CHMP Scientific Advice

Document type: Summary Qualification Document

Active substance name: not applicable – Applicant name: Synapse-Managers

Proposed indication: not applicable

Procedure number: EMA/SA/0000122652

Official address Domenico Scarlattilaan 6 ● 1083 HS Amsterdam ● The Netherlands
Address for visits and deliveries Refer to www.ema.europa.eu/how-to-find-us
Send us a question Go to www.ema.europa.eu/contact **Telephone** +31 (0)88 781 6000

List of Tables

List of Figures

List of Abbreviations

ALB	Albumin
ALBGLOB	Albumin to Globulin ratio
ALT	Alanine Aminotransferase
AST	Aspartate Aminotransferase
ALP	Alkaline Phosphatase
APTT	Activated Partial Thromboplastin Time
BASO	Basophil Count
BICARB	Bicarbonate
BID	*bis in die* (twice a day)
BW	Body Weight
CA	Calcium
CDISC-SEND	Clinical Data Interchange Standards Consortium – Standard for Exchange of Nonclinical Data
CCG	Concurrent Control Group
CHOL	Cholesterol
CK	Creatine Kinase
CL	Chloride
CoU	Context of Use
DILI	Drug Induced Liver Injury
DRF	Dose Range Finding (study)
EOS	Eosinophil Count
Fibrino	Fibrinogen
GGT	Gamma-Glutamyl Transferase
GLDH	Glutamate Dehydrogenase
GLOBUL	Globulin
GLP	Good Laboratory Practice
GLUC	Glucose
HCD	Historical Control Data
HCT	Haematocrit
HD	High Dose
HD F	High Dose Female
HD M	High Dose male
HNSTD	Highest non-severely toxic dose
K	Potassium
IHI	Innovative Health Initiative
IMI	Innovative Medicines Initiative
i.v.	intravenous(ly)
LD	Low Dose
LD F	Low Dose Female
LD M	Low Dose Male
LLN	Lower Limit of Normal
LOEL	Lowest Observed Effect Level
LOAEL	Lowest Observed Adverse Effect Level
LON	Limit(s) of Normal
LYM	Lymphocyte Count
MCH	Mean Corpuscular Hemoglobin
MCHC	Mean Corpuscular Hemoglobin Concentration
MCV	Mean Corpuscular Volume

MD	Mid Dose
MD F	Mid Dose Female
MD M	Mid Dose Male
MTD	Maximum Tolerated Dose
NEUT	Neutrophil Count
N/A	not applicable
NHP	Non-Human Primate
NOEL	No Observed Effect Level
NOAEL	No Observed Adverse Effect Level
PHOS	Phosphorous
PLAT	Thrombocyte Count
PROT	Protein
PT	Prothrombin Time
QD	*quaque die* (once a day)
QT	Qualification Team
RETI	Reticulocyte Count
RBC	Red Blood Cell Count
RDW	Red Cell Distribution Width
SME	Subject Matter Expert
STD10	Severely Toxic Dose in 10% of animals
SOP	Standard Operating Procedure
THROMNUC	Thrombocyte Count
TRIG	Triglycerides
U_	Urinary
U_CLEXR	Urinary Chloride Excretion Rate
U_CREAT	Urinary Creatinine
U_KEXR	Urinary Potassium Excretion Rate
U_PH	Urinary pH
U_PROTCRT	Urinary Protein/Creatinine Rate
U_SODMEXR	Urinary Sodium Excretion Rate
U_SPGRAV	Urinary Specific Gravity
U_Volume	Urinary Volume
ULN	Upper Limit of Normal
UREAN	Urinary Nitrogen
VCG	Virtual Control Group
WBC	White Blood Count

1. Introduction

The concept of Virtual Control Groups (VCG) was developed within the framework of the IMI consortium eTRANSAFE (https://etransafe.eu/; Steger-Hartmann et al., 2020). It describes the generation of VCGs from well-curated historical control data (HCD) after applying study-specific criteria, termed "matching criteria". The application of VCGs is intended to partially reduce the number Concurrent Control Group animals (CCG) or entirely replace CCGs used in systemic toxicity studies, hereby reducing the animal numbers in preclinical safety testing.

In 2023, the development of a prototype database for VCGs started from shared control groups data from legacy systemic toxicity studies for the most common animal species by several pharmaceutical companies participating in IMI eTRANSAFE. Statistical procedures for characterizing the control data and methods for identifying the optimal matching with the animals assigned to treatment groups have been developed since. The database, the statistical procedures and the matching methods were used for a qualification process, where several legacy toxicity studies were reanalysed upon replacement of CCGs with VCGs.

For the qualification of VCGs, three steps for the implementation of the VCG concept were identified, which are assigned to three contexts of use:

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

In February 2023, Scientific Advice from EMA for the CoU I was requested for those species most frequently used in preclinical safety (rats, mice, dogs, NHPs, and mini-pigs). The request resulted in the submission of Briefing Documents and included replies to the List of Issues. During a discussion meeting held on November 25th, 2024, the EMA Qualification Team (EMA 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, EMA QT informed VICT3R about the decision, that "an amended Context of Use 1 (VCG application in non-GLP DRF studies) could be qualified, however limited to studies in rats". EMA QT requested that the submitted documents be summarized in a Summary Qualification Document for public consultation.

The Summary Qualification Document refers to the following information previously submitted to the EMA QT:

- 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 further species (3 mice studies, 9 dog studies, 6 NHPs studies, 2 mini-pig studies) this material was not included in this current Summary Qualification Document.

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

1.1. Rationale for seeking advice

Advice was requested on 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.

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. Thus, each individual study reanalysis aimed to determine whether the replacement of CCGs with VCGs would alter the study overall conclusions and whether the performance of the proposed VCG approaches can be considered robust and reliable.

The rationale for seeking advice was to obtain feedback on the VCG concept, its implementation and the acceptability within a regulatory context.

2. VCG qualification procedure for CoU I

2.1. HCD collection

The first step in the compilation of a suitable HCD pool is the identification of the animal subpopulation to be matched. The data should be collected and curated according to the criteria describing in the developed SOP (see Annex). It should be noted that the database of control animals where the HCD is collected ("the VICT3R database") will require continuous updates with new control animal data to account for trends or shift in the various parameters e.g., genetic drift or to include new vehicles, new dosing schedules, etc. For the studies reanalysed in this document, a time window of 5 years was used. However, statistical analysis may demonstrate that periods of different length may also be acceptable.

The requirement for continuous updates of the database implies that CCGs will not be replaceable in all future 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.

2.2. Description of the VCG generation and legacy study reanalysis procedures

2.2.1. Short overview of the SOP

An SOP for the generation of VCGs was developed during the qualification procedure (see Annex). This SOP summarizes the current knowledge and experience accumulated by the applicant regarding HCD collection and the generation of VCGs, and instruction for documenting the procedures.

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

The matching criteria used in the SOP represent a subset of the criteria established by other authors (Palazzi *et al.*, 2024; Sato *et al.*, 2024) and may not represent a complete set of all factors that are controlled during a study. The relevance of some of these factors is yet unknown, while others are rarely recorded in the data collection (i.e. bioanalytical assay methods).

It is proposed that in cases of uncertainty regarding the matching between VCGs and the treatment animals, statistical analysis and visualization 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 beneficial to relax some matching criteria (such as using an HCD pool older than 5 years), 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 for details).

Examples of such statistical characterization of the HCD performed prior to the study reanalysis are described in this Summary Qualification Document. This included a general assessment of the HCD pool to identify possible trends in the data that arise from combining studies. The purpose of this assessment was to identify possible clustering of animals into subpopulations and reveal patterns that may not be immediately apparent in the HCD pool (Wolford *et al.*, 1987; Weber *et al.*, 2011; de Kort *et al.*, 2020; Gurjanov *et al.*, 2023).

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 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 for CoU I are usually non-GLP studies. The SOP created here is nevertheless considered as a guidance for these DRF studies while taking into account the necessary flexibility in the design of these non-GLP studies.

2.2.2. Selection of legacy studies for the qualification procedure

The current procedure intends to qualify the application of VCG in rat DRF studies (CoU I). Despite this focus on DRF studies, several reanalysed 4-week toxicity rat studies were included, some of which were originally performed under GLP.

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

- **Data availability**. Compared to DRF studies, 4-week 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 growth 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 4-week toxicity studies. If the evaluation of reanalysed 4-week 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).

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 underlies a high degree of variability. In consequence, the reanalysis of these studies using VCGs will require a reflection of this heterogeneity. Therefore, we did not implement a uniform procedure when reanalysing DRF rat studies but rather to allow for individual approaches that best reflect the procedures implemented at different test sites. For example, DRF studies may differ in the number of animals per experimental group.

When control and treated groups have less than three animals per gender, statistical analysis cannot be conducted, and data can only be visually inspected for trends. However, in case larger group sizes were used, both statistical analysis and visual inspection are performed. In both cases, the biological and toxicological evaluations of the results take 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, developing a strategy for qualifying the use of VCGs requires some level of backward engineering of the process by which a study director interprets the results. There is no generalized workflow for this procedure and, therefore 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.

A straightforward proposal is to provide the results of a legacy study where the CCG was replaced with VCG to the study director and have him or her assess the results without background information on the original study. However, this would ignore the 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.

With that in mind, we developed a mixed approach for reanalysis of legacy studies after replacing CCGs with the same number of VCG animals collected by applying study specific matching criteria. Dose groups and the VCG were subjected to identical statistical procedures as originally applied to the CCG data for all quantitative parameters. Numeric results of the legacy study were compared with the ones for VCGs by listing those parameters with identical behaviour in terms of significance and direction of the change (increase or decrease). These lists were then given to the study director or another experienced toxicologist (i.e., a subject matter expert; SME) for assessment of treatment relatedness of findings. To document the decisions, pull-down fields were offered with arguments against “treatment-relatedness”. For example, the expert was given to indicate whether the finding

had "no dose dependency", was "within 2-sigma range of historical control data" or had "different deviation in different sexes". 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.

Regarding the semi-quantitative or qualitative parameters not subjected to statistical analyses during routine studies (clinical observations, ophthalmology, gross pathology, histopathology), historical control data (HCD) collected for the same species, strain and test facility over the last five year was used to calculate background incidences for a specific finding. Alternatively, analysis of incidence for studies using Sprague Dawley rats included rat studies performed between 2000 and 2021.

