathological data, as well as for integration of data and conclusion into the overall study analysis.

4. Definitions

Concurrent Control Groups (CCGs): a control group of animals that are both physically present and have raw data collected during the study, i.e. the group of animals used as control in a specific study.

Virtual Control Groups (VCGs): Control groups of animals generated from selected HCD that serve to partly or entirely replace or supplement and be a representative sample of CCGs in animal toxicity studies for comparisons to the treated groups.

Historical Control Data (HCD): Data collected from legacy studies that can be utilized to construct VCGs.

5. Requirements for selecting HCD to be used for VCG generation

5.1. Data Collection

For generating VCGs, HCD usually stored in repositories of Test Facilities should be captured using the CDISC Standard for Exchange of Nonclinical Data (SEND) data structure format (16) and adhering to the corresponding Controlled Terminology or other reference terminologies such as NCI Thesaurus (10), which can be mapped to SEND. This data encompasses endpoints and parameters recorded in accordance with OECD guidelines for the conduct of animal studies (11)(12), (13).

For regulatory compliance, unique identifiers for each animal are mandatory to ensure complete traceability back to the original study. These identifiers are preserved throughout the data processing workflow, allowing regulators to verify the source and integrity of all data points included in the VCG.

Essential metadata that must be collected includes (corresponding SEND term is provided in brackets):

- Unique animal identifier linking to the original study (UUID)
- Species (SPECIES)
- Strain (STRAIN)
- Animal supplier/breeder (SPLRNAM)
- Route of administration (ROUTE)
- Dosing duration (DOSDUR)
- Study day (BWDY/LBDY)
- Body weight (BWORRES) usually used as a surrogate for age for rodents
- Initial age (AGE) for larger species
- Treatment vehicle (TRTV)
- Treatment schedule
- Study start-year (STDSTDTC)
- Test facility location (TSTFLOC)
- Time from beginning of treatment until sacrifice (TRMSAC)

5.2. Data Standard and Curation

As described above, the expected data format is SEND. If there are no controlled terminologies for specific parameters or terms (e.g. harmonized description of vehicle composition) the study director or data manager may use in-house terminologies which need to be documented and archived according to GLP requirements.

If collected parameters are not consistently formatted or measured using the same units across studies, curation steps will be required to harmonize the data. In such cases, it is mandatory to retain the original data entries alongside the curated data while also documenting the adopted curation steps.

All collected data needs to be traceable to the original study, and protocols need to be in place to assure data integrity through the process of HCD data collection.

6. Procedure for Generation of VCGs

6.1. List of matching criteria

To ensure comparability of the VCG animals and those in the treatment groups in a study, the parameters set forth in the study protocol must serve as the selection criteria for the HCD from which VCGs were built.

These include the following criteria:

a) Test facility
b) Species & strain & Sex
c) Supplier/breeder
d) Origin (particularly relevant for NHPs)
e) Housing condition (differentiate between group or single housing)
f) Route of administration (differentiate between ORAL GAVAGE, ORAL via diet, INTRAVENOUS, INTRAMUSCULAR, SUBCUTANEOUS or others)
g) Method of dosing (differentiate between bolus or continuous infusion)
h) Treatment schedule (differentiate between QD, BID or other) & dosing duration & recovery period duration
i) Type of vehicle
j) Year of study (within a given time window, e.g. the last 5 years).
k) For rodents: initial body weight or k). For large species (dog, NHPs, minipigs): age of animals and selected pre-values (for details, see 6.3)

This list of criteria represents a subset of criteria established by other authors (14) that may impact the variability of parameters measured in animal studies, chosen based on the biological plausibility of a criteria to elicit a high influence on variability. These criteria can be pre-specified during the planning phase of the study as soon as the final study protocol is available. One exception is the initial body weight for rodents (criterion k) that becomes available only at study start upon delivery of the animals to the test facility or during the acclimation phase. Therefore, initial body weight will be the last criterion applied to the HCD before constructing VCGs for rodents (6). When applying this criterion, the initial body weight of rodents in the selected HCD should reflect the initial body weight range of the animals assigned to a new study. Statistical tests or visual inspection should be performed to ensure that the initial body weight does not differ substantially from the initial weights of the treatment group animals (7). At a minimum, the initial body weight of rodents in the selected HCD must fall within the minimum and maximum initial body weight of treatment group animals.

