EUROPEAN MEDICINES AGENCY
SCIENCE MEDICINES HEALTH

1 31 March 2026
2 EMADOC-1700519818-3025927
3 Committee for Medicinal Products for Human Use (CHMP)

4 Draft Qualification opinion for Virtual Control Groups 5 (VCG) to replace Concurrent Control Groups (CCG) in rat 6 non-GLP Dose-Range Finding (DRF) studies

Draft agreed by Scientific Advice Working Party (SAWP)	15 January 2026
Adopted by CHMP for release for consultation	26 February 2026[1]
Start of public consultation	31 March 2026[2]
End of consultation (deadline for comments)	12 May 2026

7

Comments should be provided using this template. The completed comments form should be sent to ScientificAdvice@ema.europa.eu

8

Keywords	Virtual control groups, Historical control data, 3Rs, NAMs, non-GLP, repeated dose toxicity studies

9

[1] Last day of relevant committee meeting

[2] Date of publication on the EMA public website

Qualification Opinion agreed by CHMP

Based on the evidence presented by the Applicant and reviewed by the SAWP Qualification Team, CHMP considers that Virtual Control Groups (VCGs) can be used to substitute for Concurrent Control Groups (CCGs)) in non-clinical non-GLP rat dose range finding (DRF) studies to inform on dose selection for subsequent pivotal rat GLP-compliant repeated dose studies, when applied in accordance to the SOP (see Appendix 1)

Agreed Context of Use (CoU):

The Application of Virtual Control Groups (VCGs) according to the SOP (see Appendix 1) to substitute for Concurrent Control Groups in non-clinical rat non-GLP dose range finding (DRF) studies to inform on dose selection for subsequent pivotal rat GLP repeated dose studies.

EMA/CHMP qualification opinions or qualification advice of novel methodologies for medicinal product development are provided without prejudice to any requirements related to other applicable legislation (e.g. MDR/IVDR and AI Act).

Executive Summary

The Applicant [Synapse-Managers together with 5 pharmaceutical companies] requested a Qualification Opinion (QO) on the applicability of Virtual Control Groups (VCGs) in non-clinical rat dose range finding (DRF) studies, with the aim of reducing the number of control animals by partially or entirely replacing concurrent control groups (CCGs).

To this end, VCGs are derived from a historical control database based on the SENDIG standard, previously called the VICOG database[3], and currently known as the VICT3R database. This Qualification Opinion solely concerns the method of extracting VCGs to replace CCGs. The population and composition of the VICT3R database itself is out of scope of this Qualification Opinion.

The key requirement defined by the Applicant for implementation of the VCG approach is the assurance that the use of VCGs does not compromise study outcomes and, consequently, does not compromise human safety. In the context of DRF studies, the primary objective is to inform on dose selection for subsequent pivotal repeated dose studies, by establishing a No Observed Adverse Effect Level (NOAEL) and identifying preliminary adverse findings. To address this requirement, the Applicant performed reanalyses of legacy non-GLP rat DRF studies and GLP pivotal rat repeated dose toxicity studies, to evaluate whether replacement of CCGs with VCGs would alter overall study conclusions and whether the overall VCG approach can be considered robust and reliable.

VCGs were generated in accordance with a standard operating procedure (SOP), applying a statistical procedure in combination with expert judgement to characterize control data and identify the optimal match to the animals assigned to treatment groups. Matching criteria included overarching study information (facility, housing conditions, animal provider), general animal information (species, strain, sex), study conditions (route of administration, treatment schedule, dosing period, recovery, vehicle), time window and specific animal information (age, initial body weight). Initial body weight was used as matching criterion in all reanalyses.

To assess the replicability, four major categories were compared:

- Target organs of toxicity

[3] The prototype of the database, originally called VICOG, was embedded by IHI VICT3R in the VICT3R database.

- Measurable biomarkers of toxicity based on measured clinical pathology parameters
- Identification of threshold doses (e.g., NOAEL)
- Reversibility (where applicable, generally not assessed in DRF studies and only assessed in one pivotal GLP study)

Across all eight reanalysed rat studies, replication of target organ of toxicity, biomarkers and threshold dose was established. Replication of reversibility was also established, albeit in the one single study where this finding was relevant, and hence this finding is considered only supportive.

In some reanalyses, certain findings were not replicated, while in other cases the use of VCGs led to the identification of additional noteworthy findings. However, these observations did not result in a different NOAEL and would therefore not have influenced decision-making regarding dose selection for subsequent pivotal studies and hence do not impact CoU I.

In conclusion, replacing CCGs with VCGs in rat non-GLP DRF studies, when applying the procedure described in the SOP with predefined matching criteria, resulted in similar overall study conclusions and the same NOAEL as in the case of rat non-GLP DRF studies using the CCG and this in all cases evaluated.

Abbreviations:

CDISC-SEND	Clinical Data Interchange Standards Consortium – Standard for Exchange of Nonclinical Data
CCG	Concurrent Control Group
CoU	Context of Use
DILI	Drug Induced Liver Injury
DM	Demographics Domain
DRF	Dose Range Finding (study)
EMA	European Medicines Agency
GGT	Gamma-Glutamyl Transferase
GLP	Good Laboratory Practice
GLUC	Glucose
HCD	Historical Control Data
HD	High Dose
HNSTD	Highest non-severely toxic dose
ICH	International Council for Harmonization of Technical Requirements for Pharmaceuticals for Human Use
IHI	Innovative Health Initiative
IMI	Innovative Medicines Initiative
i.v.	intravenous(ly)
LD	Low Dose
LLN	Lower Limit of Normal
LOEL	Lowest Observed Effect Level
LOAEL	Lowest Observed Adverse Effect Level
LON	Limit(s) of Normal
LYM	Lymphocyte Count
MAA	Marketing Authorization Application
MCV	Mean Corpuscular Volume
MD	Mid Dose
N/A	not applicable
NHP	Non-Human Primate
NOEL	No Observed Effect Level
NOAEL	No Observed Adverse Effect Level

98	QD	quaque die (once a day)
99	QO	Qualification Opinion
100	QT	Qualification Team
101	RETI	Reticulocyte Count
102	RBC	Red Blood Cell Count
103	RDW	Red Cell Distribution Width
104	SAWP	Scientific Advice Working Party
105	SME	Subject Matter Expert
106	STD10	Severely Toxic Dose in 10% of animals
107	SOP	Standard Operating Procedure
108	TS	Trial Summary
109	VCG	Virtual Control Group
110	WBC	White Blood Count

Qualification Opinion – Overview of the Assessment

Overview of the Qualification Procedure:

In February 2024, the Applicant requested Scientific Advice from EMA for the Context of Use I (the applicability of VCGs in nonclinical dose range finding (DRF) rat studies to reduce the number of control animals by partially or entirely replacing concurrent control groups) for those non-clinical species most frequently used (rats, mice, dogs, NHPs, and mini-pigs).

The request resulted in the submission of Briefing Documents and included replies to the List of Issues formulated by the SAWP Qualification team (SAWP QT).

During a discussion meeting held on November 25th, 2024, the SAWP QT was informed that all further interaction with EMA will run through the Innovative Health Initiative VICT3R (https://www.vict3r.eu/), which started September 1st, 2024.

The final submission of the qualification documents occurred on May 30th, 2025.

Oct 1st, 2025, SAWP QT informed VICT3R about the decision, that "an amended Context of Use I (VCG application in non-GLP DRF studies) could be qualified, however limited to studies in rats". To this end, the SAWP QT requested that a Summary Qualification Document would be compiled comprising the following information:

- Briefing Document 1 submitted April 3rd, 2024
- Replies to list of issues (LoI 1) received September 18th, 2024, submitted November 13th, 2024
- Briefing Document 2 submitted May 30th, 2025

While the above-mentioned documents also included information and data for other non-clinical species (3 mice studies, 9 dog studies, 6 NHPs studies, 2 mini-pig studies) this material was not included in the Summary Qualification Document, which focuses on the rat data, and hence the data obtained in other species are not part of the current Qualification Opinion.

This Summary Qualification Document contains a detailed description of the matching procedure for the generation of VCGs summarized in a Standard Operating Procedure (SOP).

Main issues discussed during the Qualification Procedure:

The main issues discussed by the SAWP QT and pertinent to the agreed CoU are the following:

- Upon SAWP QT request, a SOP has been submitted by the Applicant, that acts as a master document which will be adapted as the VICT3R consortium continues its evidence building exercise. This master SOP is adapted by every test facility to fulfil the requirements of internal quality assurance frameworks. Deviations from the described processes are possible and are recommended to be documented according to the requirements in the individual Test Facilities (see Appendix 1).
- Concerning the matching criteria, the master SOP lists "advisable" criteria that should be identical between treatment group animals and VCG animals (i.e. test facility, species & strain & sex, supplier/breeder, origin) and, in addition, mentions that the other listed matching criteria can be relaxed, provided this is justified, to satisfy the required number of animals. The SAWP QT carefully evaluated the impact of this flexible application of the SOP on the outcome of the reanalysed studies (see non-clinical validation of the methodology). The data supporting the application of the VCG approach for rat non-GLP DRF studies (i.e. CoU I) followed the SOP with

only minimal relaxation of the listed matching criteria (see non-clinical validation of the methodology) and the SAWP QT concludes that in the context of dose setting for pivotal rat GLP repeated dose toxicity studies, this was acceptable.

- The SOP provides the option for the use of in-house historical control data (HCD) for the population of the VCGs, provided use of CDISC SEND data structure format and adherence to the corresponding Controlled Terminology or other reference terminologies which can be mapped to SEND is confirmed. The SAWP QT carefully evaluated the impact on the outcome of the reanalysed studies (see non-clinical validation of the methodology) and concluded that in the context of dose setting for pivotal rat GLP repeated dose toxicity studies, the application of the VCG approach for rat non-GLP DRF studies (i.e. CoU I) this was acceptable.
- To substantiate the request for qualification opinion for COU I, the Applicant supplemented the initially provided dataset comprising results for 5 reanalysed rat studies with results from 3 additional reanalysed studies. The key requirement for the implementation of the proposed concept is the assurance that using VCGs does not compromise study outcomes and thus does not endanger human safety. Therefore, each individual study reanalysis aimed to determine whether the replacement of CCGs with VCGs would alter the study overall conclusions (i.e. the determination of a NOAEL or highest non-severely toxic dose (HNSTD)) and if the performance of the proposed VCG approaches can be considered robust and reliable. The SAWP QT concludes that the submitted reanalysed rat studies confirm that the VCG approach allows for the identification of target organs and NOAEL/HNSTD determination, thereby providing an adequate basis to guide dose setting for subsequent rat GLP repeated dose toxicity studies.

In conclusion, while there are no regulatory barriers to the use of VCG in DRF studies (and such use is indeed encouraged in general), based on the evidence presented in the qualification opinion request and in the discussion meetings, CHMP considers that Virtual Control Groups (VCGs) can be used to substitute for Concurrent Control Groups (CCGs) in non-clinical non-GLP rat dose range finding (DRF) studies to inform on dose selection for subsequent pivotal rat GLP-compliant repeated dose toxicity studies, provided that the SOP is followed (see Appendix 1).

Rationale for the Virtual Control Group Approach

Introduction

Current European and international regulations require animal studies for the non-clinical safety assessment prior to the conduct of clinical trials and market authorisation for pharmaceuticals (ICH M3, 2008). The conventional setting of a regulatory toxicology study uses 25% of the animals as controls: OECD TG 407 "Repeated Dose 28-Day Oral Toxicity Study in Rodents" (OECD, 2008a) may be cited here as an example for subacute studies, though the pertinent guidelines for chronic studies or carcinogenicity studies essentially follow the same scheme (OECD, 2008b), with the exception that the animal numbers per group increase with the duration of the study.

The use of historical control data collected from previous studies is advised in some regulatory documents (European Commission, 2013) mainly for the purpose of performance control of the study and the assessment of outliers. The replacement of concurrent controls with historical data is currently not foreseen for animal studies. However, virtual controls, also called synthetic control arms, are a well-established procedure for randomized clinical trials (Berry et al. 2017), particularly in the field of

oncology (Kadakia KT. et al. 2021). This approach is also described in the International Council of Harmonisation (ICH) guideline E10 (ICH E10, 2001).

Compared to the setting in clinical trials, the conditions in animal toxicity studies are considerably more controlled regarding homogeneity of applied bred animals and experimental conditions. All assessed parameters are usually under quality control using standard operation procedures and subject to Good Laboratory Practice (GLP). It is therefore considered, provided additional adequate quality control procedures are implemented during data collection, curation and selection of historical control (HCD) data for virtual control groups (VCGs), concurrent control groups (CCGs) could be reduced in size or entirely replaced by VCGs. By these means VCGs could significantly contribute to 3Rs. In addition, the VCG concept could contribute to overcoming the well described issue of statistically underpowered animal studies (Kluxen et al. 2022).

The VCG approach

Within the framework of the IMI consortium eTRANSAFE (https://etransafe.eu/), the Applicant has developed the concept of VCG (Steger-Hartmann et al., 2020). It describes the generation of VCGs from well-curated HCD after applying study-specific criteria, termed matching criteria.

The application of VCGs is intended to partially reduce the number of CCG animals or entirely replace CCGs used in non-clinical repeated dose toxicity studies, hereby reducing the animal numbers in non-clinical safety testing. A concise graphical explanation of the concept is provided in Figure 1.