The described procedure is depicted in Figure 1 and has been published by Golden et al. (Golden et al. 2023).

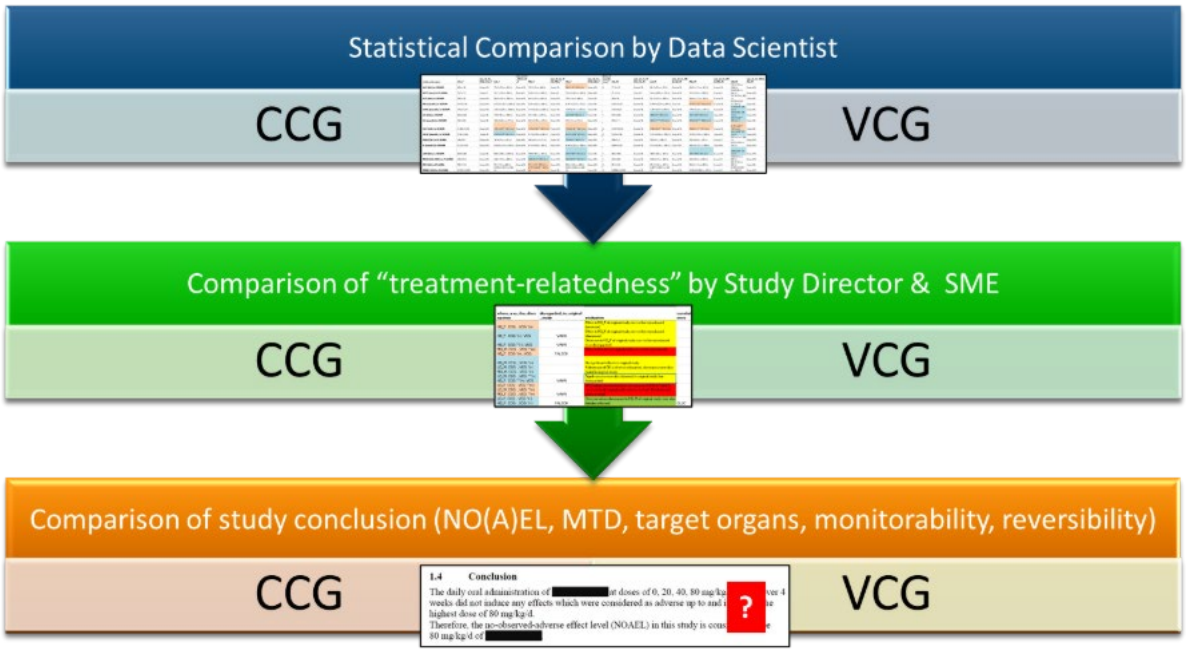


Figure 1: Description of the qualification procedure based on assessment of legacy 4-week toxicity studies, where the CCGs were replaced by VCGs. *In a first step the statistical outcome of the two studies is compared and the parameters, which show different behavior (additional statistical significances detected or lost) are marked in the table (heatmap). Subsequently, the statistical differences are evaluated by a study director of 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 outcome of the study regarding NO(A)EL (or other dose thresholds), target organs, and other key findings.*

The above depicted procedure has been applied to a total number of eight rat legacy studies. The applied protocols, the tools and methods (e.g. programming language) for the qualification procedure, i.e. the creation of the VCG from the HCD were not completely identical between the reanalysed studies. Table 1 summarizes the differences and commonalities used for the eight studies.

Table 1: *Overview of reanalysed studies including study design and further study details. F and M correspond to female and male; LD, MD and HD correspond respectively to low, mid and high dose. When only on sex was included, the suffix "_F" and "_M" were added, to the dose (e.g. HD_F).*

Study Number	Study Label	Animal Provider	Species	Strain	Age animals CCG	Dose duration (dosing period)	Recovery period duration	CCG group size	Dose groups	Route of administration	Dosing frequency	Vehicle	Housing	Study year	GLP (Y/N)	Study included CCG (Y/N)
1	BD1-A	Charles River, Sulzfeld, Germany	Rat	Crl:Wistar	7-9 weeks	4 weeks	N/A	10 M/10 F	LD, MD, HD	Oral gavage	QD	Ethanol/Kolliphor HS15/Water for Injection	Group Housed	2021	Y	Y
2	BD1-B	Charles River, Sulzfeld, Germany	Rat	Crl:Wistar	7-9 weeks	4 weeks + 2 weeks recovery	Y	10 M/10 F+ 6 M/6 F recovery	LD, MD, HD	Oral gavage	QD	Ethanol/Kolliphor HS15/Water for Injection	Group Housed	2022	Y	Y
3	BD1-C	Charles River, Sulzfeld, Germany	Rat	Crl:Wistar	7-11 weeks	4 weeks	N/A	10 M/10 F	LD_M, MD_M, HD_M LD_F, MD_F, (HD F)[1]	Oral gavage	QD	Ethanol/Kolliphor HS15/Water for Injection	Group Housed	2021	Y	Y
4	BD1-D	Charles River, Sulzfeld, Germany	Rat	Crl:Wistar	7-8 weeks	2 weeks (DRF)	N/A	6 F	LD_F, HD_F	Oral gavage	QD	Ethanol/Kolliphor HS15/Water for Injection	Group Housed	2020	N	Y
5	BD2-O	Charles River, Sulzfeld, Germany	Rat	Crl:Wistar	6-8 weeks	4 weeks	N/A	10M/10 F	LD, MD, HD	Oral gavage	QD	Ethanol/Kolliphor HS15/Water for Injection	Group Housed	2023	Y	Y
6	BD2-P	Charles River, Sulzfeld, Germany	Rat	Crl:Wistar	7-8 weeks	4 weeks	N/A	6M/6F	LD, MD, HD	Oral gavage	BID	Ethanol/Kolliphor HS 15/Water for injection	Group Housed	2022	N	Y

7	BD2-Q	Charles River, Sulzfeld, Germany	Rat	Crl:Wistar	7-8 weeks	4 weeks	N/A	10M/10F	LD, MD, HD	Oral gavage	QD	PEG 400 / Cremophor RH 40 / Imwitor / 40/35/25 (v/v/v) +0.5% SDS	Group Housed	2020	Y	Y
8	H	Envigo RMS, Inc., Indianapolis, Indiana	Rat	Hsd:Sprague Dawley	8 to 10 weeks	4 weeks	N/A	15M/15F	LD, MD, HD	Oral gavage	BID	0.5% (w/v) methylcellulose (MC; 4000 cps) and 0.1% (v/v) Polysorbate 80 (Tween 80) in reverse osmosis water	Group Housed	2019	Y	Y

[1] HD_F was prematurely sacrificed due to moribund status.

2.2.3. Matching procedures used in reanalysed legacy studies

The matching procedures are summarized in Table 2 below. The matching criteria used are part of the list of criteria described in chapter 6.1 of the SOP. A few additional criteria such as identical diet and humidity conditions were also applied in some studies.

Table 2: *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.	Treatmt. 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

Study Number	Study Label	Test Facility	Species	Strain	Sex	Animal provider	Housing	Route of Admin.	Treatmt. schedule	Dosing period	Recovery period duration	Vehicle	HCD period	Age of VCG animals	Initial BW	Other matching criteria	VCG size	VCG iteration
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

2.2.4. Procedures for comparing study results obtained using CCG and VCG

The general approach for the reanalysis of a study after replacing the CCG with 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 whole procedure is depicted in Figure 2 (from Gurjanov *et al.*, 2024a).

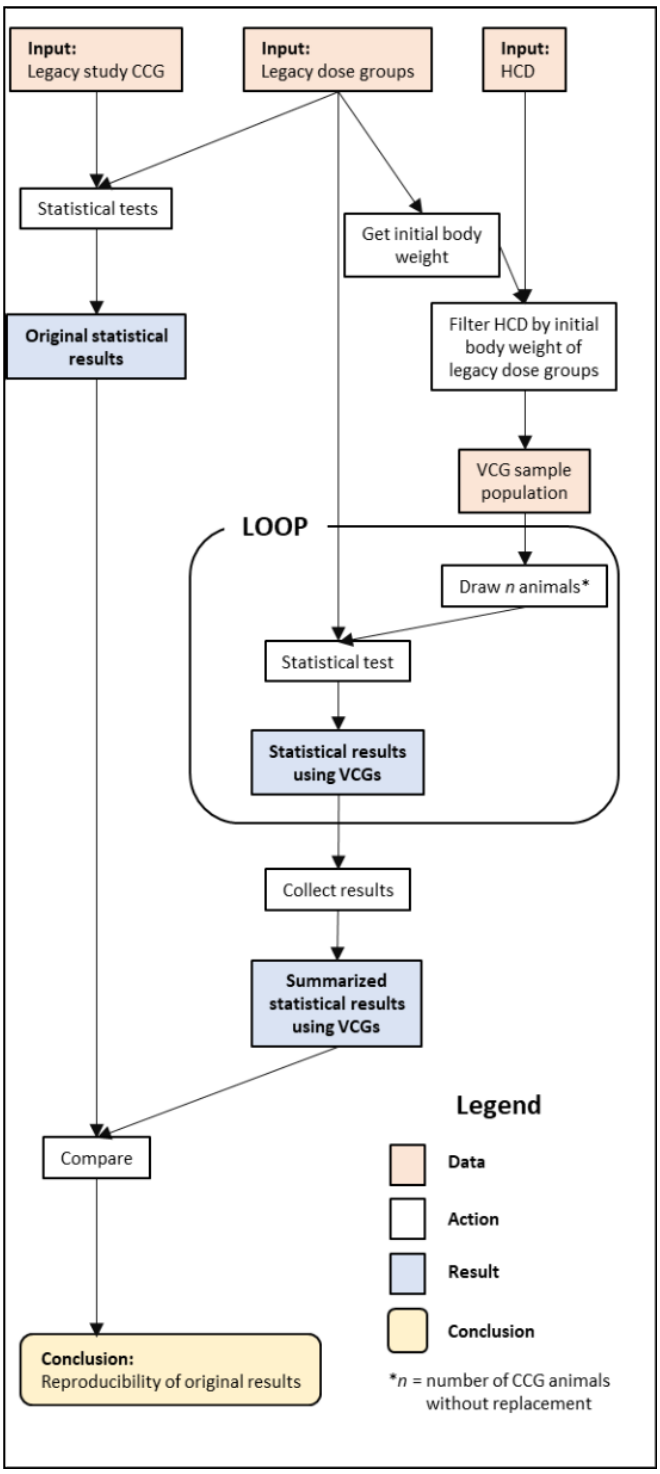
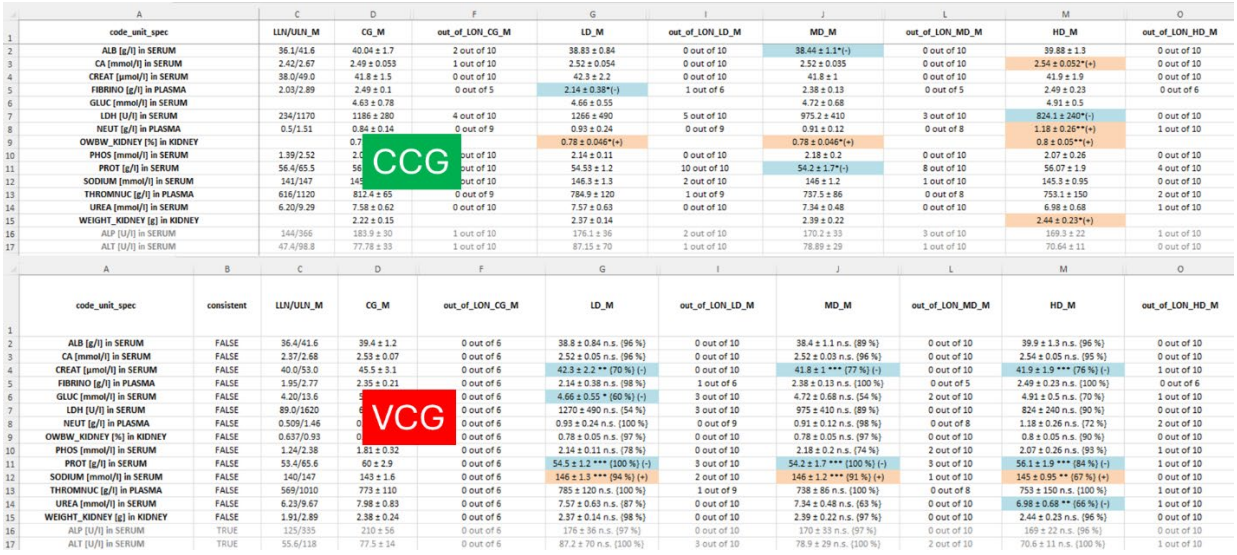