For large animals (dogs, NHPs, mini pigs) the age of the animals is to be used, if possible, as this criterion shows a better correlation with clinical pathology parameters than initial body weight. In this case, animals in the selected HCD must be at least as old as the animals in the treatment groups or of an age that matches a window pre-specified around the treated groups (i.e., treated group age +/- 25%).

In addition to the above-described criteria, study-specific parameters measured during the acclimation phase particularly for larger species can serve as additional matching criteria or as

criterion for deselecting animals due to extreme pre-values, especially if the pre-values for animals in the selected HCD show a different distribution compared to treatment groups animals (see 6.4) and specified in the protocol.

6.2. Potential modifications of the matching criteria

The use of the matching criteria listed above as exclusion/inclusion criteria for HCD selection is recommended. Particularly, it is advisable to keep parameters a-d identical between the treatment group animals and the VCG animals.

In some cases, one may relax the other criteria to satisfy the required number of animals (see section 6.3). In this case, additional analysis or scientific rationale should be provided to ensure that the selected HCD animals for constructing VCGs come from a similar population of animals as those used in the treatment group. This can be done, for example, by comparative statistical analysis of pre-values (for details, see 6.4) and may particularly apply to criteria e), f), g), h) and i).

For criteria h) to j), it is recommended to analyze time-control charts or distribution plots of parameters as described by Gurjanov et al. (5).

6.3. Required animal number in VCGs

To simulate the random allocation process of animals into treatment and control groups that is commonly performed in the preparatory phase of a new study, the number of animals in the selected HCD used to construct VCGs must be at least equal to the desired control group size.

The desired control group is of the same size as the expected number of animals in each treatment group if not otherwise stated in the study protocol, while a larger control group may be desirable to increase in certain cases the statistical power of analysis.

If the pool of animals in the matched HCD is higher than the desired control group size, animals for the VCG should be sampled randomly from this pool. However, in case the number of animals in the selected HCD for constructing VCGs is smaller than the desired control group size, the study director may relax the matching criteria to expand the selected HCD pool size (see section 6.2) and document the criteria and scientific rationale in the protocol. Examples may include QD and BID studies, studies with similar vehicles, or extending the time windows beyond 5 years in the past.

If even after relaxing the matching criteria after thorough investigation, the number of animals in the selected HCD used to construct VCGs is still smaller than the desired control group size, the CCG cannot be entirely replaced by VCGs. In such a case, the study director may consider reducing the number of animals in the CCG by replacing the reduced CCG animals with eligible VCG animals resulting in a so-called hybrid design (4).

6.4. Matching based on pre-values for larger species

Pre-values, also called baseline values or pretreatment values, are measurements or observations taken prior to the first administration of the test item, usually also before the allocation of animals into treatment and control groups. Recommendations on how to perform these measurements are not explicitly mentioned in the OECD guideline for dog repeated dose toxicity studies (11) but are mentioned in other guidelines (1), (3). However, a full list of

parameters measured is not set forth in these guidelines and no explicit recommendation on how to use the pre-values before, during and after the conduct of a study is given.

A common use of pre-values is to single-out diseased animals in combination with clinical observations prior to treatments. Animals with pre-values out of reference ranges for clinical pathology parameters are also identified, as these may indicate infections (e.g. lymphocyte count) or preexisting liver damage (e.g. liver transaminases). In addition, pre-values for specific parameters, based on a scientific understanding of the hypothesized or previous studies regarding the mechanism of action, may be used to assist study directors assigning animals to treatment and control groups, distributing animals homogenously. Not all pre-values are of equal importance for assigning animals into groups, and mostly body weight, body temperature, liver enzymes, red blood cell and lymphocyte count play a role.