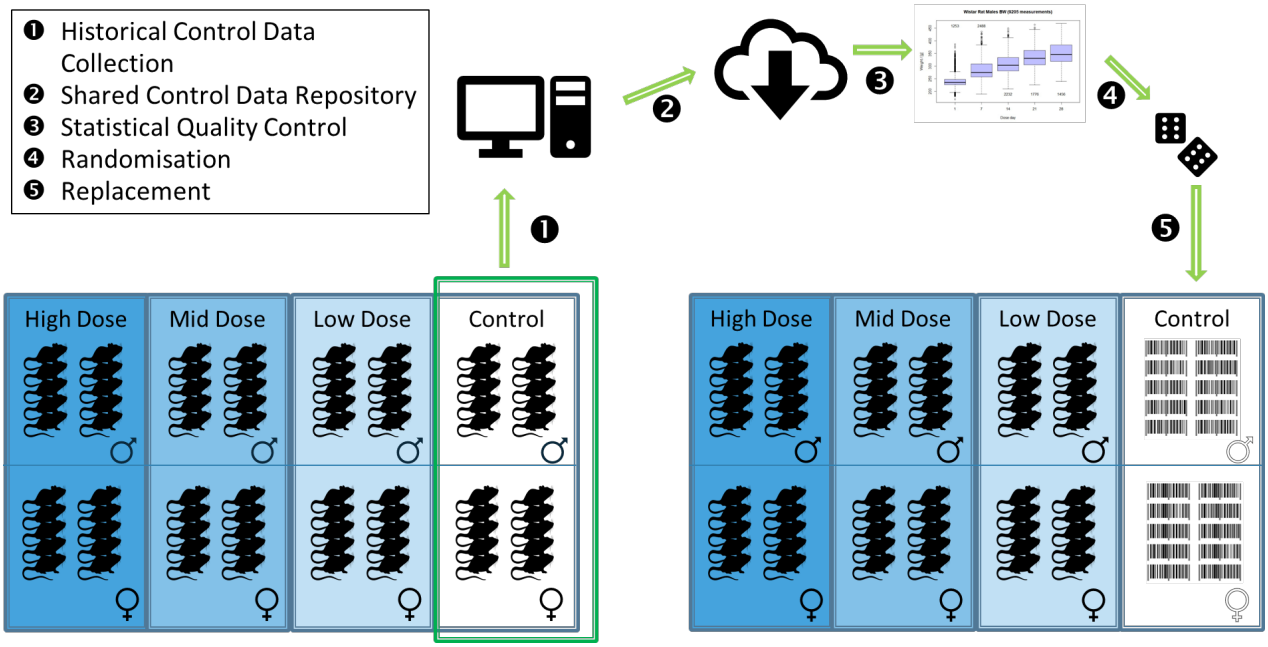


Figure 1: Graphical illustration of the VCG concept

In 2023, several pharmaceutical companies participating in IMI2 eTRANSAFE started to share data of control groups from legacy non-clinical repeated dose toxicity studies for the most common animal species used in toxicity testing and initiated the development of a prototype database for historical control data. Statistical procedures for characterizing the control data and identifying the optimal match with the animals assigned to treatment groups were developed (Sanz et al. 2023). The database, the statistical procedures and the matching methods were used for a qualification process, where several legacy repeated dose toxicity studies were reanalysed upon replacement of CCGs with VCGs.

The key requirement for the implementation of the proposed concept is the assurance that using VCGs does not compromise study outcomes and thus does not pose a threat to human safety in later clinical trials. Therefore, each individual study reanalysis aimed to determine whether the replacement of CCGs with VCGs would alter the study's overall conclusions and whether the performance of the proposed VCG approaches can be considered robust and reliable.

The Applicant has identified three steps for the broad implementation of the VCG concept, which are assigned to three contexts of use:

1. Application in dose-range finding (DRF) non-GLP repeated dose toxicity studies (CoU I)
2. Partial replacement of concurrent controls in pivotal (GLP) repeated dose toxicity studies (CoU II)
3. Full replacement of concurrent controls in pivotal (GLP) repeated dose toxicity studies (CoU III)

The current Qualification Opinion pertains only to CoU I: Application in dose-range finding (DRF) non-GLP repeated dose toxicity studies and only for studies conducted with rats.

Technical Considerations

The concept of applying a VCG relies on two distinct tools:

1) the historical control database (VICT3R database) from which historical controls data (HCD) are selected to populate the virtual control groups (VCGs), *and*
2) the standard operating procedure (SOP) which describes the method of selecting these HCD to form a study matched VCG.

The current qualification opinion specifically encompasses a SOP describing how to construct the VCG based on matching criteria and analyses to determine the predictive value of study outcomes when using VCGs, as compared with the original study CCGs.

Please note that an assessment of the historical control database (-VICT3R database) is not part of the qualification opinion, but for information a short description of this database is provided below.

<u>VICT3R database</u>

The Applicant has gathered and organised a significant volume of HCD from legacy repeated dose toxicity studies conducted with the most frequently used non-clinical species (rats, mice, dogs, minipigs and monkeys). For the current qualification opinion, only rat data are considered. The legacy data encompass all the endpoints and parameters recorded during repeated dose toxicity studies in accordance with OECD TG 407 guidelines (OECD 2008). The data were extracted from local repositories or LIMS systems of participating sponsors, carefully curated, and standardized using the CDISC SEND terminology (Clinical Data Interchange Standards Consortium – Standard for Exchange of Nonclinical Data). Subsequently, the datasets were uploaded to a newly created database: the VICT3R database. All donated data sets can be traced back to the original test facility and the study on an individual animal basis.

The data provided by the different sponsors were subjected to data curation workflows to achieve standardisation of numeric values, application of controlled terminology and data enrichment of the Demographics Domain (DM) using trial summary (TS) information from the TS Domain. The normalization of numeric values included the transformation to standard units whenever possible. For this process all data were imported into a separate database (the curation database) and after finalisation the result was transferred into the VICT3R database to result in standardised domains for

bodyweight (BW), organ weight (OM), laboratory (LB) containing the clinical chemistry, haematology, and urinalysis data and the microscopic (MI) domain.

Besides curation, a thorough characterisation of the data collected in the database for VCG selection was performed. The exploratory data analysis allowed the identification of the main sources of the endpoint variations and their corresponding magnitude. Additionally, it uncovered any hidden confounding factors that may be present. The insights gained from this exploration informed the selection of the relevant factors to consider when building VCGs.

Construction of the VCG on the basis of HCD – general considerations

The first step in the compilation of a suitable HCD pool is the identification of the animal subpopulation to be selected from the HCD. The data should be collected and curated according to the matching criteria described in the developed SOP (see Appendix 1).

It should be noted that the VICT3R database will require continuous updates with HCD to account for trends or shifts in the various experimental parameters (e.g., genetic drift or to include new vehicles, new dosing schedules, etc) (outside of the scope of the QO).

For the repeated dose toxicity studies reanalysed to support this qualification opinion, a time window of 5 years was used. However, statistical analysis may support the use of HCD outside of this time window.

The requirement for continuous updates of the VICT3R database implies that CCGs will not be replaceable in all future non-clinical repeated dose toxicity studies. The decision to use CCGs instead of VCGs will largely depend on the prior analysis of the HCD pool currently available and whether it offers enough virtual control animals matching the animals in treatment groups for any given prospective study.

Construction of the VCG on the basis of HCD – Standard Operating Procedure (SOP) (see Appendix 1)

An SOP for the generation of VCGs was developed (see Appendix 1). This SOP summarises the current knowledge and experience accumulated by the Applicant regarding HCD collection and the generation of VCGs, and instructions for documenting the procedures.

The SOP provides a list of matching criteria to identify the relevant subpopulation of control animals from legacy non-clinical toxicity studies and to ensure that possible confounding factors are accounted for (such as different testing facilities or different vehicles).

It is proposed that in cases of uncertainty regarding the matching between VCGs and the treatment animals, statistical analysis and expert judgement when statistical analysis is not feasible should be used to demonstrate that the selected VCG animals show similar distributions of key parameters (e.g. initial body weight) compared to the animals in the treatment groups. In some scenarios, it may be acceptable, provided adequately justified, to relax some matching criteria, to obtain a larger HCD pool from which VCGs can be selected, if there is statistical evidence that ranges and distributions of endpoints remain similar (see chapter 6.5 of the SOP, see Appendix 1).

If VCGs cannot be created due to failed matching or insufficient numbers of matching legacy control animals, a CCG would need to be used. This may occur if an important characteristic of the planned study (e.g., a specific type of vehicle, a new biomarker or a previously not determined endpoint) was not present in the control animals of legacy non-clinical repeated dose toxicity studies. Since the search for adequate VCG will occur before the study starts, it will still be possible to include CCGs before the start of the in-life phase of a study if the matching procedure fails.

SOPs are generally implemented in GLP settings. DRF studies, which are in scope for the CoU I of this qualification opinion, are usually non-GLP studies. The SOP created here is nevertheless considered as an appropriate guidance for these DRF studies, while considering the necessary flexibility in the design of these non-GLP studies.

Considerations on the selection of legacy non-clinical repeated dose toxicity studies for reanalysis to support the QO

The current procedure intends to qualify the application of VCG in rat DRF studies (see CoU I). Despite this focus on DRF studies, several GLP compliant 4-week rat toxicity studies were included in the reanalysed data set.

The reason for the inclusion of these studies is summarised here:

- **Data availability**. Compared to DRF studies, GLP 4-week repeated dose toxicity studies are performed in a much more standardized design, leading to larger and more homogeneous data sets, making the resulting data particularly valuable for an evaluation of the VCG concept. The design of DRF studies frequently varies in terms of the assessed organs or tissues, and the study duration, which limits the amount of matching data available for assessing a high fraction of DRF studies. This limitation will be overcome with the expansion of the VICT3R database and further statistical analyses.
- **Extrapolation of statistical conclusions**. Statistical analyses of the impact of matching criteria may shed light on the acceptable heterogeneity in study designs. For example, a thorough assessment of the effect of vehicles on study endpoints will inform about which vehicles can be grouped together without losing reliability during VCG application. Analogous analyses could be performed for other matching criteria, such as study duration. It should be noted that if the heterogeneity of a DRF study or any other study cannot be adequately addressed, the creation of a VCG will not be possible for such a study.
- **Higher comprehensiveness**. Compared to DRF studies, a broader spectrum of parameters is assessed in GLP 4-week repeated dose toxicity studies. If the evaluation of reanalysed GLP 4-week repeated dose toxicity studies confirms that original study results are not compromised by replacing the CCGs with VCGs, this conclusion could be extended to DRF studies (provided adequately matched data sets are available).

Considerations on the reanalysis of legacy DRF repeated dose toxicity studies after substitution of a CCG by a VCG

Regarding the reanalysis of DRF studies, it must be considered that the design and evaluation procedure of such studies is rarely described in guidelines and is subject to a high degree of variability. For example, DRF studies may differ in the number of animals per experimental group. In consequence, the reanalysis of these studies using VCGs will require a reflection of this heterogeneity. Therefore, individual approaches were used that best reflect the procedures implemented at different test sites.

For the submitted reanalyzed rat studies both statistical analysis and visual inspection were performed. The biological and toxicological evaluations of the results took precedence over the results of the statistical testing, aligning with the standard practices in the field (OECD 2002).

Beyond reproducing the statistical analysis of results, the strategy for qualifying the use of VCGs takes into consideration the process by which a study director interprets the results.

In this context a mere statistical approach to compare the results of the original results with those obtained with VCGs is not considered to be sufficient (Steger-Hartmann and Clark 2023). Focusing only on the comparison of those results which match or fail to match in terms of observed statistically significant changes will not provide a complete insight into the usability of VCGs. Differences in significant results would also be observed during a repetition of the study under identical conditions and are thus also foreseeable for VCGs (Steger-Hartmann et al. 2025). Besides experimental variation, this is also due inherent insufficient statistical power to assess the null hypothesis for the multitude of parameters measured in toxicity studies (Kluxen et al. 2021). In addition, qualitative parameters such as gross pathology and histopathology are not directly amenable to a statistical evaluation.

Therefore, the question, whether study results are not compromised while using VCGs must be addressed differently.

The main purposes of an animal toxicity study are the identification of the dose of the test item which causes adverse effects, the identification of the affected target organs, whether the observed toxicities can be monitored by biomarkers (clinical pathology) and depending on the study design, whether the observed adverse effects disappear or remain after stopping the administration of the test item (recovery phase). In this context, a difference in the statistical outcome between CCG and VCG might be of minor relevance or even irrelevant, if the affected parameter is not biologically related to the observed adverse effect.

Consequently, the results of the legacy study where the CCG was replaced with VCG were presented to a subject matter expert (SME), who was not the study director of the original study and the SME compared their judgment with the original outcome from the study director. Thereafter, noteworthy and treatment-related findings were presented to the study director of the original legacy study, who then evaluated whether missing or novel findings identified with the VCG reanalysis would have influenced the overall study outcome. Specifically, this assessment of replicability considered the effect of the VCG replacement on the threshold doses determined in the study (e.g. NOAEL, LOAEL), the identified target organs, and the correlation with clinical pathology parameters indicating monitorability of the target organ toxicity. This approach acknowledges prior knowledge a study director usually has on the test item, particularly regarding the pharmacological mode of action. This knowledge is essential to differentiate between excessive pharmacological effects and off-target toxicity. If this information is not considered during a blinded re-read it would significantly influence the evaluation of a study.

Method Development and Workflow

Introduction

Reanalysis of rat legacy (DRF) repeated dose toxicity studies was performed by replacing CCGs with the same number of VCG animals collected. For this, a mixed approach was designed by applying study specific matching criteria, considering the following steps.

1. The dose groups and the VCG were analysed using the same statistical procedures as those originally applied to the CCG data for all quantitative parameters.
2. Numeric results of the legacy study were compared with the ones from the VCGs by identifying parameters that showed identical behaviour in terms of statistical significance and direction of the change (increase or decrease).

3. Results were subsequently provided to the study director or another experienced toxicologist (i.e., a subject matter expert; SME) for assessment of treatment relatedness of findings. To document these decisions, pull-down fields were provided with arguments against "treatment-relatedness". For example, the expert was asked to indicate whether a finding had "no dose dependency", was "within 2-sigma range of historical control data" or had "different deviation in different sexes".
4. Subsequently, the study director or SME reviewed the novel and missed treatment-related findings after replacing CCGs with VCGs and assessed the adversity regarding dose and target organs.

For semi-quantitative or qualitative parameters that are not subjected to statistical analyses during routine studies (clinical observations, ophthalmology, gross pathology, histopathology), HCD collected for the same species, strain and test facility over the preceding five years were used to calculate background incidences for a specific finding. Alternatively, for analysis involving Sprague Dawley rats, incidence data from rat studies performed between 2000 and 2021 were included.

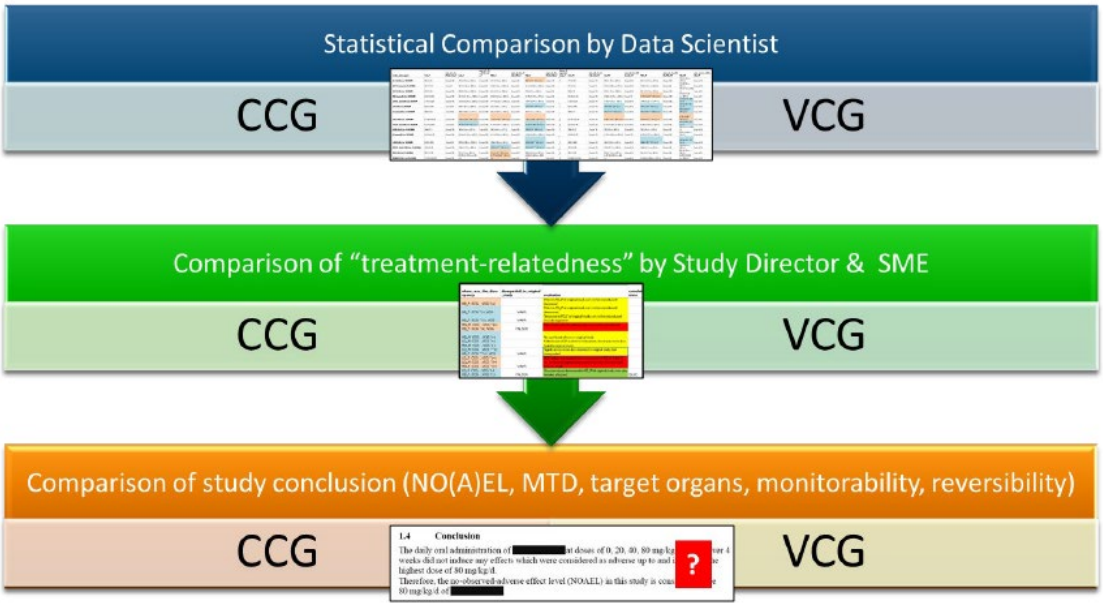


Figure 2: Overview of the qualification procedure based on assessment of legacy repeated dose toxicity studies in which CCGs were replaced by VCGs. *In the first step, the statistical outcome of the two studies is compared, and the parameters showing different behaviour (additional statistical significances detected or lost) are marked in the table (heatmap). Subsequently, these statistical differences are evaluated by a study director or SME regarding their treatment-relatedness. In addition, the study director or SME assesses whether an additional or a lost treatment-related finding has an influence on the overall study outcome, including determination of the NO(A)EL (or other dose thresholds), target organs, and other key findings.*

<u>Procedures for comparing study results obtained using CCG and VCG</u>

The general approach for the reanalysis of a study following the replacement of the CCG with a VCG has been briefly described above and is based on the publications Golden *et al.* (2024) and Gurjanov *et al.* (2024a). The workflow of the overall procedure is depicted in Figure 3 (from Gurjanov *et al.*, 2024a).