Figure 2*: *Visual description for the comparison of statistical results using CCGs and VCGs. *HCD was used as reference data for assessing test substance-relatedness based on limits of normal and for calculation of background incidences for qualitative findings (clinical observations, histo-pathology).*

After replacing the CCG with VCG in the data set of legacy studies under reanalysis, treatment groups were statistically compared with the new control values. For that, an R-script was developed 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, 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 our 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, we reported quantitative results 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 3). 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	5	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	6	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	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	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 3*: 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 his or her judgment with the original outcome and commented the observed differences (see Figure 4). 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 4*: Screenshots of the csv files comparing and assessing the treatment-relatedness between VCG and the original results (example from study 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 5 (from Gurjanov *et al*., 2024a).

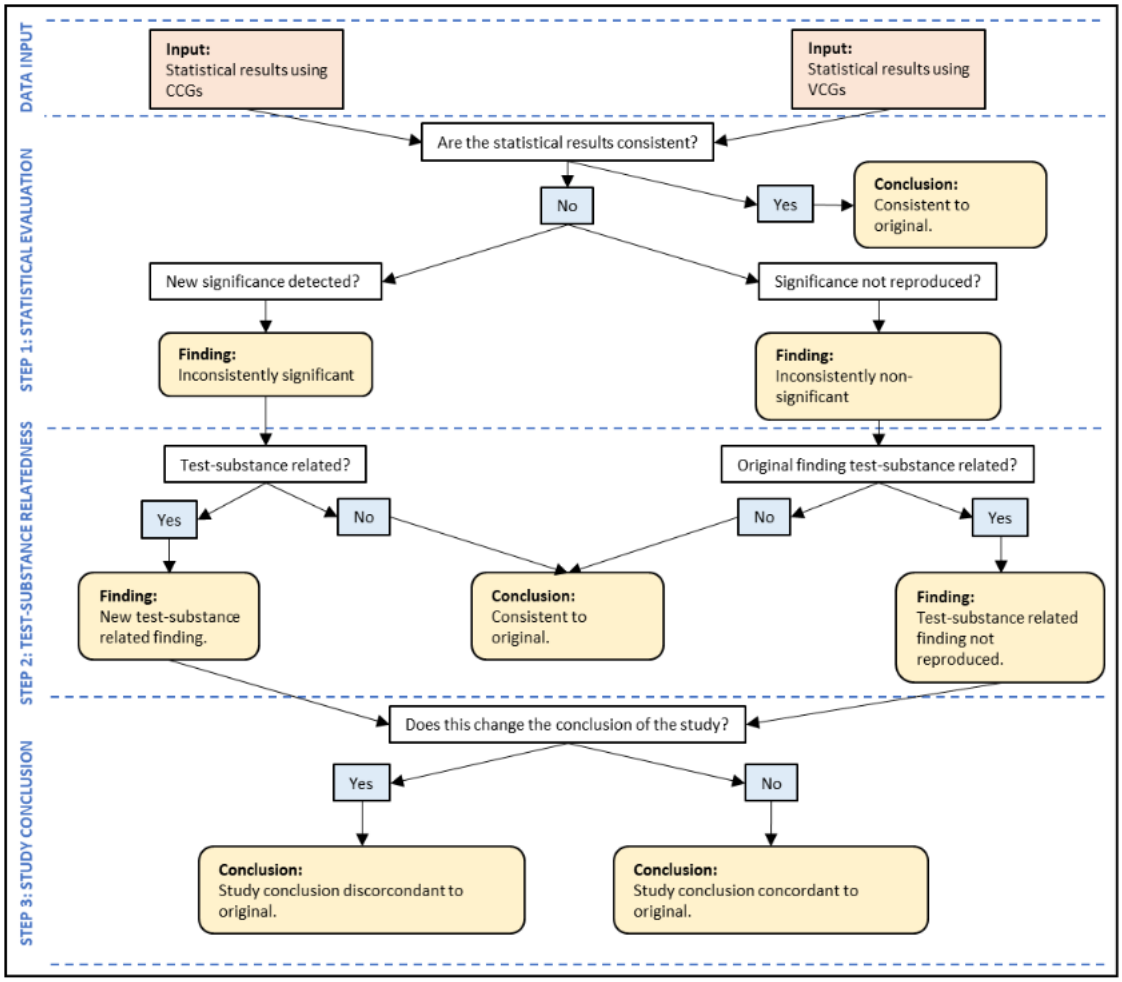


Figure 5: Workflow describing the performance assessment for study outcome using VCGs

3. Application of VCGs for CoU I – results of reanalysed studies

3.1. Results of study reanalysis BD1-A

3.1.1. Study description and original results

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.

3.1.2. Results 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 GGT 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.

A side-by-side comparison between noteworthy findings in both CCG and VCGs analysis for legacy study A are presented in Table 3.

Table 3: *Table of noteworthy findings in legacy study BD1-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	Inconsistent			
Glucose (whole blood)	-	M	HD	Partially consistent	-	M/F	HD/LD
Total bilirubin	+	F	LD	Inconsistent			
gamma-Glutamyl transferase				Inconsistent	+	M+F	LD
Absolute liver weight	+	F	MD	Consistent	+	F	MD
Relative liver weight	+	F	MD	Partially consistent	+	F	LD
Noteworthy pathological findings							
None							

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

3.2. Results of study reanalysis BD1-B

3.2.1. Study description and original results

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.

This DRF study investigated possible effects of a potential anti-cancer drug including 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).

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 tingible 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.

3.2.2. Results 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.

A side-by-side comparison between noteworthy findings in both CCG and VCGs analysis for legacy study B are presented in Table 4.

Table 4: *Table of noteworthy findings in legacy study BD1-B. 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 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	Partially consistent	+	M/F	LD/MD
Chloride	-	M	MD	Partially consistent	-	M	LD
Protein/ Creatinine ratio	+	M+F	MD	Partially consistent	+	M+F	LD
Protein	-	M	HD	Inconsistent			
Alanine aminotransferase				Inconsistent	+	F	MD
Basophils				Inconsistent	+	M+F	LD
gamma-Glutamyl transferase				Inconsistent	+	F	LD
Glucose (whole blood)				Inconsistent	+	M+F	LD

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 haematopoiesis	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 %)

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

3.3. Results of study reanalysis BD1-C

3.3.1. Study description and original results

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.

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 the 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.

3.3.2. Results 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).

3.4. Results of study reanalysis BD1-D

3.4.1. Study description and original results

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.

This study investigated the feasible doses for an oncology drug candidate to be applied in a subsequent 4-week toxicity study under GLP.

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

3.4.2. Results after replacing CCGs with VCGs

The effects observed at HD, 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.

3.5. Results of study reanalysis BD2-O

3.5.1. Study description and original results

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 (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 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.

3.5.2. Results after replacing CCGs with VCGs

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 5.

Table 5: *Discrepancies of clinical chemistry findings between CCGs and VCGs for study BD2-O. 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	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 6**.

Table 6*: Comparative summary of noteworthy findings for study BD2-O (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							
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 + Haematocrit	(-)	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) • Hypertrophy (minimal/slight) • Accumulation of lipid/centrilobular (minimal/slight)	M/F	HD	0/20 0/20 0/20	3.84% 0.76% 2.69%		
Kidney	• Degeneration/vacuol. collecting duct (minimal/moderate) • Dilatation, cortical tubule (minimal/slight)	M/F	HD	0/20 1/20	<0.38%[3] 1.92%		
Bone marrow (femur/sternum)	• Cellularity, increased: foamy macrophages (minimal/moderate) • Cellularity, increased: myelopoiesis (minimal)	M/F M	HD HD	0/20 0/20	0.38% 0.38%		

Spleen	• Cellularity, increased: foamy macrophage (minimal/moderate) • Cellularity, increased: germinal center (number/size; mostly minimal/slight) • Extramedullary haematopoiesis, 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.

3.6. Results of study reanalysis BD2-P

3.6.1. Study description and original results

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.