Therefore, to simulate the process by which pre-values affect CCG construction in animal studies when constructing VCGs, pre-values may be considered. For example, animals in the selected HCD with pre-values out of the range of pre-values of animals assigned to the treatment groups may be replaced in the VCG with better fitting animals in the pool. The decision to include or exclude these animals should be taken by the study director and documented in the final report. Furthermore, the similarity of pre-values may also be used as evidence when assessing whether certain matching criteria can be relaxed for HCD selection.

6.5. Assessing the relevance of the matching criteria

The animals selected for VCGs should relate to the treatment group population in the same way as the CCG. In particular, the CCG, treatment group and VCG animals should represent samples from the same animal population, for this purpose treatment group animal data before the start of the treatment should be used. Based on additional data collection and analyses, the VICT3R consortium will recommend in the future which matching criteria will have to be considered and which can be disregarded or grouped together.

It is advisable that the study director investigate the most relevant criteria for the study's conduct. For large animals, the following parameters and endpoints may be considered: study year, age and body weight of the animals, levels of several liver enzymes, hematocrit, and counts of erythrocytes and lymphocytes.

When age of the VCG animals differ significantly from the age of the treatment group animals, it is advisable to consider the following aspects for the assessment of similarity between VCGs and treatment animals:

- Ranges for selected endpoints should be similar in VCG and treated animals. The endpoints to be considered in rodents are limited to body weights since pre-values are usually not available for these species. For larger animals, pre-values should be used for such comparisons. Visualization of the data (e.g. distribution plots) is advisable in this context to facilitate data interpretation (5). However, at this stage, there are no stringent criteria for assessing similarity, but the ranges should show a good overlap, and the distributions should not be skewed.
- VCG data should fall into the same historic reference value range which will be used to assess the study.

Similar considerations as above may be used to assess whether grouping animals with differences in criteria c) to j), i.e. grouping different housing conditions, dosing regimens or vehicles.

6.6. Documentation and Reporting of VCGs

When reporting, VCGs in prospective studies must be handled identical to CCGs. The following steps need to be followed:

a. VCG animal data needs to be traceable back to the original study and animal number available in the GLP archives. If the HCD used to generate VCGs does not originate from the same Test Facility where the prospective study is executed, a framework needs to be in place to be able to trace it back to the original provider.
b. The procedure and specific details of the applied criteria for selecting the HCD animals used to construct VCGs should be described in the study plan and subsequently included in the final report. The report should also contain reasons for modifying matching criteria if applicable (see 6.2).
c. The statistical analyses performed for characterizing the VCGs (see 6.5) should be described and include an assessment of similarity of the selected HCD animals used for constructing VCGs to the animals of the treatment groups. This description must also include reasons for excluding VCG animals, if applicable.
d. Statistical methods employed for testing differences between VCG, and treatment groups should be equivalent to those that would be used for comparing CCG and treatment groups.
e. If statistical analyses are performed between VCGs and treatment groups, this should ideally be run in the validated laboratory integrated management system (LIMS) of the test facility, i.e. the VCGs must be uploaded to the LIMS. If this is technically not possible, the analyses must be performed in a validated environment with adequate documentation.
f. VCG data must be reported in Appendices I and II of the final report together with statistical results identical to the procedures established for CCGs documentation.
g. The report should contain a general statement on the validity of the generated VCGs.

7. Conclusions

This SOP provides a framework for generating and applying VCGs as well as guidance for reporting in animal studies. The SOP primarily focuses on systemic toxicity studies performed according to regulatory requirements.

Application to other study types is expected to follow the same principles and can be further refined as the area evolves.

8. References

(1) EMA (2010). Guideline on repeated dose toxicity. 18 March 2010 CPMP/SWP/1042/99 Rev 1 Corr* Committee for Human Medicinal Products (CHMP)
(2) EMA (2024). Reflection paper on establishing efficacy based on single-arm trials submitted as pivotal evidence in a marketing authorisation application: Considerations on evidence from single-arm trials. (EMA/CPMP/458061/2024). https://www.ema.europa.eu/en/documents/scientific-guideline/reflection-paper-

establishing-efficacy-based-single-arm-trials-submitted-pivotal-evidence-marketing-authorisation-application_en.pdf.