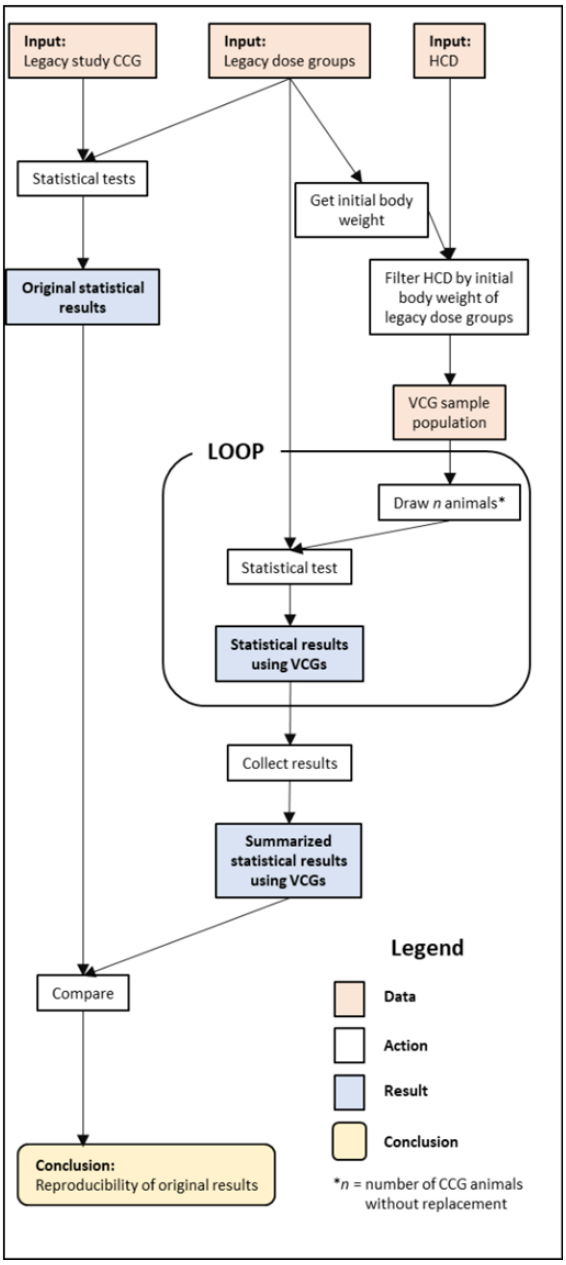
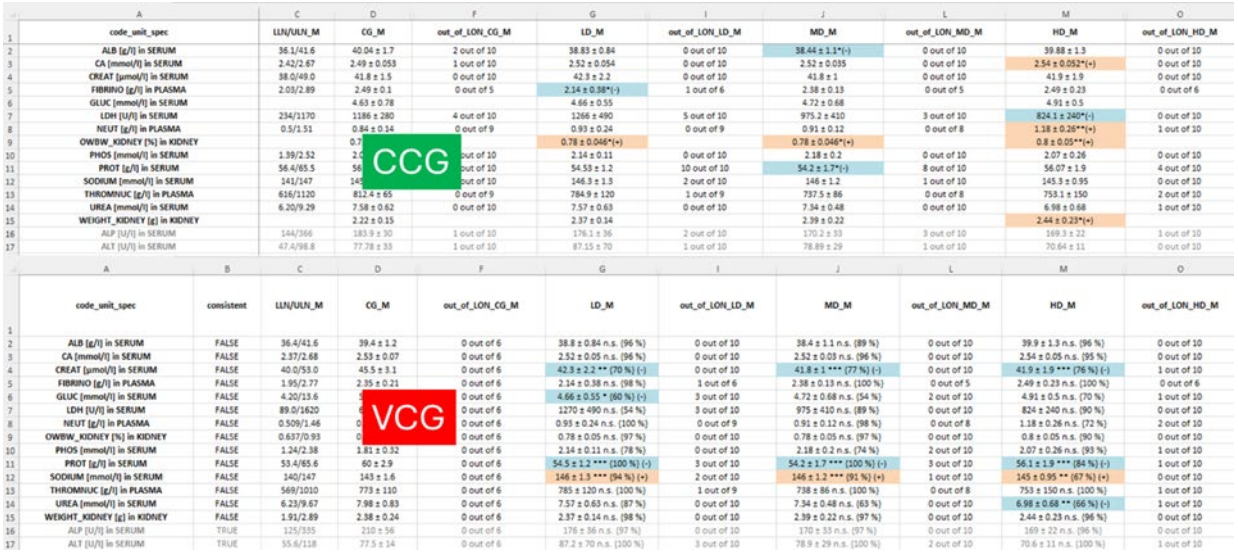


Figure 3: Visual description for the comparison of statistical results using CCGs and VCGs. *HCD was used as reference data for assessing test substance-relatedness based on limits of normal and for calculation of background incidences for qualitative findings (clinical observations, histopathology).*

After replacing the CCG with a VCG in the data set of legacy non-clinical repeated dose toxicity studies under reanalysis, treatment groups were statistically compared with the new control values from the VCG. For all studies except study H, an R-script was applied to further filter for control animals within the initial body weight range of the treatment groups animals, and to calculate statistically significant changes in treatment groups against the generated control group. Initial body weight filtering led to a final pool of animals from which VCGs were generated.

In each legacy study reanalysis (except for study H), a VCG is generated 100 times by sampling animals from the available pool and statistical analysis performed independently for each iteration. For each parameter measured, a statistical test (Dunnett test or U-test) was applied as per the original legacy study under reanalysis with VCGs, effectively mimicking the original process for data analysis. A quality assurance step is performed with the original CCG group by comparing statistical results produced with the R-scripts with the outcome in the original study report.

Identifying statistically significant differences between dose and control groups is insufficient to determine biological relevance. Therefore, statistical results from the reanalysis of the legacy studies after replacing CCG with VCGs were presented to an SME, who assessed test treatment-relatedness of findings. The goal of the SME when preparing the study assessment is to identify possible reasons to why a statistically significant difference between groups is not relevant and can therefore be disregarded. To facilitate and document these decisions, quantitative results were reported in an Excel table sorted in a way that parameters with mismatching results between the CCG and VCGs analysis were on the top and displayed in bold font (see Figure 4). The Excel table was then provided to SME to assess the treatment-relatedness of the findings.



code_unit_spec	LLN/ULN_M	CG_M	out_of_LON_CG_M	LD_M	out_of_LON_LD_M	MD_M	out_of_LON_MD_M	HD_M	out_of_LON_HD_M
ALB [g/l] in SERUM	36.1/41.6	40.04 ± 1.7	2 out of 10	38.83 ± 0.84	0 out of 10	38.44 ± 1.1*(-)	0 out of 10	39.88 ± 1.3	0 out of 10
CA [mmol/l] in SERUM	2.42/2.67	2.49 ± 0.053	1 out of 10	2.52 ± 0.054	0 out of 10	2.52 ± 0.035	0 out of 10	2.54 ± 0.052*(+)	0 out of 10
CREAT [µmol/l] in SERUM	38.0/49.0	41.8 ± 1.5	0 out of 10	42.3 ± 2.2	0 out of 10	41.8 ± 1	0 out of 10	41.9 ± 1.9	0 out of 10
FIBRINO [g/l] in PLASMA	2.03/2.89	2.49 ± 0.1	0 out of 5	2.14 ± 0.38*(-)	1 out of 6	2.38 ± 0.13	0 out of 5	2.49 ± 0.23	0 out of 6
GLUC [mmol/l] in SERUM		4.63 ± 0.78		4.66 ± 0.55		4.72 ± 0.68		4.91 ± 0.5	
LDH [U/l] in SERUM	234/1170	1186 ± 280	4 out of 10	1266 ± 490	5 out of 10	975.2 ± 410	3 out of 10	824.1 ± 240*(-)	0 out of 10
NEUT [g/l] in PLASMA	0.5/1.51	0.84 ± 0.14	0 out of 9	0.93 ± 0.24	0 out of 9	0.91 ± 0.12	0 out of 8	1.18 ± 0.26**(+)	1 out of 10
OWBW_KIDNEY [%] in KIDNEY		0.7		0.78 ± 0.046*(+)		0.78 ± 0.046*(+)		0.8 ± 0.05**(+)	
PHOS [mmol/l] in SERUM	1.39/2.52	2.0	out of 10	2.14 ± 0.11	0 out of 10	2.18 ± 0.2	0 out of 10	2.07 ± 0.26	0 out of 10
PROT [g/l] in SERUM	56.4/65.5	56	out of 10	54.53 ± 1.2	10 out of 10	54.2 ± 1.7*(-)	8 out of 10	56.07 ± 1.9	4 out of 10
SODIUM [mmol/l] in SERUM	141/147	145	out of 10	146.3 ± 1.3	2 out of 10	146 ± 1.2	1 out of 10	145.3 ± 0.95	0 out of 10
THROMNUC [g/l] in PLASMA	616/1120	812.4 ± 65	0 out of 9	784.9 ± 120	1 out of 9	737.5 ± 86	0 out of 8	753.1 ± 150	2 out of 10
UREA [mmol/l] in SERUM	6.20/9.29	7.58 ± 0.62	0 out of 10	7.57 ± 0.63	0 out of 10	7.34 ± 0.48	0 out of 10	6.98 ± 0.68	1 out of 10
WEIGHT_KIDNEY [g] in KIDNEY		2.22 ± 0.15		2.37 ± 0.14		2.39 ± 0.22		2.44 ± 0.23*(+)	
ALP [U/l] in SERUM	144/366	183.9 ± 30	1 out of 10	176.1 ± 36	2 out of 10	170.2 ± 33	3 out of 10	169.3 ± 22	1 out of 10
ALT [U/l] in SERUM	47.4/98.8	77.78 ± 33	1 out of 10	87.15 ± 70	1 out of 10	78.89 ± 29	1 out of 10	70.64 ± 11	0 out of 10

code_unit_spec	consistent	LLN/ULN_M	CG_M	out_of_LON_CG_M	LD_M	out_of_LON_LD_M	MD_M	out_of_LON_MD_M	HD_M	out_of_LON_HD_M
ALB [g/l] in SERUM	FALSE	36.4/41.6	39.4 ± 1.2	0 out of 6	38.8 ± 0.84 n.s. (96 %)	0 out of 10	38.4 ± 1.1 n.s. (89 %)	0 out of 10	39.9 ± 1.3 n.s. (96 %)	0 out of 10
CA [mmol/l] in SERUM	FALSE	2.37/2.68	2.53 ± 0.07	0 out of 6	2.52 ± 0.05 n.s. (96 %)	0 out of 10	2.52 ± 0.03 n.s. (96 %)	0 out of 10	2.54 ± 0.05 n.s. (95 %)	0 out of 10
CREAT [µmol/l] in SERUM	FALSE	40.0/53.0	45.5 ± 3.1	0 out of 6	42.3 ± 2.2 ** (70 %) (-)	0 out of 10	41.8 ± 1 *** (77 %) (-)	0 out of 10	41.9 ± 1.9 *** (76 %) (-)	1 out of 10
FIBRINO [g/l] in PLASMA	FALSE	1.95/2.77	2.35 ± 0.21	0 out of 6	2.14 ± 0.38 n.s. (98 %)	1 out of 6	2.38 ± 0.13 n.s. (100 %)	0 out of 5	2.49 ± 0.23 n.s. (100 %)	0 out of 6
GLUC [mmol/l] in SERUM	FALSE	4.20/13.6		0 out of 6	4.66 ± 0.55 * (60 %) (-)	3 out of 10	4.72 ± 0.68 n.s. (54 %)	2 out of 10	4.91 ± 0.5 n.s. (70 %)	1 out of 10
LDH [U/l] in SERUM	FALSE	89.0/1620		0 out of 6	1270 ± 490 n.s. (54 %)	3 out of 10	975 ± 410 n.s. (89 %)	0 out of 10	824 ± 240 n.s. (90 %)	0 out of 10
NEUT [g/l] in PLASMA	FALSE	0.509/1.46		0 out of 6	0.93 ± 0.24 n.s. (100 %)	0 out of 9	0.91 ± 0.12 n.s. (98 %)	0 out of 8	1.18 ± 0.26 n.s. (72 %)	2 out of 10
OWBW_KIDNEY [%] in KIDNEY	FALSE	0.637/0.93		0 out of 6	0.78 ± 0.05 n.s. (97 %)	0 out of 10	0.78 ± 0.05 n.s. (97 %)	0 out of 10	0.8 ± 0.05 n.s. (90 %)	0 out of 10
PHOS [mmol/l] in SERUM	FALSE	1.24/2.38	1.81 ± 0.32	0 out of 6	2.14 ± 0.11 n.s. (78 %)	0 out of 10	2.18 ± 0.2 n.s. (74 %)	2 out of 10	2.07 ± 0.26 n.s. (93 %)	1 out of 10
PROT [g/l] in SERUM	FALSE	53.4/65.6	60 ± 2.9	0 out of 6	54.5 ± 1.2 *** (100 %) (-)	3 out of 10	54.2 ± 1.7 *** (100 %) (-)	3 out of 10	56.1 ± 1.9 *** (84 %) (-)	1 out of 10
SODIUM [mmol/l] in SERUM	FALSE	140/147	143 ± 1.6	0 out of 6	146 ± 1.3 *** (94 %) (+)	2 out of 10	146 ± 1.2 *** (91 %) (+)	1 out of 10	145 ± 0.95 ** (67 %) (+)	0 out of 10
THROMNUC [g/l] in PLASMA	FALSE	569/1010	773 ± 110	0 out of 6	785 ± 120 n.s. (100 %)	1 out of 9	738 ± 86 n.s. (100 %)	0 out of 8	753 ± 150 n.s. (100 %)	1 out of 10
UREA [mmol/l] in SERUM	FALSE	6.23/9.67	7.98 ± 0.83	0 out of 6	7.57 ± 0.63 n.s. (87 %)	0 out of 10	7.34 ± 0.48 n.s. (63 %)	0 out of 10	6.98 ± 0.68 ** (66 %) (-)	1 out of 10
WEIGHT_KIDNEY [g] in KIDNEY	FALSE	1.91/2.89	2.38 ± 0.24	0 out of 6	2.37 ± 0.14 n.s. (98 %)	0 out of 10	2.39 ± 0.22 n.s. (97 %)	0 out of 10	2.44 ± 0.23 n.s. (96 %)	0 out of 10
ALP [U/l] in SERUM	TRUE	125/335	210 ± 56	0 out of 6	176 ± 36 n.s. (97 %)	0 out of 10	170 ± 33 n.s. (97 %)	0 out of 10	169 ± 22 n.s. (96 %)	0 out of 10
ALT [U/l] in SERUM	TRUE	55.6/118	77.5 ± 14	0 out of 6	87.2 ± 70 n.s. (100 %)	3 out of 10	78.9 ± 29 n.s. (100 %)	2 out of 10	70.6 ± 11 n.s. (100 %)	1 out of 10

Figure 4: Screenshots of the two csv files provided to the study director for the comparative assessment of the results. *The upper part of the table shows the results as reported for the original study, where statistically significant increases of a parameter are highlighted with light red and decreases with light blue. The lower table lists the parameters after replacing the CCGs with VCGs (the example is taken from study BD2-Q).*

Treatment-relatedness was assessed by the SME with identical criteria as used for the original study. For example, statistically significant changes were attributed to the treatment with the test item if they showed dose-dependency, if the changes were above or below 2σ of the historical control ranges, or if the direction of the changes was of toxicological relevance (e.g. LDH decreases are usually not considered toxicologically relevant).