3.6.2. Results after replacing CCGs with VCGs

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 7.

Table 7: *Discrepancies of clinical chemistry findings between CCGs and VCGs for study 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 8.

Table 8: *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/FLD				
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 findings:

- 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 anemia, 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 like the original report.

3.7. Results of study reanalysis BD2-Q

3.7.1. Study description and original results

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.

3.7.2. Results after replacing CCGs with VCGs

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 9**.

Table 9: *Discrepancies of clinical chemistry findings between CCGs and VCGs for study 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 ***10***.

Table 10: *Comparative summary of noteworthy findings for study BD2-Q (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							
None							
Quantitative parameters							
Water intake[1]	(+)	F	LD	Inconsistent	(+)	M	LD
None							
Organ Weights							
Kidney[1] (abs.) (rel.)	 (+) (+)	 M M	 HD HD	 Inconsistent 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]

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

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

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

3.8. Further statistical characterization of VCGs used in studies BD2-O, BD2-P and BD2-Q

To further assess whether differences might exist between the chosen VCGs and the concurrent controls measurements conducted during in-life phase of studies in control animals were compared. This comparison provides additional information about whether the chosen animals belong to the same population. For instance, if, after combining control animals from multiple studies, one would observe a bi-modal distribution for one of the biologically relevant endpoints, this may indicate hidden confounders and would require a more thorough investigation. Similarly, if time control charts show a trend or a sudden drop of a certain endpoint, this may also indicate presence of a confounder and that the grouped animals may not belong to the same population.

Figure 6 and

Figure ***7*** show the distribution of initial body weights for all animals per study and over the course of the selected time period. Despite the selection of lighter, i.e. younger animals in some studies (e.g. BD2-013, male animals) there is no general trend regarding body weight within the chosen time period.

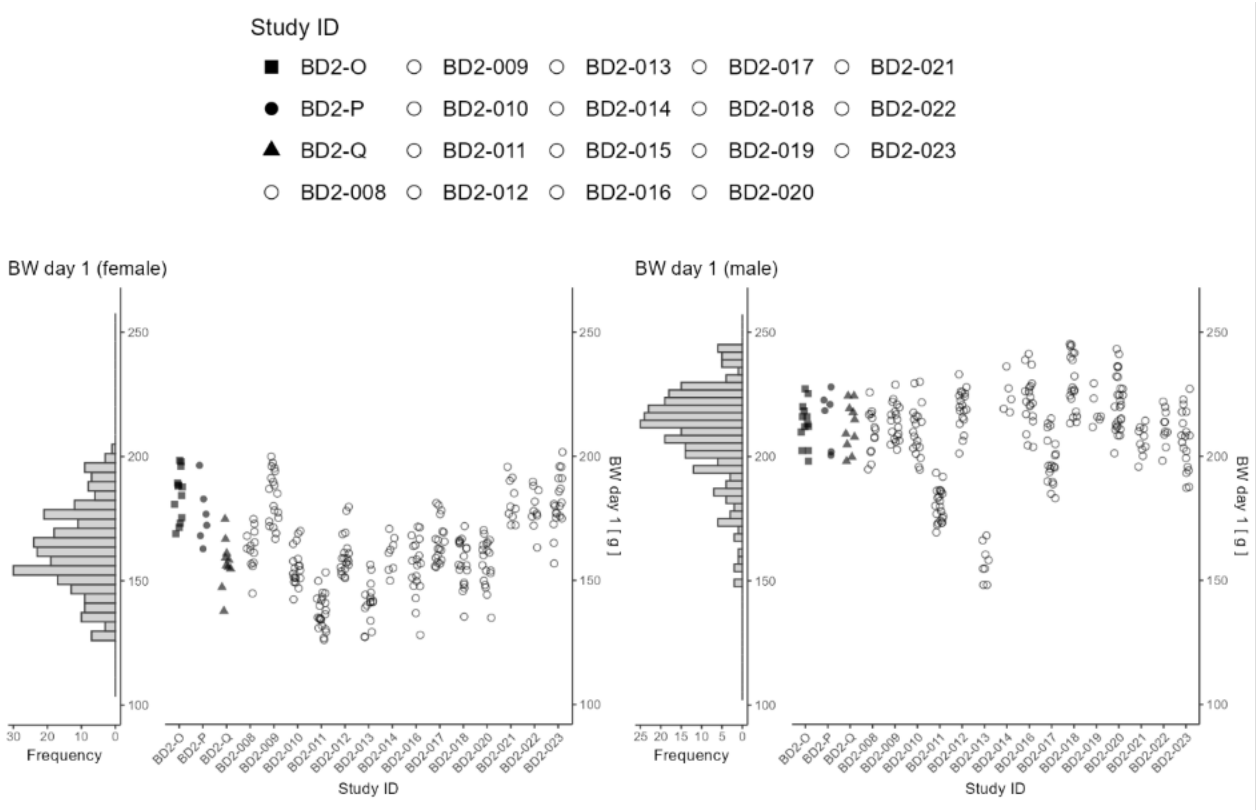


Figure 6: Body weight distribution of control animals across studies (rat). *Body weight of control group animals across BD2-O to BD2-Q rat studies and additional studies forming the HCD pool (displayed with open dots). Body weight of the rats was measured on day 1, before start of treatment, across all eligible studies.*

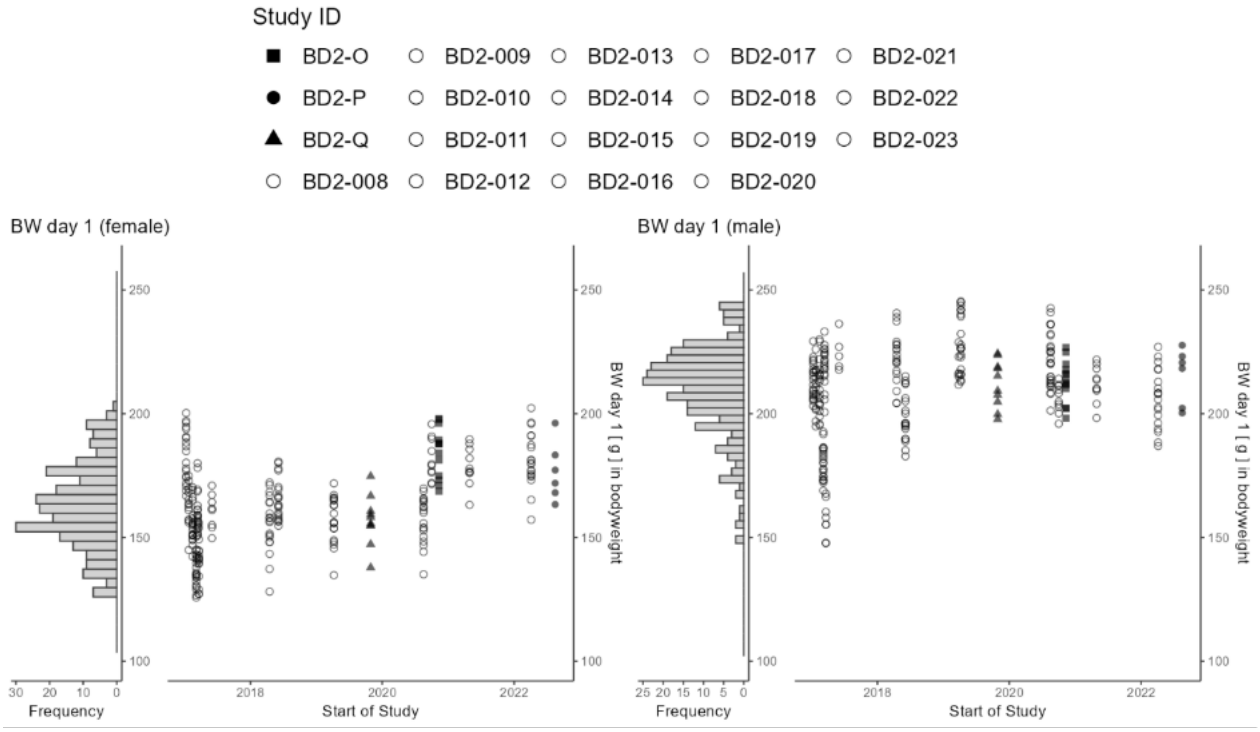
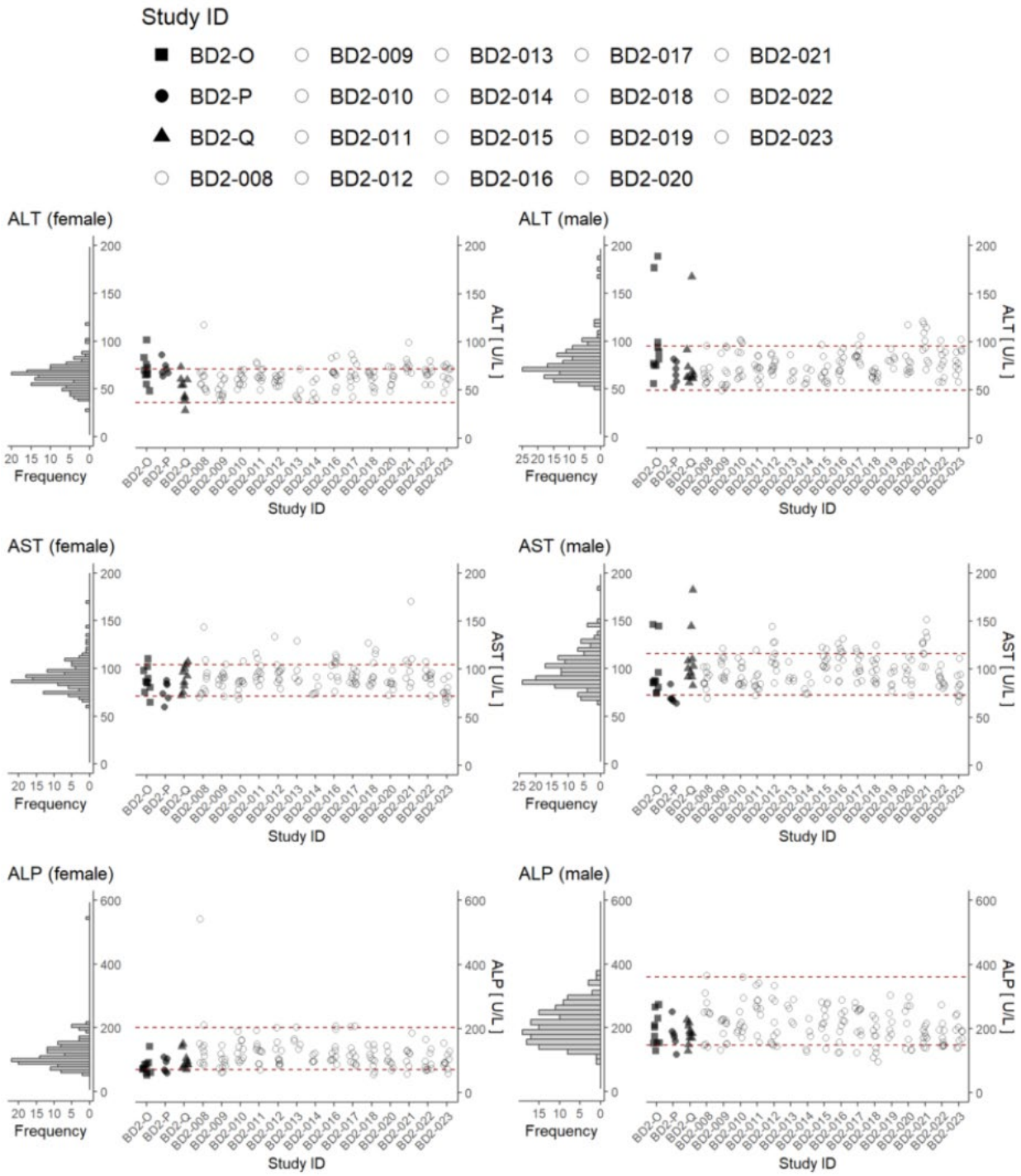