(3) EPA (1998). Health Effects Test Guidelines OPPTS 870.3150 90-Day Oral Toxicity in Nonrodents.

(4) Golden E, Allen D, Amberg A, Anger LT, Baker E, Baran SW, Bringezu F, Clark M, Duchateau-Nguyen G, Escher SE, Giri V, Grevot A, Hartung T, Li D, Lotfi L, Muster W, Snyder K, Wange R, Steger-Hartmann T. Toward implementing virtual control groups in nonclinical safety studies. ALTEX. 2024;41(2):282-301. doi: 10.14573/altex.2310041. Epub 2023 Dec 1. PMID: 38043132.

(5) Gurjanov A, Kreuchwig A, Steger-Hartmann T, Vaas LAI. Hurdles and signposts on the road to virtual control groups-A case study illustrating the influence of anesthesia protocols on electrolyte levels in rats. Front Pharmacol. 2023 Apr 20;14:1142534. doi: 10.3389/fphar.2023.

(6) Gurjanov, A., Vieira-Vieira, C., Vienenkoetter, J., Vaas, L.A.I., Steger-Hartmann, T. (2024). Replacing concurrent controls with virtual control groups in rat toxicity studies. Regulatory Toxicology and Pharmacology, 148, 105592.

(7) Gurjanov A, Vaas LAI, Steger-Hartmann T. The road to virtual control groups and the importance of proper body-weight selection. ALTEX. 2024;41(4):660-665. doi: 10.14573/altex.2403141. Epub 2024 May 22. PMID: 38809255.

(8) ICH M3 (2008). Guideline M3(R2) on non-clinical safety studies for the conduct of human clinical trials and marketing authorisation for pharmaceuticals. https://www.ema.europa.eu/en/ich-m3-r2-nonclinical-safety-studies-conduct-human-clinical-trials-pharmaceuticals-scientific.

(9) ICH E10 (2001). Topic E 10 Choice of Control Group in Clinical Trials. Note or Guidance on Choice of Control Group in Clinical Trials. (CPMP/ICH/364/96).

(10) National Cancer Institute (2025): Enterprise Vocabulary Services: https://evs.nci.nih.gov/evs-download/thesaurus-downloads (last accessed: May 27th, 2025).

(11) OECD (1998), Test No. 409: Repeated Dose 90-Day Oral Toxicity Study in Non-Rodents, OECD-Guidelines for the Testing of Chemicals, Section 4, OECD Publishing, Paris, https://doi.org/10.1787/9789264070721-en.

(12) OECD (2008), Test No. 407: Repeated Dose 28-day Oral Toxicity Study in Rodents, OECD Guidelines for the Testing of Chemicals, Section 4, OECD Publishing, Paris, https://doi.org/10.1787/9789264070684-en.

(13) OECD (2008b). Test No. 452: Chronic Toxicity Studies. OECD Guidelines for the Testing of Chemicals, Section 4. OECD Publishing, Paris. doi:10.1787/9789264071209-en.

(14) Palazzi, X., Anger, L.T., Boulineau, T., Grevot, A., Guffroy, M., et al. (2024). Points to consider regarding the use and implementation of virtual controls in nonclinical general toxicology studies. Regulatory Toxicology and Pharmacology, 150, 105632.

(15) Sato, G., Nakajima, M., Sakai, K., Togashi, Y., Yamamoto, M., et al. (2024). Potential issues associated with the introduction of virtual control groups into non-clinical toxicology studies. Translational and Regulatory Sciences, 6(1), 1-9.

(16) Standard for Exchange of Nonclinical Data (SEND) Implementation Guide. Version v3.1.1. https://www.cdisc.org/standards/foundational/send/sendig-v3-1-1

(17) Steger-Hartmann, T., Kreuchwig, A., Vaas, L., Wichard, J., Bringezu, F., Amberg, A., et al. (2020). Introducing the concept of virtual control groups into preclinical toxicology testing. ALTEX, 37(3), 343-349.