This assessment was done by an SME, who was not the study director of the original study and the SME compared their judgment with the original outcome and commented the observed differences (see Figure 5). Noteworthy or treatment-related findings identified by the SME were collected. To facilitate the overview of the results, a standard comparative table has been developed based on Gurjanov *et al.* (2024a). This table also includes an assessment of the replicability of findings as discussed in Steger-Hartmann *et al.* (2025).

Parameters with statistically significant changes (either in CCG or in VCG)	Assessment "test substance related" VCG	Assessment "test substance related" CCG (original study)	Evaluation
ALP [U/L] in SERUM	false	false	ALP decrease was not observed in original study
ALT [U/L] in SERUM	*true*	false	Dose-dependent increase of ALT in MD_F and HD_F not observed in original study
AST [U/L] in SERUM	false	false	AST decrease was not observed in original study
ATYP [G/L] in PLASMA	false	false	Increases in LD_F, MD_F and LD_M were not found in original study
BASO [G/L] in PLASMA	*true*	false	Despite a clear lack of dose-dependency, effects were observed in all groups, partly above 2sigma of HCD.
BILI [umol/L] in SERUM	true	true	BILI increase in LD_M not observed in original study
CA [mmol/L] in SERUM	false	false	Ca increase in MD_M was not observed in original study
CK [U/L] in SERUM	*true*	false	CK decrease was only observed in MD_M, effect could be due to high water intake
CL [mmol/L] in SERUM	true	true	CL reproduced, now starting at LD_M
CREAT [umol/L] in SERUM	false	false	CREAT increases were not observed in original study
GGT [U/L] in SERUM	*true*	false	GGT increase in females was not detected in original study
GLUC [mmol/L] in SERUM	*true*	false	No significant changes in GLUC detected in original study
HCT [L/L] in PLASMA	false	false	No changes in HCT detected in original study
HGB [G/L] in PLASMA	false	false	No changes in HGB detected in original study
K [mmol/L] in SERUM	false	false	Changes in LD-F and MD_F of original study was disregarded

Figure 5: Screenshots of the csv files comparing and assessing the treatment-relatedness between VCG and the original results (example from study BD1-B). *Mismatches, which were considered of toxicological relevance are highlighted with a red box (e.g. ALT, BASO, CK, GGT and GLUC considered to be treatment-related elevated in the analysis with VCG but not in the original study). Comments highlighted in green indicate identical assessments between CCG and VCG, despite differences in the statistical outcome for individual groups.*

Thereafter, noteworthy and treatment-related findings were presented to the study director of the original legacy study, who then evaluated whether missing or novel findings identified with the VCG reanalysis would have influenced the overall study outcome. Specifically, this assessment of replicability considered the effect of the VCG replacement on the threshold doses determined in the study (NOEL, NOAEL, LOEL, LOAEL, STD10), the identified target organs, and the correlation with clinical pathology parameters indicating monitorability of the target organ toxicity. The workflow for the comparison of the overall study outcome is depicted in Figure 6 (from Gurjanov *et al*., 2024a).

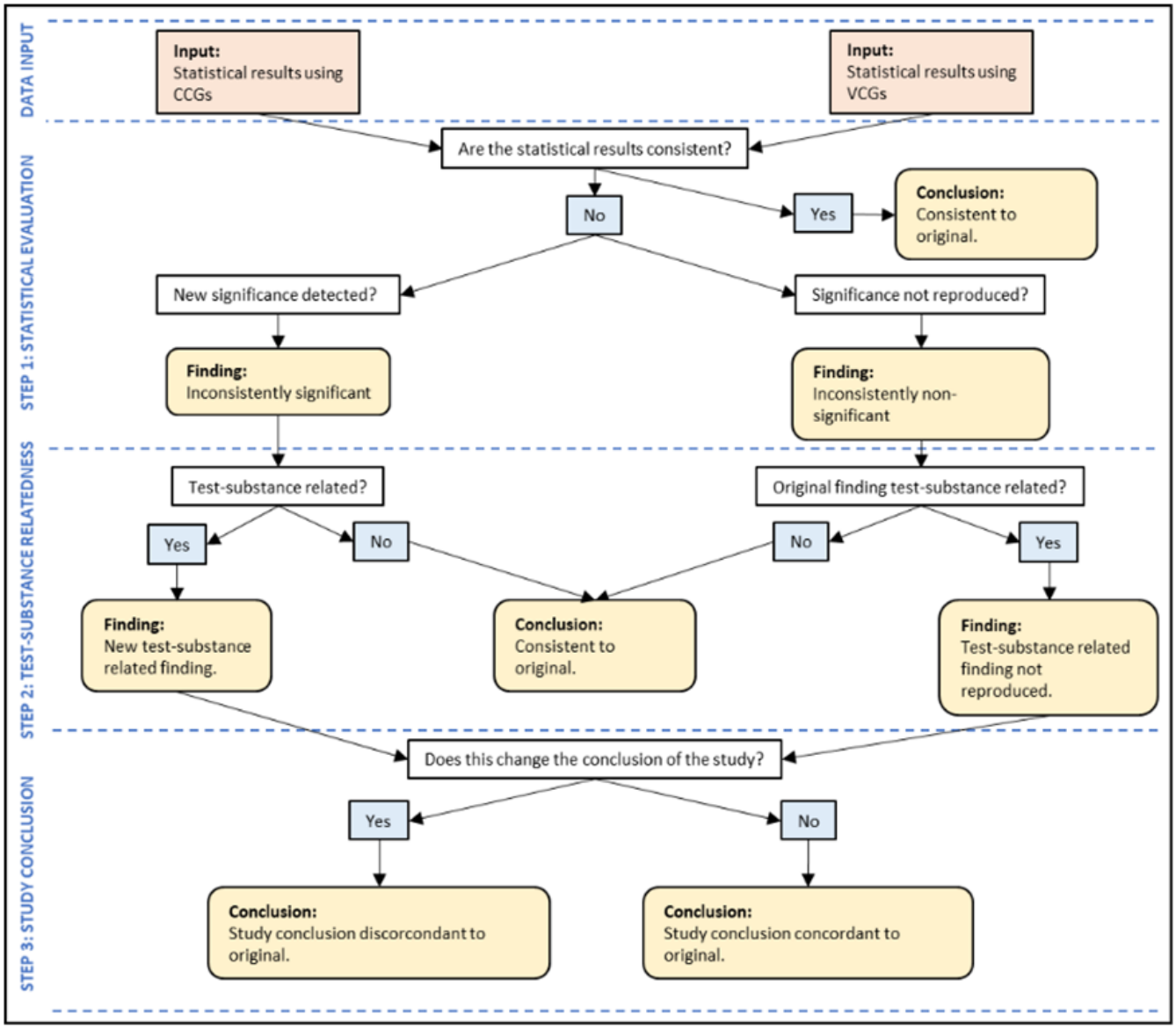


Figure 6: Workflow describing the performance assessment for study outcome using VCGs.

Non-clinical Validation of the VCG Approach

1. SOP – VCG selection procedure

Upon request of the SAWP QT, a SOP was developed (see Annex 1) that describes the recommended processes for generation, application and documentation of VCGs in preclinical studies. This master SOP is adapted by every test facility to fulfil the requirements of internal quality assurance frameworks. For CoU I the matching procedures were aligned with the recommendations of the SOP to the maximum extent possible. Deviations from the described processes are possible and minimal and are/should be documented according to the requirements in the individual Test Facilities.

The SAWP QT discussed the following deviations:

- Concerning the matching criteria, the master SOP lists "advisable" criteria which should be identical between treatment group animals and VCG animals (i.e. Test facility, Species & strain & Sex, Supplier/breeder, Origin (particularly relevant for NHPs)). In addition, the SOP mentions that the

other listed matching criteria can be relaxed to satisfy the required number of animals. This flexibility should be supported by either additional analyses or scientific rationale to ensure that the selected HCD animals for constructing VCGs come from a similar population of animals as those used in the treatment group. While flexible application of the matching criteria is allowed as per master SOP to expand the selected HCD pool size, the impact of this on the overall outcome of the VCG use is not clearly established, although the provided analysed studies provide reassurance in the context of setting a NOAEL/HNSTD in the non-GLP DRF setting.

- The master SOP provides a time window for the selection of HCD to be used as VCG, whilst mentioning that data outside of this time window may still be used, if statistical analyses show that animals from studies older than 5 years show the same distribution compared to the animals of the treatment groups or the animals assigned to a new prospective study. This is acknowledged and can be supported for CoU I, when statistical analysis pertains to all measured endpoints in the DRF non-GLP study.

- The SOP provides the option for the use of in-house historical control data (HCD), provided they confirm with CDISC SEND data structure format and adhere to the corresponding Controlled Terminology or other reference terminologies which can be mapped to SEND. The data analyses demonstrate a variability of data sources – from in house HCD to use of VICT3R database. In the frame of qualification, the use of a central repository of the HCD would be preferred to minimize variability and allow for broader access to HCD and thus more broadly foster the VCG generation.

In conclusion, the SAWP QT agreed that the flexible application of the SOP requirements to the non-GLP DRF studies, subject of CoU I, can be considered acceptable considering the lack of standardization in the design of the non-GLP DRF studies.

A set of main matching criteria is provided in the SOP which are recommended for building VCGs. For the reanalysis of the eight legacy rat repeated dose toxicity studies detailed in Appendix 2 the following criteria were applied in all eight studies (see Table 1):

- test facility
- species
- strain
- sex
- animal provider
- housing
- dosage duration [dosing period]
- recovery period duration [where applicable]
- 5-year range
- diet type
- average temperature
- humidity
- light cycle

The following deviations are observed (see Table 1):

- Regarding the criteria "treatment schedule" in study BD2-P and study H, the treatment schedule was BID, whereas the control animals used to generate VCGs had a QD treatment schedule.
- Regarding the criteria "vehicle" the vehicle was PEG 400 in BD2-Q, whereas the control animals used to generate VCGs were treated with PEG 400/Cremophor RH 40/Imwitor/40/35/25 (v/v/v) +0.5% SDS.

All studies used initial body weight as additional matching criteria with treatment group animals (Gurjanov et al. 2024b).

545 Table 1: Overview of the different matching procedures used for the reanalysis of the studies (M: Matched; Not M: Not Matched; N/A: Not applicable).

Study Number	Study Label	Test Facility	Species	Strain	Sex	Animal provider	Housing	Route of Admin.	Treatment. schedule	Dosing period	Recovery period duration	Vehicle	HCD period	Age of VCG animals	Initial BW	Other matching criteria	VCG size	VCG iteration
1	BD1-A	M	M	M	M	M	M	M	M	M	N/A	M	2017-2022	6-11 weeks	Within min and max values for animals in original dose groups	Diet Type, Temp, Humidity, Hours of Light	Equal to CCG size	100
2	BD1-B	M	M	M	M	M	M	M	M	M	M	M	2017-2022	6-11 weeks	Within min and max values for animals in original dose groups	Diet Type, Temp, Humidity, Hours of Light	Equal to CCG size	100
3	BD1-C	M	M	M	M	M	M	M	M	M	N/A	M	2017-2022	6-11 weeks	Within min and max values for animals in original dose groups	Diet Type, Temp, Humidity, Hours of Light	Equal to CCG size	100
4	BD1-D	M	M	M	M	M	M	M	M	M	N/A	M	2017-2022	6-11 weeks	Within min and max values for animals in original dose groups	Diet Type, Temp, Humidity, Hours of Light	Equal to CCG size	100
5	BD2-O	M	M	M	M	M	M	M	M	M	N/A	M	2017-2022	6-11 weeks	Within min and max values for animals in original dose groups	Diet Type, Temp, Humidity, Hours of Light	Equal to CCG size	100
6	BD2-P	M	M	M	M	M	M	M	Not M: HCD is QD, study is BID	M	N/A	M	2017-2022	6-11 weeks	Within min and max values for animals in original dose groups	Diet Type, Temp, Humidity, Hours of Light	Equal to CCG size	100
7	BD2-Q	M	M	M	M	M	M	M	M	M	N/A	Not M: HCD is PEG400, study is PEG 400 / Cremophor RH 40 / Imwitor / 40/35/25 (v/v/v) +0.5% SDS	2017-2022	6-11 weeks	Within min and max values for animals in original dose groups	Diet Type, Temp, Humidity, Hours of Light	Equal to CCG size	100
8	H	M	M	M	M	M	M	M	Not M: HCD is QD, study is BID	M	N/A	M	2016-2021	7-10 weeks	Within min and max values for animals in original dose groups	Diet Type, Temp, Humidity, Hours of Light	to CCG size	2

546

1. Criteria for assessing the replicability of study results

A process for VCG qualification was established, which combines statistical approaches and the procedure for comparing original study results with those obtained using VCGs (Golden et al. (2024)). Considering the heterogenous nature of DRF studies (e.g., non-regulated design and flexible evaluation), a certain degree of variability occurred across the different test facilities and sponsors regarding the application of statistical analysis and involved experts.