Figure 7: Body weight distribution of control animals across the chosen time span (rat).
Body weight of control group animals across BD2-O to BD2-Q rat studies and additional studies forming the HCD pool (displayed with open dots). Body weight of the rats was measured on day 1, before start of treatment, across all eligible studies.

In reflection of the developed SOP and the described the requirements to assess the relevance of matching criteria (see chapter 6.5 of the SOP) the HCD used for the generation of VCGs for reanalysis of the studies BD2-O to BD2-Q were characterized.

The following figures (Figure 8 to Figure 15) display the ranges of values for selected parameters across eligible studies and across the chosen period.

In general, the figures illustrate that there are no obvious trends over the chosen time period.


Study ID
BD2-O
BD2-P
BD2-Q
BD2-008
BD2-009
BD2-010
BD2-011
BD2-012
BD2-013
BD2-014
BD2-015
BD2-016
BD2-017
BD2-018
BD2-019
BD2-020
BD2-021
BD2-022
BD2-023
ALT (female)
ALT (male)
AST (female)
AST (male)
ALP (female)
ALP (male)
Frequency
Study ID
ALT [U/L]
AST [U/L]
ALP [U/L]

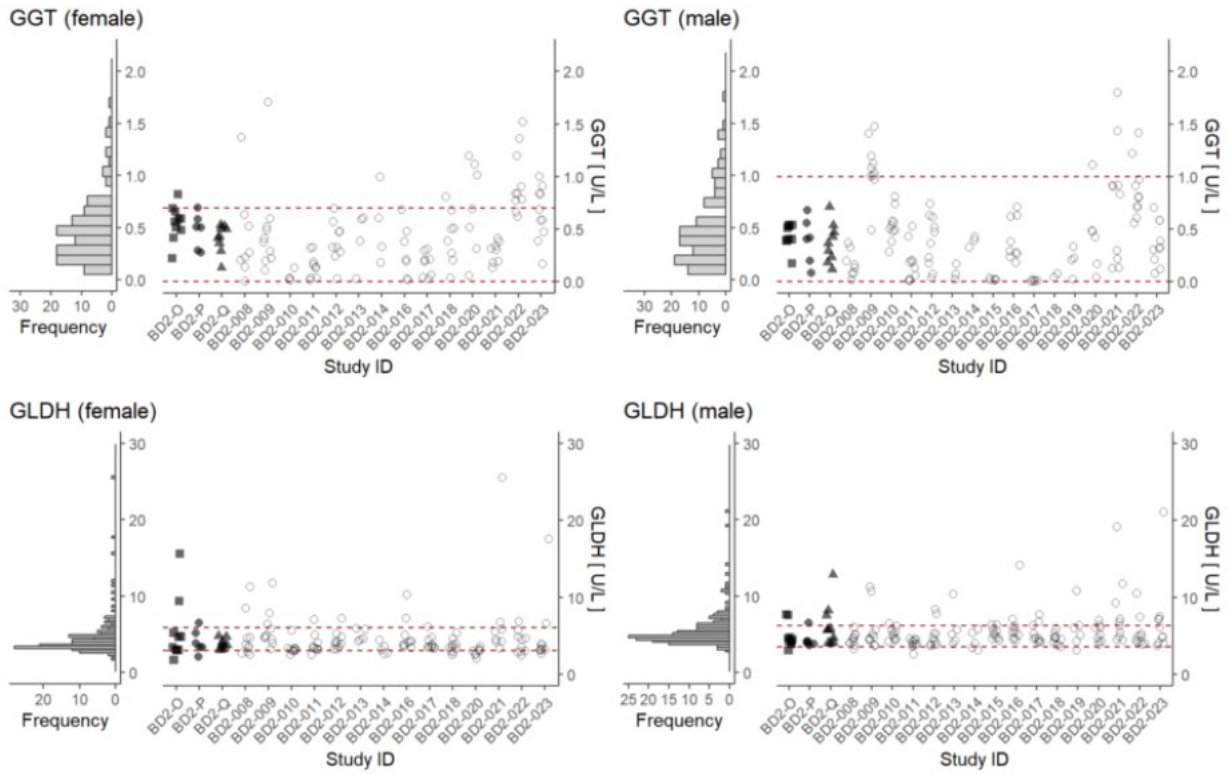
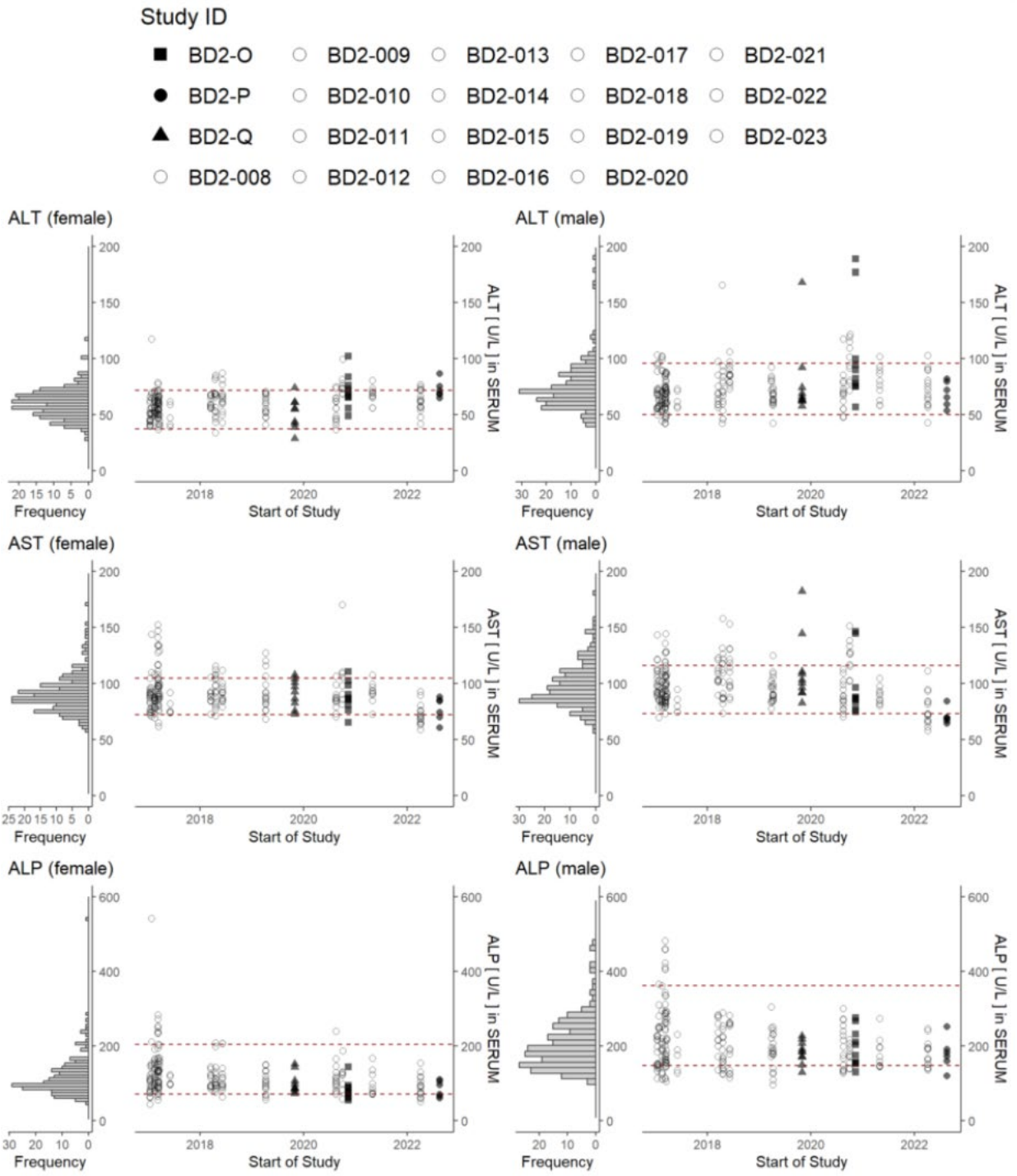


Figure 8: Distribution of liver enzyme values in control animals across eligible studies.
Female and male animals are represented on the left and the right side, respectively. Here the studies are NOT arranged by the start of study year. A single dot represents liver enzyme measurement for a single animal on day 1 of study, before the start of treatment. Red dashed lines indicate the historic reference values used in the original studies and represent the upper (95% percentile) and lower (5% percentile) limit of normal, respectively. Note that the historic reference values are NOT calculated from the data analysed here.