Each individual study was reanalysed by a study director and/or a subject matter expert (e.g., a clinical pathologist) identified by each respective study contributor ("sponsor"). Hence, company-specific profiles and procedures contributed to differences in result generation, evaluation, final comparisons and conclusions.

In the process of comparing results obtained with CCGs and VCGs, the impact of missing or additionally detected noteworthy or treatment-related changes on the overall study conclusion was assessed, focusing on the four main objectives of an animal safety study:

- Identification of target organ(s) of toxicity
- Identification of measurable markers of toxicity (monitorability)
- Identification of threshold dose(s)
- Identification of reversibility (recovery)

1.1. Replication of target organ(s) of toxicity

Was there any new identified or missed target organ after replacing the CCG with VCG?

While the probability of overriding necropsy or histopathology findings or a prominent change in organ weight observed in the original study is generally low, it could occur for two theoretical cases:

- The background incidences in the VCGs could be higher than in the CCGs, putting a histology finding in the treatment group into question.
- The replacement of CCGs with VCGs yields a remarkable increase or decrease of clinical pathology finding or organ weight related to a target organ toxicity despite lacking histology findings.

The replication of target organs of toxicity was considered successful when the identified target organs were the same as those specified in the original study report or if no target organ of toxicity was identified either with CCG or with VCG.

Conversely, replication failed when a specific target organ of toxicity could no longer be identified with VCGs or if a new target organ was found with VCGs.

1.2. Replication of biomarkers (monitorability)

Are there measurable parameters (biomarkers) in the in-life phase useful for monitoring the onset of toxicity?

In this context, a biomarker was defined as any quantitative parameter determined in the in-life phase of the study (e.g., cardiovascular or clinical pathology parameter) which can be related to the identified target organ of toxicity. Examples include a transaminase increase in case of histopathology liver findings or heart rate and/or blood pressure changes in case of histopathology heart findings. Such

biomarkers will allow the monitoring of toxicological effects during the in-life phase of a preclinical study, but also in the clinical setting, particularly if they occur at doses below the histology findings.

The replication of biomarkers was considered successful when the originally observed parameters were replicated with VCG. The replication was also considered successful if no biomarkers were identified in the original study and the same held true with VCG. In the opposite case, this criterion was not satisfied if a finding for a quantitative parameter was described in the original study report and could not be replicated with VCGs.

1.3. Replication of threshold doses

Was there any change of threshold doses such as NOEL, NOAEL, LOEL, LOAEL, HNSTD or STD10 after replacing the CCG with VCG?

The replication of threshold doses was considered successful when the findings determining a toxic does remain equal to those indicated in the original study report. The replication of a threshold dose was also considered successful if a toxicologically relevant finding was neither explicitly assessed nor determined in the original study. A threshold dose was therefore also considered replicable if the VCG did not result in new toxicologically relevant findings, which would have resulted in a threshold dose identification not seen with CCG.

Notably, the described replicability of threshold doses is important in the light of the purpose of DRF studies. The upper dose level for later studies derived from results of DRF studies is evidently not impaired using VCGs. Since the upper dose 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 has an influence in these findings.

1.4. Replication of reversibility of findings (recovery)

Do adverse findings regress over time, and animals recover from toxicity?

The reversibility of findings refers to the ability of an organ, tissue, or a measured parameter to recover and return to the normal state after withdrawal of the test substance. The replicability of the recovery was assessed after the end of the recovery period, replacing CCGs with VCGs for one study and hence these data are only considered as supportive (see Table 1).

The reversibility of findings (recovery) was considered successful if no new noteworthy finding was detected in the recovery phase or if the recovery-related noteworthy findings remained equal to those in the original study report.

1.5. Replicability assessment

Table 3 provides a summary of the replicability of the four key study objectives described above.

Table 3: Overview of the replicability of results from reanalysed studies. Column "Replicability of Threshold Dose" – was the NOEL, NOAEL, LOEL; LOAEL, HNSTD or STD changed after replacing the CCG with VCG? "Yes" or "No". Column "Replication of Target Organ of Toxicity" – was there any new target organ identified or missed after replacing the CCGs with VCGs? "Yes" to be used also in case if both CCG and VCG identified no target organ tox; "No": if a target organ tox was no longer identified with VCGs or if a new target organ tox was found with VCGs. Column "Replication of Biomarkers (monitorability)": all quantitative parameters (CV, clinical pathology) which are related to the target organ toxicity: "Yes" in case the originally observed parameters were replicated with VCG; "No" if the parameters were not replicated with VCGs. If in the original study no biomarkers were identified and the same was true also after reanalysis with VCG, the answer is also "yes". N/A: not applicable.

Study Number	Study Label	Replication of Target Organ(s) of Toxicity	Replication of Biomarkers (monitorability)	Replication of Threshold Dose	Replication of Reversibility Results (Recovery)	Further observations and comments
1	BD1-A	Y	Y	Y	N/A	NOAEL = HD
2	BD1-B	Y	Y	Y	Y	STD10>HD, some clinical pathology parameters observed with VCG in the recovery group point towards a persisting pharmacodynamic effect (increase in Gluc, decrease in LDH), which was not evident with CCG.
3	BD1-C	Y	Y	Y	N/A	NOAEL = MD
4	BD1-D	Y	Y	Y	N/A	NOAEL = HD
5	BD2-O	Y	Y	Y	N/A	STD10 > HD. The few observed inconsistency of quantitative parameters between VCG and CCG were of minor relevance and without influence on the overall outcome of the study.
6	BD2-P	Y	Y	Y	N/A	LOAEL = LD. The few observed inconsistency of quantitative parameters between VCG and CCG were of minor relevance and without influence on the overall outcome of the study.
7	BD2-Q	Y	Y	Y	N/A	NOEL=LD, NOAEL=HD; dose-dependent significant changes observed only with VCG were considered to have no biological relevance.
8	H	Y	Y	Y	N/A	NOAEL = MD

Overall, the replacement of CCGs with VCGs allowed a good replication of the target organs of toxicity identified in the original study as well as the related biomarkers of toxicity in all 8 reanalysed rat studies.

All 8 studies also replicated the original threshold dose, e.g. the application of VCGs did not result in changes of the dose considered to be the NOEL, NOAEL, LOEL, LOAEL, HNSTD or STD10 respectively of the original study. Replication of reversibility of results was successful in the single reanalysed studies which included a recovery period.

One of the main reasons for achieving this acceptable replicability is the fact that parameters determining dose thresholds such as mortality, severe clinical observations, and histopathologic lesions are not affected by replacing CCGs with VCGs. Differences between treatment groups and controls in quantitative parameters, such as body weight, weight gain, food and water consumption, and clinical pathology parameters may be affected by VCGs. However, these differences did not reach an extent which would have led to different conclusion regarding the NOAEL or other threshold doses.

Importantly, the grouping of certain matching criteria (treatment schedule for study BD2-P and study H, vehicle for study BD2-Q) did not result in an impairment of replicability.

In summary, for the 8 reanalysed rat repeated dose toxicity studies described in this document, the overall conclusions (i.e. the study summary and conclusions) remained unchanged, thereby showing that the performance of the VCG is fully or sufficiently consistent with the scientific interpretation made in the original report.

1.6. Deviations of replicability

With respect to the performance of VCGs, deviating observations were made and are discussed in more detail here.

- In the study BD2-O, the replacement of CCG with VCG resulted in no new noteworthy findings. However, not all findings identified as noteworthy in the original reports could be replicated pertaining to increases in leucocyte, lymphocyte and basophil counts of the HD male animals, as well as a decrease in glucose and an increase in protein of the HD female animals. Leucocyte and lymphocyte count as well as glucose decrease could also not be replicated in study BD2-P VCGs HD male and female animals respectively. However, these missed findings occurred at a dose identical or higher than those with treatment-related histological findings, i.e., this observation did not affect the conclusion regarding biomarker and monitorability of the study.
- For study BD2-P, the comparison of VCG with CCG resulted in two potentially new noteworthy findings while some findings identified as noteworthy in the original reports could not be replicated when using a VCG. However, a careful analysis of both studies has shown that these missed replications did not alter the final conclusions of this study.
- In the study BD2-Q, the replacement of CCG with VCG resulted in two new statistically significantly changed findings for all dose groups and both sexes. The extent of these changes was small and within LON. Hence, these new findings were considered to have no biological relevance and no impact on the study conclusion.
- Small effect fluctuations, probably due to technical variability and inter-animal variability, are inherent to non-clinical studies and are therefore unavoidable (Steger-Hartmann et al. 2025). These discrepancies are considered to be related to the inherent variability and experimental

imprecision, which would most probably also occur if the studies were repeated with the same experimental setting.

- In study BD2-Q, the replacement of CCGs with VCGs resulted in statistically significant differences of quantitative parameters which were not identified in the original study. However, a further evaluation of these statistically significant findings showed that they are not biologically relevant or not treatment-related.
- Statistically significant treatment effects were not always considered decisive, i.e. a statistically significant change could be dismissed by the SME, and conversely, a statistically non-significant change could be deemed relevant. The main reason was the emphasis placed on dose-dependency, magnitude of the effect, limits of normal and biological relevance. This reflects standard practices in the field.

2. Conclusions on the replicability assessment

To substantiate the request for qualification opinion for CoU I the Applicant provided the results of eight reanalysed rat repeated dose toxicity studies. The key requirement for the implementation of the proposed VCG concept is the assurance that using VCGs does not compromise study outcomes and thus does not endanger human safety. Thus, each individual study reanalysis aimed to determine whether the replacement of CCGs with VCGs would alter the study overall conclusions and if the performance of the proposed VCG approach can be considered robust and reliable. To this end the replicability analysis focused on identification of target organ(s) of toxicity, measurable markers of toxicity (monitorability), threshold dose(s), and reversibility (recovery). Since the data were to scarce (only 1 study) to assess reversibility, these data are considered as supportive only.

The submitted reanalysed studies confirm that the VCG approach allows for the identification of target organs and NOAEL/HNSTD setting guiding the dose setting for later GLP repeated dose toxicity studies.

Life-cycle management

When VCGs are used in the context of a Marketing Authorization Application (MAA), any updates of the procedures described in this Qualification Opinion should be submitted at time of MAA. Such updates, whether to be included in the MAA or in a follow-up Qualification Opinion, should be clearly identified, listed and justified.

Structural changes to the SOP arising from prospective insights gained during the ongoing development of future Contexts of Use should be addressed through an update of the Qualification Opinion.

Overall conclusion:

Based on the evidence presented by the Applicant and reviewed by the Qualification Team, CHMP considers that Virtual Control Groups (VCGs) can be used to substitute for Concurrent Control Groups (CCGs) in non-clinical non-GLP rat dose range finding (DRF) studies to inform on dose selection for subsequent pivotal rat GLP-compliant repeated dose studies, when applied in accordance to the SOP (see Appendix 1).

Appendices to be published:

1. Standard Operating Procedure (published separately)
2. Individual reanalysed rat repeated dose toxicity studies
3. Summary Qualification Document (published separately)
4. Written responses to Lists of Issues (published separately)

Appendix 1: Standard Operating Procedure (published separately)

Appendix 2: Individual reanalysed rat repeated dose toxicity studies

In this section, an overview is provided of the eight rat studies for which a VCG approach was applied retrospectively to support the current context of use (CoU I). These comprise studies BD1-A through BD1-D, study H and studies BD2-O, BD2-P and BD2-Q. The objective of these reanalysed studies was to assess whether a NOAEL using a VCG approach would be comparable to the NOAEL established in the original study, with the intention that this NOAEL would inform dose selection for future (pivotal) GLP repeated dose toxicity studies.

With respect to the reanalysis of DRF studies, it should be noted that the design and evaluation procedure of such studies are only broadly described in relevant guidelines and are subject to variability. Consequently, reanalysis using VCGs necessarily requires consideration of this heterogeneity. In addition, reanalysis procedures may differ depending on the species used and on data collection practices for pre-values (parameter concentrations measured before dosing period starts). For large animal studies, pre-values are typically measured, whereas in small animals (i.e. rodents) limited blood volume precludes the collection of such data. Procedures also vary according to group size. When control and treated groups comprise only one or two animals per sex, statistical analysis is not feasible, and evaluation is based on visual inspection for trends. For larger group sizes, both statistical analysis and visual inspection were performed. In all cases, biological and toxicological interpretation of the findings takes precedence over statistical testing, in line with standard practices in non-clinical safety assessment.

In view of these considerations, the Applicant chose not to apply a single, fully harmonized reanalysis procedure across all DRF studies. Instead, individual approaches were adopted that best reflect the evaluation procedures routinely implemented for DRF studies at the respective test sites.

Study BD1-A (rat, 4-week, GLP)

Original study

The test item was investigated for its cumulative toxicity by administration once daily through oral (gavage) route for a 4-week period to 10 male and 10 female rats [Crl:WI] per group with a suspension of the test item in Ethanol/Kolliphor® HS15/Water for Injection (10/40/50 v/v/v) as vehicle at three doses (LD, MD, HD) over a period of approximately 4 weeks (28 to 30 administrations) with an administration volume of 10 mL/kg. Animals underwent necropsy one day after end of treatment. A control group of 10 males and 10 females was treated likewise with an equivalent volume of the vehicle.

In this 4-week toxicity study, no mortality, no clinical observations during the in-life phase, and no histopathological findings were noted that were regarded as treatment-related. Changes of four quantitative parameters (water intake, blood glucose, total bilirubin, and absolute plus relative liver weights) were found to be noteworthy. While these findings were associated with the treatment, they were not judged as being adverse. Therefore, the high dose was determined as NOAEL.

A rise of bilirubin may be a marker for cholestatic DILI (FDA, 2009; Trost, 2014). However, this parameter was only slightly increased (within the upper limit of normal (ULN)), no other clinical chemistry parameter associated with DILI was increased (e.g., ALT, AST, and ALP) and despite an increase in relative and absolute liver weight no histological changes in the liver were observed.

Study conclusions after replacing CCGs with VCGs

While the increased water intake was not seen with VCGs, the findings in glucose were additionally visible in females starting from the lowest dose. Regarding relative liver weight only mid dose group and high dose group showed an increase, whereas with VCGs the low dose animals were affected too. Of note is that the significant change in bilirubin in female high dose was not reproduced with VCGs. However, a new noteworthy finding was observed: a significant increase in gamma-GT was now present throughout all doses in all sexes. GGT is a sensitive—but not specific—marker for cholestatic DILI (Trost, 2014; Robles-Diaz et al., 2015). The GGT values in all doses was only slightly beyond the ULN. Given the small effect and the lack of concomitant histological findings, the increases in GGT were considered irrelevant. The conclusion of the legacy study A therefore remains unchanged: 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.

Table A: Table of noteworthy findings in legacy study A. CCG: concurrent control group, VCG: virtual control group, M: males, F: females, HCD: historical control data. LD: low dose, MD: mid dose, HD: high dose.

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**
Water consumption	+	M+F	HD	not reprod.			
Glucose (whole blood)	-	M	HD	already at LD	-	M/F	HD/LD
Total bilirubin	+	F	LD	not reprod.			
gamma-Glutamyl transferase				New	+	M+F	LD
Absolute liver weight	+	F	MD	Consistent	+	F	MD
Relative liver weight	+	F	MD	already at LD	+	F	LD
Noteworthy pathological findings							
None							

Study BD1-B (rat, 4-week, GLP)

Original study

This study investigated possible effects of a potential anti-cancer drug. Unlike the other two legacy studies, this study has a 2-week recovery group to assess the reversibility of potential noteworthy findings. Cancer therapy is often accompanied with marked side effects resulting frequently in poor general condition of the animals. Therefore, instead of a NOAEL usually the severely toxic dose (STD10) is the threshold of interest, i.e., the dose at which 10 % of animals show poor general condition leading to premature sacrifice (Maziasz et al., 2010).

The test item was investigated for its cumulative toxicity by administration once daily through oral (gavage) route for a 4-week period to 10 male and 10 female rats [Crl:WI] per group with a suspension of the test item in Ethanol/Kolliphor® HS15/Water for Injection (10/40/50 v/v/v) as vehicle at three doses (LD, MD, HD) over a period of approximately 4 weeks (28 to 30 administrations) with an administration volume of 10 mL/kg. A control group of 10 males and 10 females was treated likewise with an equivalent volume of the vehicle. In addition, 6 animals per sex and dose were treated with HD together with a control group of 6 animals for approx. 4 weeks followed by a 2-week recovery period. Animals underwent necropsy one day after end of treatment.

Dosing up to HD was well tolerated, and no test item-related effects on mortality, clinical observations, ophthalmology, organ weights nor gross observations at necropsy were detected.

Animals of both sexes dosed at HD showed a slightly retarded body weight gain during the first week of dosing. Correspondingly, mean food intake was reduced in males at HD and females at MD and above during the first week of dosing. Water consumption was statistically significantly increased in all treated groups without a clear dose-dependent increase among the different groups.

Haematology revealed a slight statistically significant decrease in prothrombin time. Clinical chemistry revealed statistically significant deviations in several parameters: protein was statistically significantly decreased at HD in males. Total bilirubin was statistically significantly increased at MD and above in both sexes. Chloride was statistically significantly decreased starting at MD in males.

Urinalysis revealed a statistically significantly increased urinary protein to creatinine ratio at LD and above.

Histopathological evaluation revealed moderate de- and regeneration in the Harderian gland considered to be treatment-related in HD animals. After 2 weeks recovery, only slight to minimal changes were detected in two females indicating partial reversibility. Up to slight thyroid follicular hypertrophy was recorded treated and vehicle control groups with slightly higher incidences in males starting at MD and females starting at LD when compared to vehicle-treated controls.

After 2 weeks of recovery, no difference in incidences was seen between vehicle and treated groups. Minimally to slightly increased numbers of thymic tangible body macrophages were recorded in HD animals. After recovery, no relevant differences regarding the numbers of tangible body macrophages between control and treated animals were seen. In the spleen, a higher incidence of extramedullary haematopoiesis was observed in LD animals and above. The severity grade did not exceed a total severity score of slight. After the treatment-free period, only one female displayed minimal extramedullary haematopoiesis.

Study conclusions after replacing CCGs with VCGs

Neither mortality nor poor general condition was observed during the dosing period. Replacing CCGs with VCGs is not bound to change these findings. Therefore, the conclusion of this study regarding STD10 remained unchanged after replacing the CCG with VCGs.

807 Table B: Table of noteworthy findings in legacy study B. CCG: concurrent control group, VCG: virtual
808 control group, M: males, F: females, HCD: historical control data. LD: low dose, MD: mid dose, HD:
809 high dose.

Mortality							
None							
Clinical findings							
None							
Quantitative parameters							
Original noteworthy findings using the 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 gain (1st week)	-	M+F	HD	consistent	-	M+F	HD
Food consumption (1st week)	-	M/F	HD/MD	consistent	-	M/F	HD/MD
Water consumption	+	M+F	LD	consistent	+	M/F	MD/LD
Total bilirubin	+	M+F	MD	now in LD (M)	+	M/F	LD/MD
Chloride	-	M	MD	now in LD	-	M	LD
Protein/ Creatinine ratio	+	M+F	MD	now in LD	+	M+F	LD
Protein	-	M	HD	not reprod.			
Alanine aminotransferase				new	+	F	MD
Basophils				new	+	M+F	LD
gamma-Glutamyl transferase				new	+	F	LD
Glucose (whole blood)				new	+	M+F	LD

810

Noteworthy pathological findings					
Organ/tissue	**Finding**	**Sex (M/F)**	**Starting dose**	**Background incidences in CCG**	**Background incidences in HCD**
Thyroid gland	Follicular cell hypertrophy	M+F	LD	M: 0 out of 10 (0 %) F: 0 out of 10 (0 %)	M: 11 out of 116 (9 %) F: 2 out of 109 (2 %)
Thymus	Tingible body macrophages, increased	M+F	HD	M: 0 out of 10 (0 %) F: 0 out of 10 (0 %)	M: 0 out of 116 (0 %) F: 0 out of 109 (0 %)
Harderian gland	Degeneration/regeneration	M+F	HD	M: 0 out of 10 (0 %) F: 0 out of 10 (0 %)	M: 1 out of 116 (1 %) F: 1 out of 109 (1 %)
Spleen	Extramedullary hematopoiesis	M+F	LD	M: 0 out of 10 (0 %) F: 1 out of 10 (10 %)	M: 22 out of 116 (19 %) F: 35 out of 109 (32 %)

811

Study BD1-C (rat, 4-week, GLP)

Original study conclusions

This study examined the effects of a cardiovascular drug candidate, which elicits hemodynamic effects. This mode of action led to several treatment-related effects. Due to general poor condition of animals in the high dose group this dose was classified as severely adverse. Several smaller findings in body weight, clinical observations, food and water intake, clinical pathology, organ weights, and histopathology were also observed in the mid dose group. One animal in the mid dose group was prematurely sacrificed due to poor general condition. However, no findings were observed in the histopathological examinations that would indicate a connection to the treatment. Since findings in the mid dose were rather irrelevant and the histopathological findings were non-adverse, the NOAEL was set to the mid dose group in this study.

The test item was investigated for its cumulative toxicity by administration once daily through oral (gavage) route for a 4-week period to 10 male and 10 female rats [Crl:WI] per group with a suspension of the test item in Ethanol/Kolliphor® HS15/Water for Injection (10/40/50 v/v/v) as vehicle at three doses (LD, MD, HD) over a period of approximately 4 weeks (28 to 30 administrations) with an administration volume of 10 mL/kg. Animals underwent necropsy one day after end of treatment. A control group of 10 males and 10 females was treated likewise with an equivalent volume of the vehicle.

Since findings in the mid dose were rather irrelevant and the histopathological findings were non-adverse, the NOAEL was set to the mid dose group in this study.

Study conclusions after replacing CCGs with VCGs

After replacing the CCGs with VCGs, conclusions did not change and were rather reinforced by using the background incidences for salivation, change in faeces, and high stepping gait. Several treatment-related quantitative parameter changes in the original study were not reproduced using virtual controls:

- Food intake decrease was only reproduced in males, but not in females.
- Serum protein levels increase was not reproduced in females. Rather, a decrease in serum protein was seen in both males and females (dose dependency observed only in males). Of note is that a significant increase in protein/creatinine ratio was seen in the high dose in males in the original study which was, however, not reproduced using virtual controls.
- Changes in calcium level were also not reproduced.
- Chloride levels decrease in the female animals in the original study was also observed in male animals of all dose groups in the VCGs statistical analysis.
- Liver weight increase in female animals was not reproduced after replacing CCGs with VCGs but relative liver weight increases were still observed.

The overall assessment of findings included histopathological and clinical observations. Despite the list of clinical pathology findings not reproduced in the VCGs analysis, the study director came to the same conclusion as in the original study report that the NOAEL is the mid dose.

The replacement of CCGs with VCGs did not result in any new noteworthy findings.

Replacing CCGs with VCGs had no effect on the determination of NOAEL.

The results of this analysis have also been published in Gurjanov et al. (2024a).

Study BD1-D (rat, 2-week [DRF], non-GLP)

Original study conclusions

This study investigated the appropriate doses for an oncology drug candidate which could be applied for a subsequent 4-week toxicity study under GLP.

The test item was investigated for its cumulative toxicity by administration once daily through oral (gavage) route for a 2-week period to 6 female rats [Crl:WI] per group with a suspension of the test item in Ethanol/Kolliphor® HS15/Water for Injection (10/40/50 v/v/v) as vehicle at two doses (LD, MD) with an administration volume of 10 mL/kg. Animals underwent necropsy one day after end of treatment. A control group of 6 females was treated likewise with an equivalent volume of the vehicle.

Dosing over 2 weeks did not induce any treatment-related findings in female rats up to and including a dose of 3 mg/kg. At 9 mg/kg, a slightly reduced body weight gain as well as a transiently reduced food intake were seen. Neutrophils were slightly increased at this dose level, but no histopathological correlate was detected. Significantly elevated liver enzymes (EROD, GS-T, CA-T) were observed but these changes were not considered as adverse, since they were within historical control value ranges.

Study conclusions after replacing CCGs with VCGs

The effects observed at the dose of 9 mg/kg, i.e., a slight reduction of body weight gain, transiently reduced food intake as well as the increase in neutrophils were reproduced with VCGs. Notably, the significantly increased EROD values were also reproduced, though still within the range of historical control ranges, whereas GS-T and CA-T showed no significant increase.

Study H (rat, 4-week, GLP)

Original study conclusion

In this toxicity study compound was administered twice daily via oral gavage for 28 days to Sprague-Dawley rats. Compound-related microscopic findings consisted of hypertrophy of the centrilobular hepatocytes in the liver of animals in the high dose, which correlated with increased liver organ weights in animals in the mid dose; and erosion/ulcer and/or inflammation in the non-glandular stomach of males starting in the low dose and females in the mid dose. These changes exhibited complete or partial reversal during the recovery phase. Finally, these changes in the liver and non-glandular stomach were considered non-adverse due to their low severity grades and lack of clearly correlative clinical pathology findings. Compound-related clinical observations were observed in females in the highest dose on day 1 which included hypoactive behaviour, squinting eyes, pale appearance of the entire body and/or piloerection. Additional minor changes were noted in mean body weight gains and mean food consumption and clinical pathology parameters. The NOAEL for this study was the mid dose.

Study conclusion after replacing CCGs with VCGs

The retrospective re-analysis of the study results was done by an SME/toxicologist/pathologist, who was not the study director of the original study. The SME compared his or her judgment with the original outcome and commented the observed differences. So, to conclude, there were no changes after replacing the CCGs by a randomly selected VCG from the Sanofi VCG pool. E.g., for the liver microscopic findings, none of the observed findings from the treatment groups were seen in both male and female control groups (see CCG and the example of a VCG in Figure E1). Consequently, interpretation and assessment of these liver findings come to the same conclusion whether using the CCG or VCG. Furthermore, the additionally provided VCG Min-Max and VCG incidences support the conclusion that the hepatocyte necrosis seen in one male rat in the mid dose should not be considered

as a compound-related effects (see relative high incidence of 1.4% and maximal occurrence up to 2 findings per study in Figure E1). In contrast, the minimal and mild centrilobular hypertrophy clearly exceed the VCG Min-Max and VCG Incidences and support the conclusion that these findings are compound-related.

The same was true for finding of the nonglandular stomach: there were no changes after replacing the CCGs by VCG Random from the Sanofi VCG pool. Findings in the nonglandular stomach are much rarer, confirmed also by the lower VCG Min-Max and VCG Incidence. So again, interpretation and assessment of these findings come to the same conclusion whether using the CCGs or VCGs.

Only for the kidney the CCGs significantly differed from the VCGs, so none of the observed findings from the high dose groups were reported in the VCGs whereas they were seen in CCG. But when using the VCG Min-Max and VCG Incidences, the interpretation and assessment come to the same conclusion that these kidney findings are not compound-related because of the relative high incidences and maximal occurrence of these findings in control groups per study in the same range as the high dose group.

All other microscopic findings occurred only spontaneously in one animal and or with only minimal severity grade and were considered as non-adverse in the CCGs as well as in the VCGs.

Accordingly, changes for the other clinical observations, clinical pathology and other parameters were analysed in the same way. Table E2 shows examples for which statistical analyses differ between the CCG and a randomly selected VCG from the Sanofi VCG pool as example, ordered by the most prominent changes for male and female rats. An additional graphical analysis is shown in Figure E3 for a few examples. Differences were seen in the statistical analysis, for example for ALP in male high dose animals (see also Figure 15: statistically significant in CCG high dose with $p=0.0002$, but not significant in the VCG anymore with $p=0.23$); but since microscopic changes in the liver of high dose animals was anyway considered as adverse effects, the missing significance in the VCGs did not influence the overall study conclusion.

In summary, after taking all these changes into consideration, the NOAEL of this legacy rat study H remains unchanged. The NOAEL was considered to be the mid dose, and first adverse effects were seen in the high dose.

Figure H1: Overview of main microscopic results from the legacy rat study H at study end after 28-day treatment (findings are highlighted in red, green indicates no finding).

			Male							Female						
Organ	Findings	Severity	CCG	VCG Random	VCG Min-Max	VCG Incidence	Low Dose	Mid Dose	High Dose	CCG	VCG Random	VCG Min-Max	VCG Incidence	Low Dose	Mid Dose	High Dose
LIVER	Unremarkable animals		10	10	-	-	10	9	5	10	10	-	-	10	10	2
	Hypertrophy, hepatocyte, centrilobular	Minimal	0	0	0-2	0.25%	0	0	2	0	0	0-2	0.26%	0	0	3
	Hypertrophy, hepatocyte, centrilobular	Mild	0	0	0-1	0.33%	0	0	3	0	0	0	< 0.1%	0	0	5
	Necrosis, hepatocyte	Minimal	0	0	0-2	1.4%	0	1	0	0	0	0-1	0.52%	0	0	0
STOMACH Nonglandular	Unremarkable animals		10	10	-	-	9	8	9	10	10	-	-	10	8	9
	Erosion/ulcer	Minimal	0	0	0-1	0.15%	0	1	0	0	0	0-1	0.16%	0	0	1
	Erosion/ulcer	Mild	0	0	0-1	< 0.1%	1	0	1	0	0	0	< 0.1%	0	0	0
	Hyperplasia, epithelium	Mild	0	0	0	< 0.1%	0	0	0	0	0	0	< 0.1%	0	0	1
	Inflammation, mixed cell	Minimal	0	0	0-1	0.27%	0	0	1	0	0	0-1	0.19%	0	0	0
	Inflammation, mixed cell	Mild	0	0	0-1	0.17%	0	0	0	0	0	0	< 0.1%	0	0	1
	Inflammation, neutrophil	Minimal	0	0	0	< 0.1%	1	1	0	0	0	0	< 0.1%	0	2	0
KIDNEY	Unremarkable animals		7	10	-	-	10	10	7	5	10	-	-	10	10	5
	Degeneration/necrosis, tubule	Minimal	1	0	0-1	9.5%	0	0	3	4	0	0-4	4.4%	0	0	2
	Dilatation, tubule(s)	Minimal	2	0	0-2	1.7%	0	0	2	1	0	0-3	0.63%	0	0	4

Table H2: Overview of results for legacy study H, where the statistical analyses differ comparing CCGs with VCGs for male and female animals. Findings highlighted in red; no finding in green.