Study ID
BD2-O
BD2-009
BD2-013
BD2-017
BD2-021
BD2-P
BD2-010
BD2-014
BD2-018
BD2-022
BD2-Q
BD2-011
BD2-015
BD2-019
BD2-023
BD2-008
BD2-012
BD2-016
BD2-020
ALT (female)
ALT (male)
AST (female)
AST (male)
ALP (female)
ALP (male)
Frequency
Start of Study
ALT [U/L] in SERUM
AST [U/L] in SERUM
ALP [U/L] in SERUM

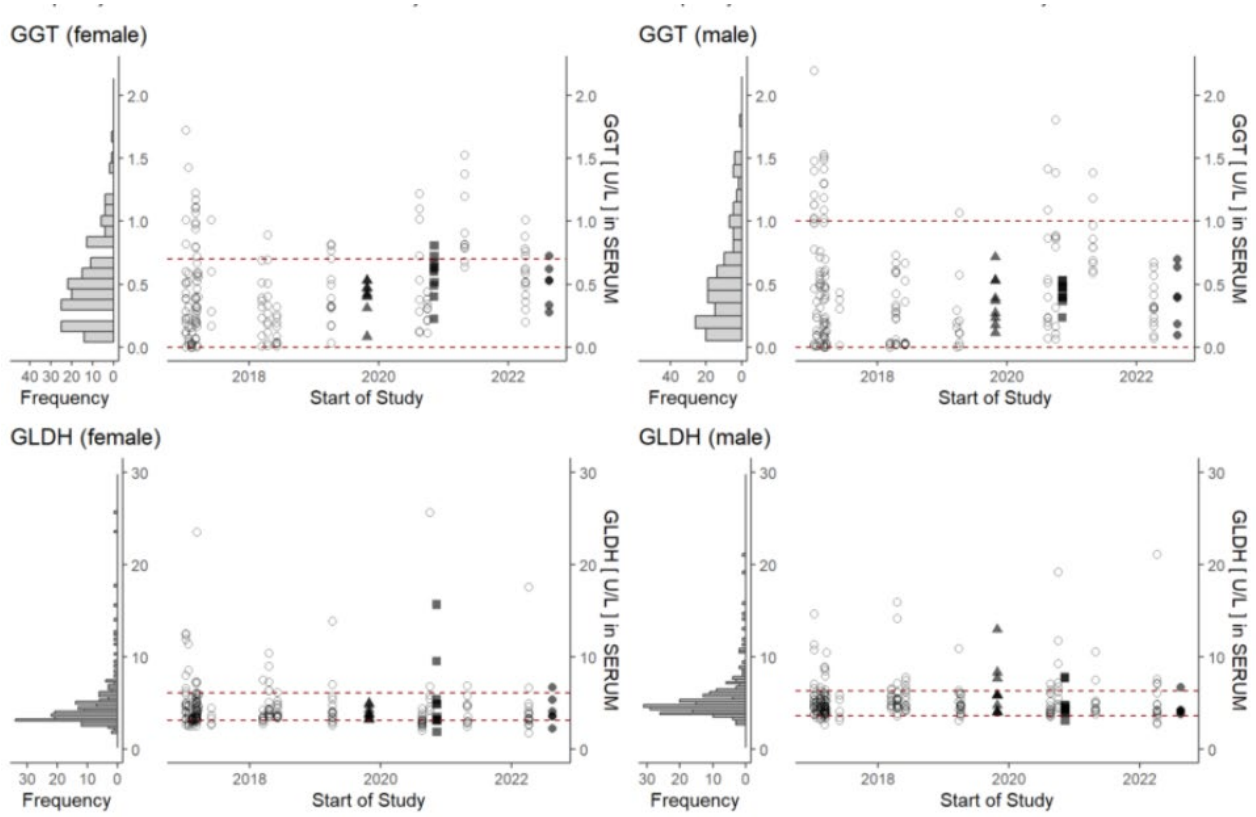


Figure 9: Distribution of liver enzyme values in control animals across chosen time period. *Female and male animals are represented on the left and the right side, respectively. In each figure, the measurements are arranged by the start of study year. A single dot represents liver enzyme measurement for a single animal on day 1 of study, before the start of treatment. Red dashed lines indicate the historic reference values used in the original studies and represent the upper (95% percentile) and lower (5% percentile) limit of normal, respectively. Note that the historic reference values are NOT calculated from the data analysed here.*

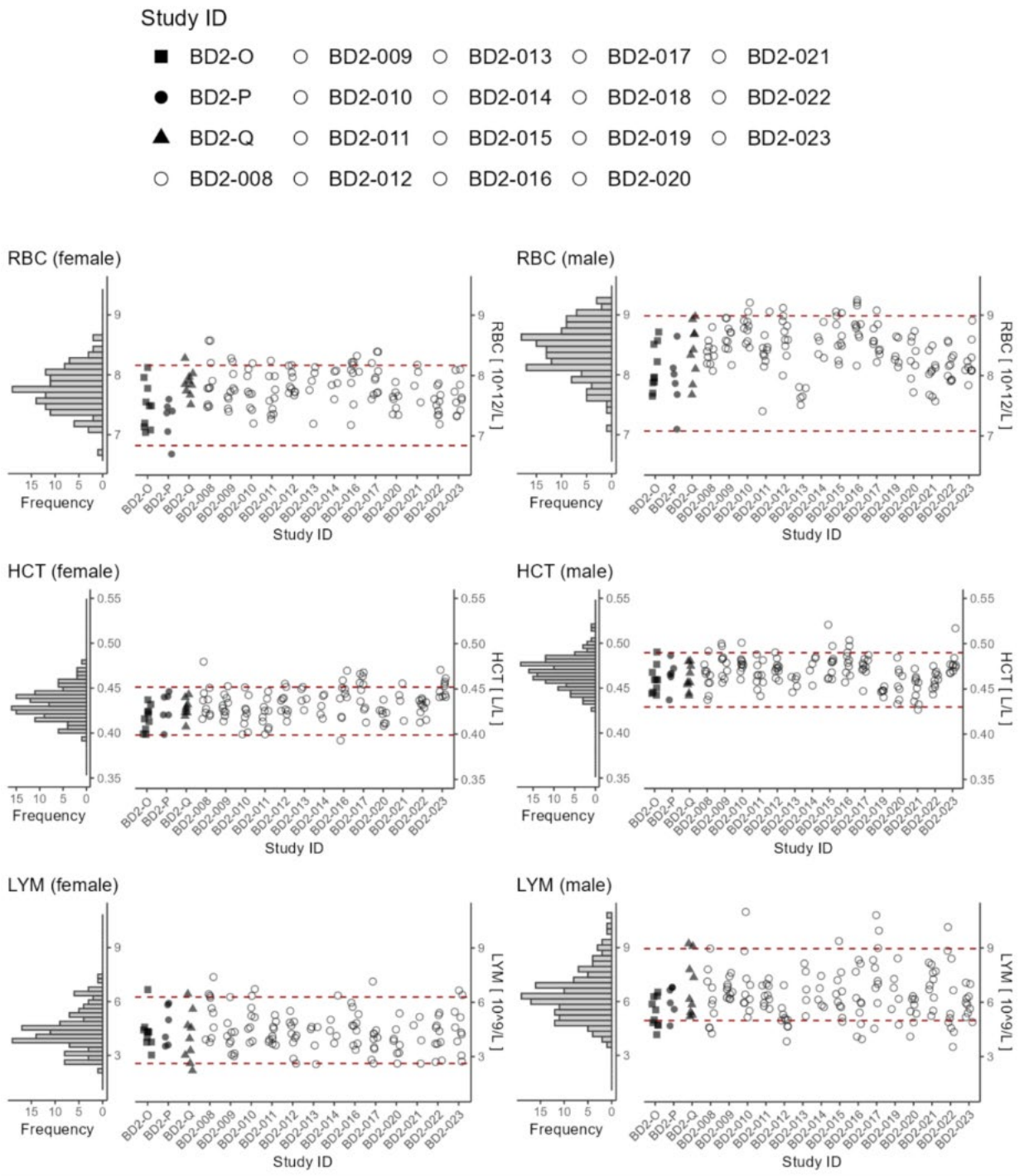


Figure 10: Distribution of RBC, HCT and LYM in control rats across studies. *Distribution of erythrocyte counts (RBC), haematocrit (HCT) and lymphocyte counts (LYM) in control animals across eligible studies. Female and male animals are represented on the left and the right side, respectively. Here the studies are NOT arranged by the start of study year. A single dot represents liver enzyme measurement for a single animal on day 1 of study, before the start of treatment. Red dashed lines indicate the historic reference values used in the original studies and represent the upper (95% percentile) and lower (5% percentile) limit of normal, respectively. Note that the historic reference values are NOT calculated from the data analysed here.*

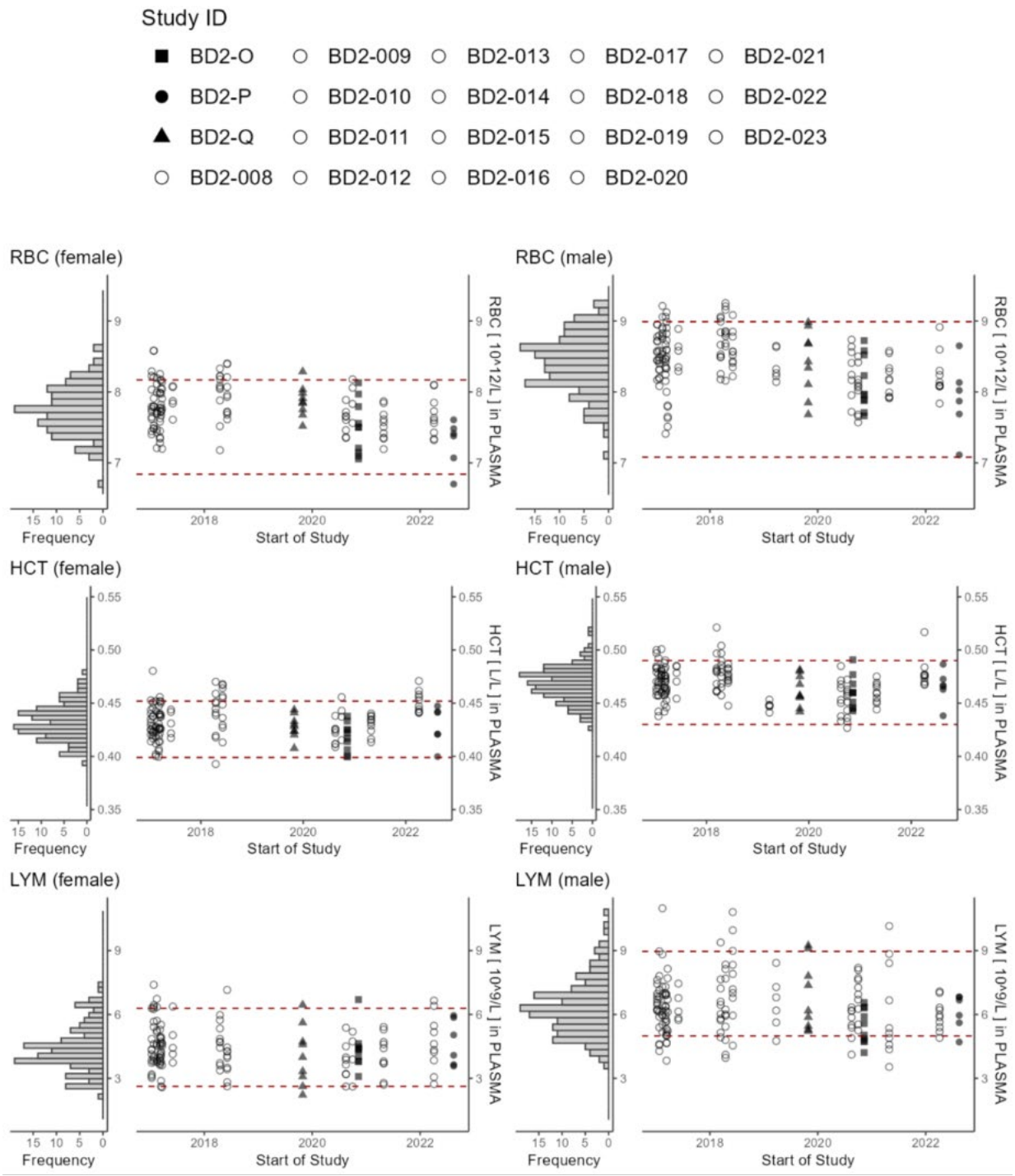


Figure 11: Distribution of RBC, HCT, LYM in control rats across time. *Distribution of erythrocyte counts (RBC), haematocrit (HCT) and lymphocyte counts (LYM) in control animals across a chosen time period. Female and male animals are represented on the left and the right side, respectively. In each figure, the measurements are arranged by the start of study year. A single dot represents liver enzyme measurement for a single animal on day 1 of study, before the start of treatment. Red dashed lines indicate the historic reference values used in the original studies and represent the upper (95% percentile) and lower (5% percentile) limit of normal, respectively. Note that the historic reference values are NOT calculated from the data analysed here.*

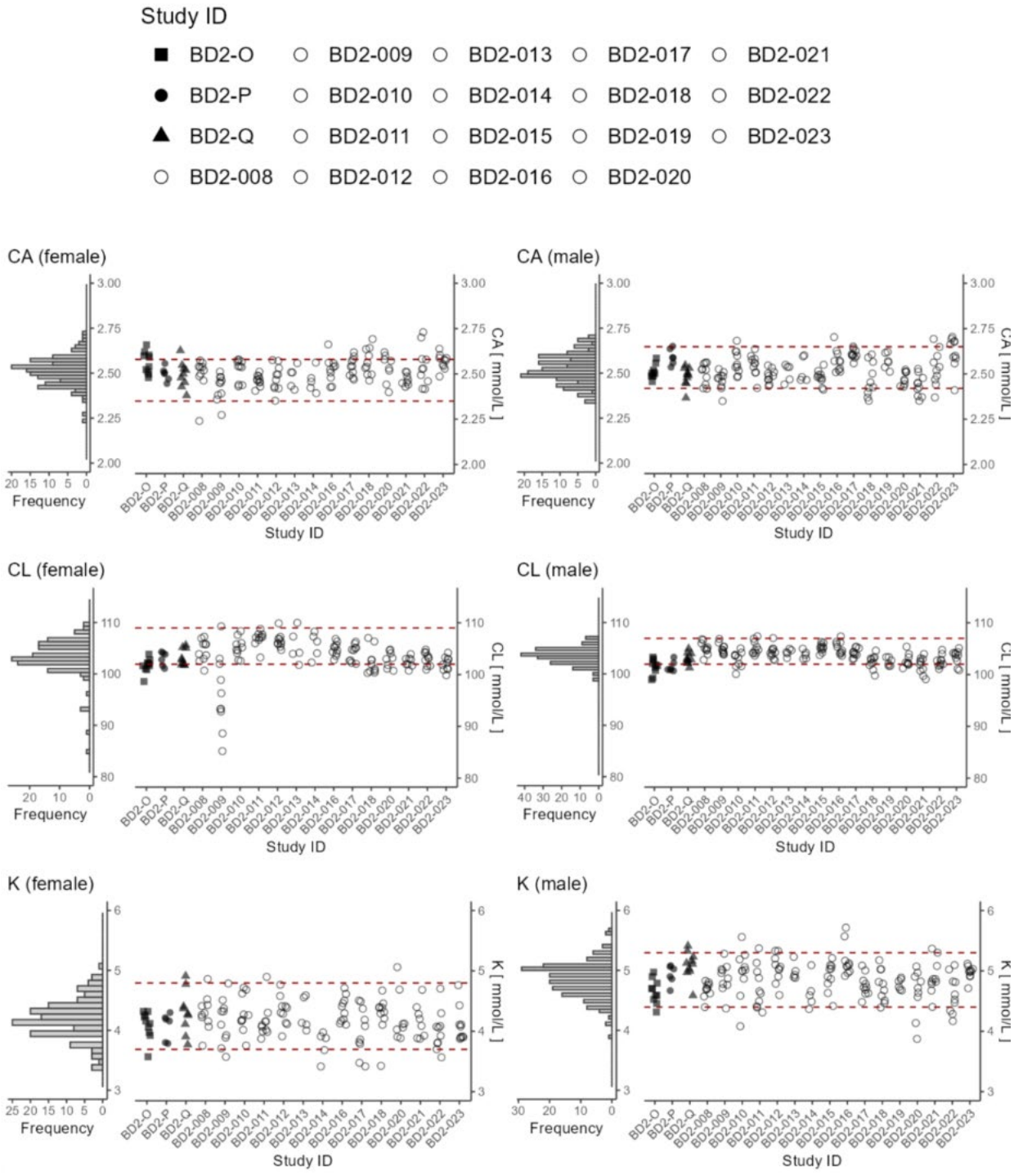


Figure 12: Electrolytes (CA, CL, K) distribution in control rats. *Distribution of the electrolytes calcium (CA), chloride (CL) and potassium (K) in control animals across eligible studies. A single dot represents liver enzyme measurement for a single animal on day 1 of study, before the* start of *treatment. Red dashed lines indicate the historic reference values used in the original studies and represent the upper (95% percentile) and lower (5% percentile) limit of normal, respectively. Note that the historic reference values are NOT calculated from the data analysed here.*

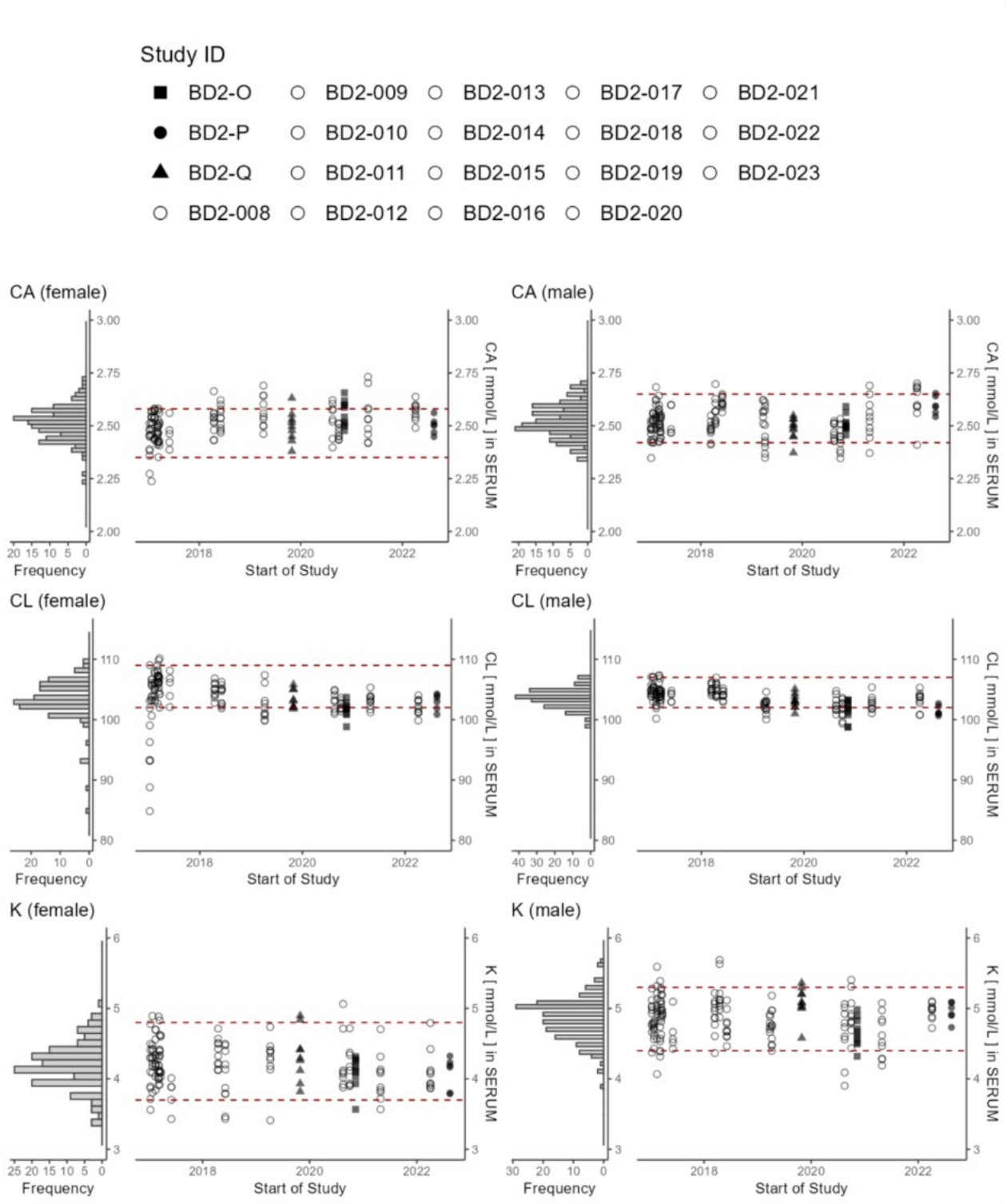


Figure 13: Electrolytes (CA, CL, K) distribution in control rats across time. *Distribution of the electrolytes calcium (CA), chloride (CL) and potassium (K) in control animals across the time span of eligible studies. A single dot represents liver enzyme measurement for a single animal on day 1 of study, before the start of treatment. Red dashed lines indicate the historic reference values used in the original studies and represent the upper (95% percentile) and lower (5% percentile) limit of normal, respectively. Note that the historic reference values are NOT calculated from the data analysed here.*