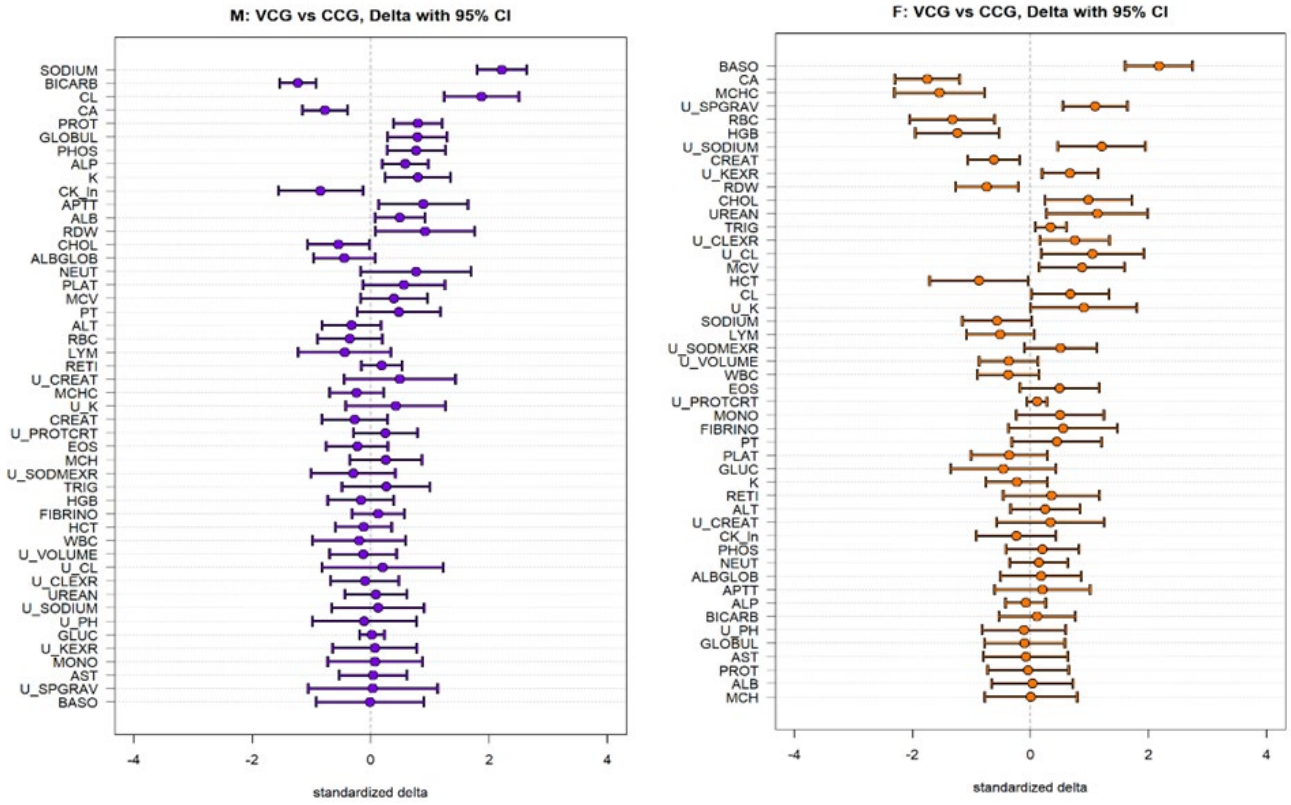
Male

	CCG Mean (SD)	VCG Mean (SD)	Delta VCG-CCG	p-value (VCG vs CCG)	p-value (low vs CCG/VCG)	p-value (mid vs CCG/VCG)	p-value (high vs CCG/VCG)
SODIUM [mmol/L]	144.7 (0.8165)	149.3 (1.447)	4.667	<0.0001	0.98 / <0.0001	0.91 / <0.0001	0.37 / <0.0001
BICARB [mmol/L]	25.07 (0.7988)	22.13 (1.125)	-2.933	<0.0001	0.58 / <0.0001	0.41 / <0.0001	0.94 / <0.0001
CL [mmol/L]	100.8 (1.014)	104.1 (1.846)	3.333	<0.0001	>0.9999 / <0.0001	0.59 / <0.0001	0.041 / <0.0001
CA [mg/dL]	11.68 (0.2597)	11.35 (0.1598)	-0.3333	0.00031	0.50 / 0.27	>0.9999 / 0.023	0.98 / 0.0065
PROT [g/dL]	6.647 (0.2066)	6.907 (0.1438)	0.26	0.00049	>0.9999 / 0.037	0.41 / 0.58	0.92 / 0.0094
GLOBUL [g/dL]	2.4 (0.1512)	2.553 (0.106)	0.1533	0.0036	0.43 / 0.60	0.10 / >0.9999	0.93 / 0.014
PHOS [mg/dL]	8.093 (0.5418)	8.627 (0.3369)	0.5333	0.0036	>0.9999 / 0.026	>0.9999 / 0.033	0.18 / 0.80
ALP [U/L]	92.85 (10.12)	104.7 (10.31)	11.89	0.0050	>0.9999 / 0.030	0.34 / 0.48	0.00022 / 0.23
K [mmol/L]	5.2 (0.2591)	5.473 (0.2434)	0.2733	0.0059	0.72 / 0.0016	0.95 / 0.0073	0.72 / 0.15
CK (ln) [U/L]	6.677 (0.5987)	6.205 (0.4461)	-0.4713	0.022	0.63 / 0.37	0.75 / 0.31	0.95 / 0.013

Female

	CCG Mean (SD)	VCG Mean (SD)	Delta VCG-CCG	p-value (VCG vs CCG)	p-value (low vs CCG/VCG)	p-value (mid vs CCG/VCG)	p-value (high vs CCG/VCG)
BASO [10^9/L]	0.01692 (0.006304)	0.04667 (0.01291)	0.02974	<0.0001	0.095 / <0.0001	0.87 / <0.0001	0.19 / <0.0001
CA [mg/dL]	11.92 (0.3406)	11.17 (0.287)	-0.7533	<0.0001	0.99 / <0.0001	0.98 / <0.0001	0.77 / <0.0001
MCHC [g/dL]	33.1 (0.6137)	32.21 (0.4655)	-0.8857	0.00036	0.77 / 0.012	0.55 / 0.029	0.0035 / >0.9999
U_SPGRAV [NA]	1.012 (0.003361)	1.021 (0.0073)	0.008714	0.00047	0.20 / 0.051	0.89 / 0.00012	0.33 / <0.0001
RBC [10^12/L]	8.383 (0.3524)	7.775 (0.4964)	-0.6084	0.00088	>0.9999 / 0.00049	>0.9999 / 0.00056	0.82 / <0.0001
HGB [g/dL]	15.56 (0.6959)	14.66 (0.6254)	-0.9015	0.0015	0.80 / 0.041	0.99 / 0.010	>0.9999 / 0.0042
U_SODIUM [mmol/L]	21.18 (8.376)	38.87 (17.74)	17.68	0.0028	0.42 / 0.062	0.71 / 0.00053	0.88 / 0.0049
CREAT [mg/dL]	0.7267 (0.04577)	0.68 (0.0414)	-0.04667	0.0067	0.99 / 0.16	0.94 / 0.51	0.83 / 0.67
U_KEXR [mmol]	0.3721 (0.1186)	0.526 (0.1658)	0.1539	0.0078	0.96 / 0.0025	0.18 / 0.32	0.028 / 0.85
RDW [%]	11.89 (0.2783)	11.45 (0.5027)	-0.439	0.0081	0.0018 / 0.79	0.25 / 0.55	0.71 / 0.0033

Figure H3: Overview of main clinical pathology results from the legacy rat study E at study end after 28-day treatment



Study BD2-O

Study description and original results (study BD2-O)

The test item was investigated for its cumulative toxicity by administration once daily through oral (gavage) route for a 4-week period to 10 male and 10 female rats [Crl:WI] per group with a suspension of the test item in Ethanol/Kolliphor® HS15/Water for Injection (10/40/50 v/v/v) as vehicle at three doses (LD, MD, HD) over a period of approximately 4 weeks (28 to 30 administrations) with an administration volume of 10 mL/kg Animals underwent necropsy one day after end of

treatment. A control group of 10 males and 10 females was treated likewise with an equivalent volume of the vehicle.

Effects of the test item were evaluated using clinical parameters (mortality, general observation, ophthalmoscopy, body weight, food and water consumption), clinical pathology (hematology, clinical chemistry, urinalysis) and full postmortem examination including necropsy, organ weight analysis and microscopic examination.

The daily oral administration of the test item to rats over 4 weeks revealed effects on body weight development, food intake, and histopathological changes indicative of phospholipidosis at the HD. The STD10 exceeded the HD.

Results after replacing CCGs with VCGs (study BD2-O)

The replacement of CCGs with VCGs resulted in the following differences regarding the statistical significance of quantitative parameters, which were not considered noteworthy in the original report.

An overview of discrepancies between findings observed with CCG and VCG is provided in Table BD2-O1.

Table BD2-O1: Discrepancies of clinical chemistry findings between CCGs and VCGs for BD2-O

Parameter	Statistical significance in original study	Statistical significance with VCG	Assessment
Ca	MD_M (+)	MD_M, MD_F (+)	No dose dependency
Cl	LD_F (-)	LD_M, MD_M, HD_M (-)	Above LLN
Creatinine	--	LD_F (-)	No dose dependency
K	HD_M (-)	MD_M, HD_M (-)	Above LLN
Phosphor	MD_F (+)	--	No dose dependency
Erythrocytes	--	HD_F (-)	Above LLN
T3	MD-F, MD-M (+)	--	No dose dependency
T4	MD-F (+)	--	No dose dependency
TSH	LD_M (+)	--	No dose dependency
Thrombocytes	--	LD-M, MD-M, HD-M (+)	Below ULN

An overview of noteworthy findings comparing CCG and VCG is provided in Table BD2-O2.

Table BD2-O2: Comparative summary of noteworthy findings for study BD2-O (CCGs vs. VCGs)

Mortality							
None							
Clinical findings							
None							
Quantitative parameters							
Original noteworthy findings using the 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 gain	(-)	M/F	HD	Consistent	(-)	M/F	HD
Food intake	(-)	M/F	HD	Consistent	(-)	M/F	HD
Hemoglobin + Hematocrit	(-)	F	HD	Consistent	(-)	F	HD
Leucocytes[1]	(+)	M	HD	Inconsistent	--	--	--
Lymphocyt.[1]	(+)	M	HD	Inconsistent	--	--	--

Basophils[1]	(+)	M	HD	Inconsistent	--	--	--
Fibrinogen	(-)	M	HD	Consistent	(-)	M	HD
Asp. AT	(+)	M	HD	Consistent	(+)	M	HD
Alanine AT	(+)	M	HD	Consistent	(+)	M	HD
Glutamate dehydrog.	(+)	M/F	HD	Consistent	(+)	M	HD
Total bilirubin	(+)	M/F	LD/MD	Partially consistent	(+)	M/F	MD
Triglycerides	(-)	M	HD	Partially consistent	(-)	M/F	HD
Glucose (whole blood)	(-)	M/F	MD/HD	Inconsistent	--	--	--
Protein (quant.)	(+)	F	HD	Inconsistent	(-)[2]	M/F	MD/LD
Protein/ creatinine (urine)	(+)	M/F	HD/MD	Partially consistent	(+)	F	HD

Organ Weights

Parameter name	Increase (+) decrease (-)	Sex (M/F)	Starting dose	Compared to CCG	Increase (+) decrease (-)	Sex (M/F)	Starting dose
Liver (rel.)	(+)	M/F	HD	Consistent	(+)	M/F	HD
Kidneys (rel.)	(+)	M	HD	Inconsistent	--	--	--
Brain (rel.)	(+)	M/F	HD	Consistent	(+)	M/F	HD
Spleen (rel.) Spleen (abs.)	(+) (+)	M/F F	HD HD	Consistent Inconsistent	(+) --	M/F --	HD --
Thymus (abs.)	(-)	F	HD	Inconsistent	--	--	--

Noteworthy necropsy findings

Liver: Pale discoloration	M	HD
Spleen: Pale discoloration, swelling	M	HD

Noteworthy histopathology findings

Organ/tissue	Finding	Sex (M/F)	Starting dose	Background incidences in CCG	Background incidences in VCG [%]
Liver	• Hepatocellular degeneration/necrosis (minimal/moderate)	M/F	HD	0/20	3.84%
	• Hypertrophy (minimal/slight)			0/20	0.76%
	• Accumulation of lipid/centrilobular (minimal/slight)			0/20	2.69%
Kidney	• Degeneration/vacuol. collecting duct (minimal/moderate)	M/F	HD	0/20	<0.38%[3]
	• Dilatation, cortical tubule (minimal/slight)			1/20	1.92%
Bone marrow (femur/sternum)	• Cellularity, increased: foamy macrophages (minimal/moderate)	M/F	HD	0/20	0.38%
	• Cellularity, increased: myelopoiesis (minimal)	M	HD	0/20	0.38%

Spleen	• Cellularity, increased: foamy macrophage (minimal/moderate) • Cellularity, increased: germinal center (number/size; mostly minimal/slight) • Extramedullary hematopoiesis, increased (minimal/moderate)	M/F M/F F	HD HD HD	0/20 0/20 5/20	0.38% 0.38% >10%
Mesent. lymph node	• Cellularity, increased: foamy macrophage (minimal/slight)	M/F	HD	0/20	0.39%
Iliac lymph node	• Cellularity, increased: foamy macrophage (minimal/slight)	M/F	HD	0/18	1.17%
Mand. lymph node	• Cellularity, increased: foamy macrophage (minimal/slight)	M/F	HD	0/18	3.14%
Thymus	• Tingible body macrophages, increased (minimal/slight) • Cellularity, decreased: cortex (minimal/slight)	M/F F	HD HD	1/20 0/20	1.92% <0.38%[3]
Ovaries	• Aggregate: foamy macrophage (minimal/moderate)	F	HD	0/10	<0.38%
Uterus	• Aggregate: foamy macrophage (minimal/slight	F	HD	0/10	<0.38%[3]
Brown adipose tissue	• Vacuolation: macrovesicular change (minimal/slight)	F	HD	N/A	N/A

[1] The values for these three parameters were within LON but were nevertheless considered noteworthy.
[2] The values of the CCGs were below LLN for M and close to LLN for F.
[3] No such finding was reported in the pool of 260 HCD animals, i.e. <1/260=<0.38%.

The replacement of CCGs with VCGs resulted in no new noteworthy finding. Some findings identified as noteworthy in the original reports could not be replicated:

- A significant increase of leucocytes in HD male animals was not found with VCGs. The observed changes were above the ULN.
- Significant increases of lymphocytes and basophils in HD male animals were not found with VCGs. The observed changes were minor and still below the ULN.
- Significantly decreased glucose values in MD males and HD females were not replicated with VCGs. The changes were considered of minor quantity in the original report.
- The significant increase in urinary protein (quantitative) of the HD females was not replicated with VCGs. Instead, a significant decrease was observed already in LD females. The CCGs had protein values close to or below the LLN, i.e. the original finding is most probably an artifact.
- The significant increase of relative kidney weights was not replicated with VCGs.
- The significant increase of the absolute spleen weight and the decrease of absolute thymus weight were not replicated with VCGs.

Noteworthy histopathology findings were not altered by the background incidences of the VCGs.

Replacing CCGs with VCGs had no effect on the threshold dose determination, i.e. the STD10 was still above the HD.

Study BD2-P

Study description and original results (study BD2-P)

The test item was investigated for its cumulative toxicity by administration twice daily (BID) through oral (gavage) route for a 4-week period to 6 male and 6 female rats [Crl:WI] per group with a

suspensions in Ethanol/Kolliphor HS 15/Water for injection (1/4/5; v/v/v) as vehicle at three doses (LD, MD, HD) over a period of approximately 4 weeks (28 to 30 administrations) with an administration volume of 5 mL/kg. Animals underwent necropsy one day after the end of treatment. A control group of 6 males and 6 females was treated likewise with an equivalent volume of the vehicle.

Effects of the test item were evaluated using clinical parameters (mortality, general observation, body weight, food and water consumption), clinical pathology (haematology, clinical chemistry) and full postmortem examination including necropsy, organ weight analysis and microscopic examination.

Adverse findings were noted on the skin both macroscopically and microscopically in both genders starting at the LD, i.e. the LOAEL was the LD.

Results after replacing CCGs with VCGs (study BD2-P)

The replacement of CCGs with VCGs resulted in the following differences regarding the statistical significance of quantitative parameters which were not identified as noteworthy in the original study.

An overview of discrepancies between findings observed with CCG and VCG is provided in Table BD2-P1.

Table BD2-P1: Discrepancies of clinical chemistry findings between CCGs and VCGs for BD2-P. CCG: concurrent control group, VCG: virtual control group, M: males, F: females, HCD: historical control data. LD: low dose, MD: mid dose, HD: high dose.

Parameter	Statistical significance in original study	Statistical significance with VCG	Assessment
Ca	LD_F, MD_F (+)	LD_F, MD_F, LD_M, MD_M, HD_M (+)	The increases are close to ULN, but show a clear exposure-relationship (exposure was not dose-linear)
Cholesterol	LD_M, HD_M (-)	LD_M (-)	Above LLN
Eosinophils	MD_F (-)	--	Above LLN
Hemoglobin	--	HD_M (-)	Above LLN
MCHC	--	LD_M, MD_M, HD_M (-)	Above LLN
MCV	--	LD_M, MD_M, HD_M, LD_F (+)	Above ULN
Liver weight (rel.)	--	MD_M (+)	Below ULN & no dose-dependency
Liver weight (abs.)	--	MD_M (+)	No dose-dependency (small change)
Phosphor	MD-M, MD-F (+)	--	Below ULN
Erythrocytes	--	MD_M, HD_M (-)	Above LLN
Sodium	MD_M, HD_M, MD_F, HD_F (+)		Below ULN (small changes)
Thrombocytes	--	HD_F (-)	Above LLN (small change)
Triglycerides	MD_F (-)	--	Above LLN

An overview of noteworthy findings comparing CCG and VCG is provided in Table BD2-P2.

Table BD2-P2: Comparative summary of noteworthy findings for study BD2-P (CCGs vs. VCGs). CCG: concurrent control group, VCG: virtual control group, M: males, F: females, HCD: historical control data. LD: low dose, MD: mid dose, HD: high dose.

Mortality
None
Clinical findings

Skin lesions (scissures, wounds, scab formation) M/F LD/MD							
Quantitative parameters							
Original noteworthy findings using the 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**
Water intake	(+)	F	MD	Inconsistent	--	--	--
Leucocytes	(+)	M	HD	Inconsistent	--	--	--
Lymphocytes	(+)	M	HD	Inconsistent	--	--	--
Glucose (whole blood)	(-)[1]	F	HD	Inconsistent	(+)	M/F	LD
Organ Weights							
Parameter name	**Increase (+) decrease (-)**	**Sex (M/F)**	**Starting dose**	**Compared to CCG**	**Increase (+) decrease (-)**	**Sex (M/F)**	**Starting dose**
Spleen (rel.)	(-)	F	LD	Partially consistent	(-)	M/F	LD
Spleen (abs.)	(-)	F	LD	Partially consistent	(-)	M/F	LD
Thymus (rel.)	(-)	F	MD	Inconsistent	--	--	--
Thymus(abs.)	(-)	F	MD	Inconsistent	--	--	--
Noteworthy necropsy findings							
Skin:sores M/F LD							
Noteworthy histopathology findings							
Organ/tissue	**Finding**	**Sex (M/F)**	**Starting dose**	**Background incidences in CCG**	**Background incidences in VCG**		
Skin	• Erosion/ulceration	M/F	LD	0/12	Crust: 0.38%		
	• Inflammation	M/F	MD	0/12	<0.38%[2]		
	• Reactive hyperplasia/hyper-keratosis	M/F	LD	0/12	< 0.38%[2]		
	• Bacteria	M/F	LD	0/12	< 0.38%[2]		

[1] The glucose values for CCGs were above ULN.

[2] No such finding was reported in the pool of 262 HCD animals, i.e. <1/262=<0.38%.

The replacement of CCGs with VCGs resulted in two potentially new noteworthy finding:

- Increases in Calcium were already observed in the original study but not considered noteworthy. With VCGs, these increases became more evident. Considering the non-linear dose-exposure relationship, they are now considered noteworthy.
- The increase in MCV observed with VCGs is above ULN and could be considered noteworthy. However, since no other haematological parameter indicates anaemia, it is regarded as biologically irrelevant.

Some findings identified as noteworthy in the original reports could not be replicated:

- Significant increases of leucocytes and lymphocytes in HD male animals were not found with VCGs. The observed changes in the original study were below the ULN.

- The significant decrease of glucose reported for HD females in the original study were not replicated with VCG. Instead, a significant increase of glucose was observed already at LD for both genders. The CCG values for glucose were above ULN, which explains this inverse relationship.
- Absolute and relative thymus weight changes observed in the original study were not replicated with VCGs.

Noteworthy histopathology findings were not altered by the background incidences of the VCGs.

Replacing CCGs with VCGs had no effect on the threshold dose determination, i.e. the LOAEL was the LD similar to the original report.

Study BD2-Q

Study description and original results (study BD2-Q)

The test item was investigated for its cumulative toxicity by administration once daily through oral (gavage) route for a 4-week period to 10 male and 10 female rats [Crl:WI] per group with a suspension of the test item in PEG 400 / Cremophor RH 40 / Imwitor / 40/35/25 (v/v/v) +0.5% SDS as vehicle at three doses (LD, MD, HD) over a period of approximately 4 weeks (28 to 30 administrations) with an administration volume of 5 mL/kg. Animals underwent necropsy one day after the end of treatment. A control group of 10 males and 10 females was treated likewise with an equivalent volume of the vehicle.

Effects of the test item were evaluated using clinical parameters (mortality, general observation, ophthalmoscopy, body weight, food and water consumption), clinical pathology (haematology, clinical chemistry, urinalysis) and full postmortem examination including necropsy, organ weight analysis and microscopic examination.

The daily oral administration of the test item to male and female rats over a period of approx. 4 weeks revealed no test item-related in life findings up to HD. In histopathology mild lymphocytic apoptosis/necrosis of thymus was obvious in males and females at the MD and higher. These findings are often unspecific, stress-induced, and toxicologically not relevant. Therefore, the NOAEL was determined to be the HD; the NOEL was the LD.

Results after replacing CCGs with VCGs (study BD2-Q)

The replacement of CCGs with VCGs resulted in the following differences regarding the statistical significance of quantitative parameters, which were not considered noteworthy in the original report.

An overview of discrepancies between findings observed with CCG and VCG is provided in Table BD2-Q1.

Table BD2-Q1: Discrepancies of clinical chemistry findings between CCGs and VCGs for BD2-Q.

Parameter	Statistical significance in original study	Statistical significance with VCG	Assessment
Protein (serum)	MD_M (-)	LD_M, MD-M, HD_M (-) LD_F, MD_F, HD_F (-)	Effects were above LLN.
Sodium (serum)	--	LD_M, MD-M, HD_M (+) LD_F, MD_F, HD_F (+)	Effects were below ULN.

An overview of noteworthy findings comparing CCG and VCG is provided in Table BD2-Q2.

1049 Table BD2-Q2: Comparative summary of noteworthy findings for study BD2-Q (CCGs vs. VCGs).

Mortality							
None							
Clinical findings							
None							
Quantitative parameters							
Original noteworthy findings using the CCG				**Noteworthy findings after replacing CCG with VCGs**			
Water intake[1]	(+)	F	LD	Inconsistent	(+)	M	LD
Organ Weights							
Kidney[1]							
(abs.)	(+)	M	HD	Inconsistent	--	--	--
(rel.)	(+)	M	HD	Inconsistent	--	--	--
Noteworthy necropsy findings							
None							
Noteworthy histopathology findings							

Organ/tissue	**Finding**	**Sex (M/F)**	**Starting dose**	**Background incidences in CCG**	**Background incidences in VCG [%]**
Thymus	Lymphocytic apoptosis/ necrosis (grade 1-2)[3]	M/F	MD	5/20	<0.39%[4]

1050 [1] The slight increase in water intake in females of all dose groups was not considered as toxicologically relevant.
1051 [2] The kidney weight increase was with 10% small, within LLN, and had no histological correlate. They were
1052 therefore considered to be related to the pharmacological mode of action of the test item, not being adverse.
1053 [3] Finding was considered non-adverse.
1054 [4] No such finding was reported in the pool of 258 HCD animals, i.e. <1/258=<0.39%.

1055 The replacement of CCGs with VCGs resulted in two new statistically significantly changed findings,
1056 namely a small decrease in protein (serum) in all doses and both sexes and a small increase in sodium
1057 (serum) also for all dose groups and both sexes. Due to the small effect size of these changes, which
1058 were both within LON, the findings were considered to have no biological relevance.

1059 Replacing CCGs with VCGs had no effect on the threshold dose determination, i.e. NOAEL was still the
1060 HD and NOEL the LD as in the original report.

References:

Berry, D. A., Elashoff, M., Blotner, S. et al. (2017). Creating a synthetic control arm from previous clinical trials: Application to establishing early end points as indicators of overall survival in acute myeloid leukemia (AML). J Clin Oncol 35, Suppl, 7021. doi:10.1200/jco.2017.35.15_suppl.7021.

EMA (2010). Guideline on repeated dose toxicity. 18 March 2010 CPMP/SWP/1042/99 Rev 1 Corr* Committee for Human Medicinal Products (CHMP). https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-repeated-dose-toxicity-revision-1_en.pdf

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-paperestablishing-efficacy-based-single-arm-trials-submitted-pivotal-evidence-marketingauthorisation-application_en.pdf.

EPA (1998). Health Effects Test Guidelines OPPTS 870.3150 90-Day Oral Toxicity in Nonrodents. https://www.regulations.gov/document/EPA-HQ-OPPT-2009-0156-0011.

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

Gurjanov, A., Kreuchwig, A., Steger-Hartmann, T., Vaas, L.A.I. (2023) 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.

Gurjanov, A., Vieira-Vieira, C., Vienenkoetter, J., Vaas, L.A.I., Steger-Hartmann, T. (2024a). Replacing concurrent controls with virtual control groups in rat toxicity studies. Regulatory Toxicology and Pharmacology, 148, 105592. doi.org/10.1016/j.yrtph.2024.105592

Gurjanov, A., Vaas, L.A.I., Steger-Hartmann T. (2024b) 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.

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-conducthuman-clinical-trials-pharmaceuticals-scientific.

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). doi.org/10.1787/9789264070684-en.

Kluxen, F. Weber, K., Strupp, C., Jensen, S.M., Hothorn, L.A., Garcin, J.-C., Hofmann, T. (2021). 'Using historical control data in bioassays for regulatory toxicology'. *Regulatory Toxicology and Pharmacology*, 2021;125, 105024. Available at: https://doi.org/10.1016/j.yrtph.2021.105024.

Maziasz, T., Kadambi, V.J., Silverman, L., Fedyk, E., Alden, C. (2010). Predictive toxicology approaches for small molecule oncology drugs. Toxicologic Pathology 38(1), 148-164. https://doi.org/10.1177/0192623309356448

National Cancer Institute (2025): Enterprise Vocabulary Services: https://evs.nci.nih.gov/evs-download/thesaurus-downloads.

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, doi.org/10.1787/9789264070721-en.

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, doi.org/10.1787/9789264070684-en.

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.

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. doi.org/10.1177/01926233241309905.

Robles-Diaz, M., Garcia-Cortes, M., Medina-Caliz, I., Gonzalez-Jimenez, A., Gonzalez-Grande, R., Navarro, J.M., et al. (2015). The value of serum aspartate aminotransferase and gamma-glutamyl transpeptidase as biomarkers in hepatotoxicity. *Liver International* 35(11)**,** 2474-2482. https://doi.org/10.1111/liv.12834.

Sanz, F., Pognan, F, Steger-Hartmann, T., Diaz, C. Et al. (2023). eTRANSAFE: data science to empower translational safety assessment. Nature Reviews Drug Discovery, 22: 605-606. https://doi.org/10.1038/d41573-023-00099-5.

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. doi.org/10.33611/trs.2023-009.

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

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. doi.org/10.14573/altex.2001311.

Trost, D.C. (2014). Hepatotoxicity, in: A. Lawrence Gould, *Statistical Methods for Evaluating Safety in Medical Product Developmen*t. Wiley VCH, Hoboken, New Jersey, pp. 229-270. https://doi.org/10.1002/9781118763070.ch9.