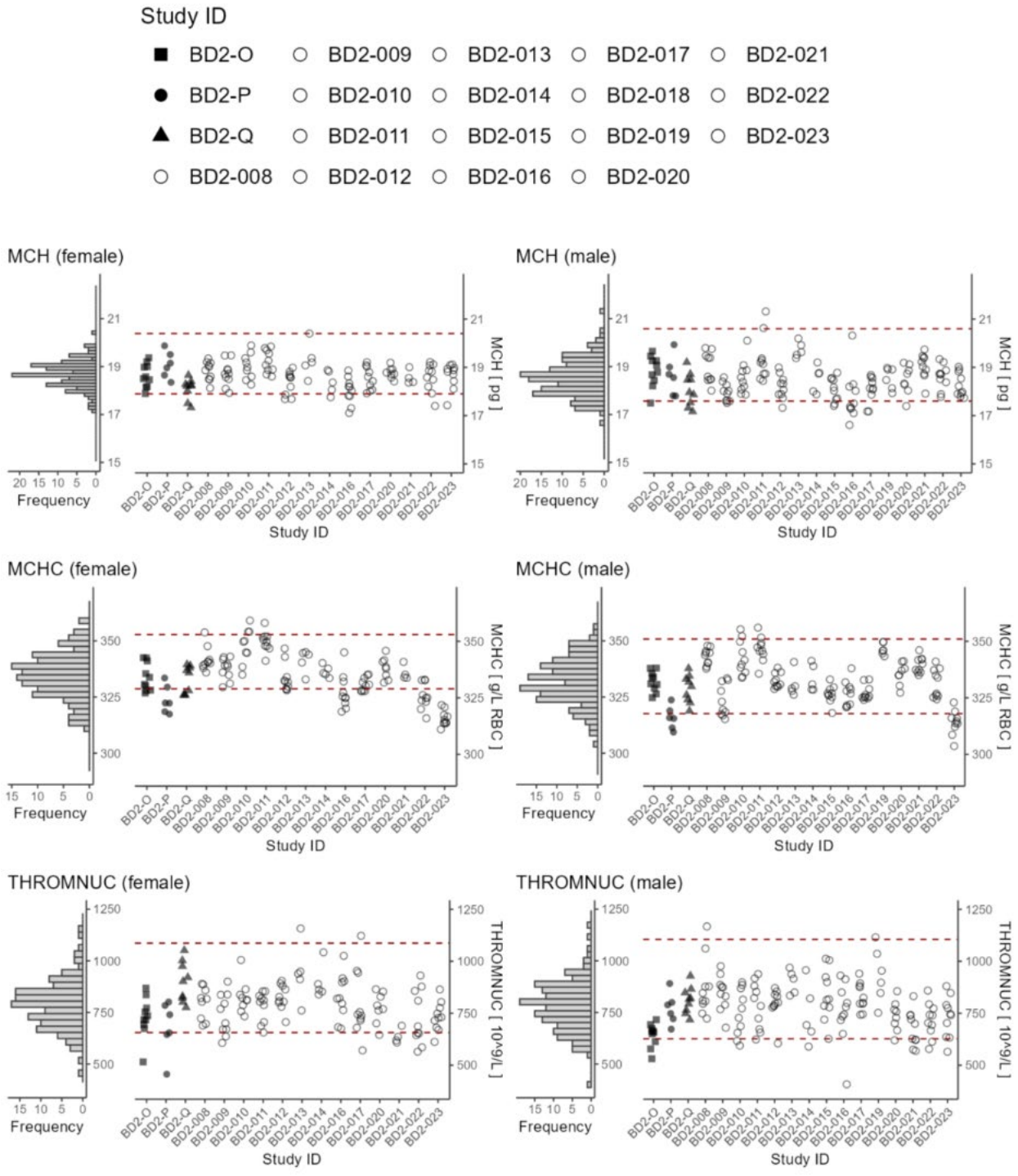


Figure 14: MCH, MCHC, THROMNUC distributions in control rats. *Distribution of the mean corpuscular hemoglobin (MCH), mean corpuscular hemoglobin concentration (MCHC) and thrombocyte counts (THROMNUC) in control animals across eligible studies. Electrolytes (CA, CL, K) distribution in control rats. A single dot represents liver enzyme measurement for a single animal on day 1 of study, before the start of treatment. Red dashed lines indicate the historic reference values used in the original studies and represent the upper (95% percentile) and lower (5% percentile) limit of normal, respectively. Note that the historic reference values are NOT calculated from the data analysed here.*

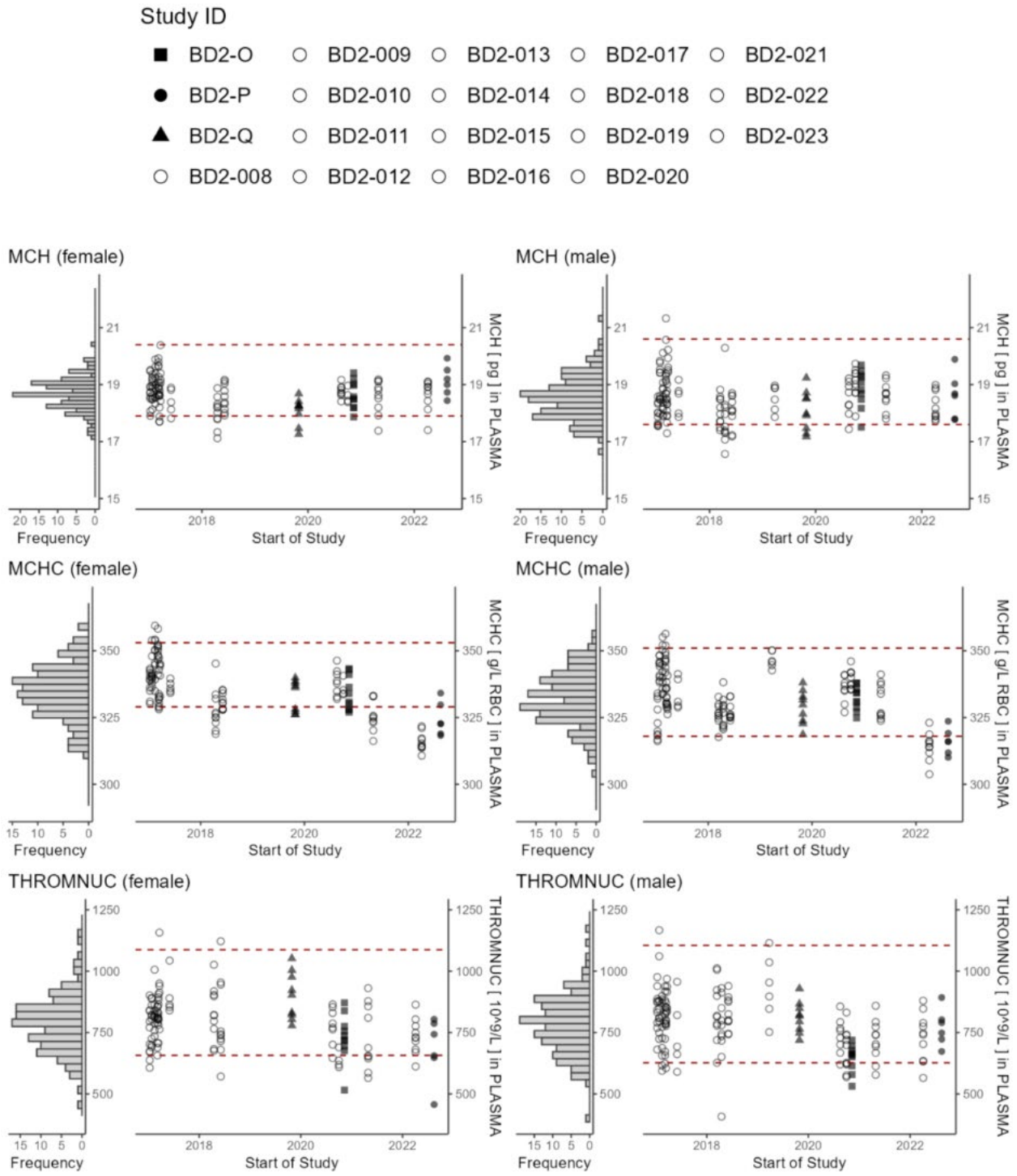


Figure 15: MCH, MCHC, THROMNUC distributions in control rats across the time span of studies. *Distribution of the mean corpuscular haemoglobin (MCH), mean corpuscular haemoglobin concentration (MCHC) and thrombocyte counts (THROMNUC) in control animals across eligible studies. A single dot represents liver enzyme measurement for a single animal on day 1 of study, before the start of treatment. Red dashed lines indicate the historic reference values used in the original studies and represent the upper (95% percentile) and lower (5% percentile) limit of normal, respectively. Note that the historic reference values are NOT calculated from the data analysed here.*

3.9. Further statistical analysis of initial body weight distribution for study BD2-O, study BD2-P and study BD2-Q

Matching with initial body weight is a key factor in the selection of appropriate VCGs (Gurjanov *et al.*, 2024b).

Figure 16 shows the distribution of initial body weights for all animals over the course of the selected time period. The specific distribution of body weight distribution for animals of study BD2-O is indicated by the red lines.

Figure 17 shows the distribution of the body weight in the control data before and after matching with the initial body weight of the treatment animals for study BD2-O.

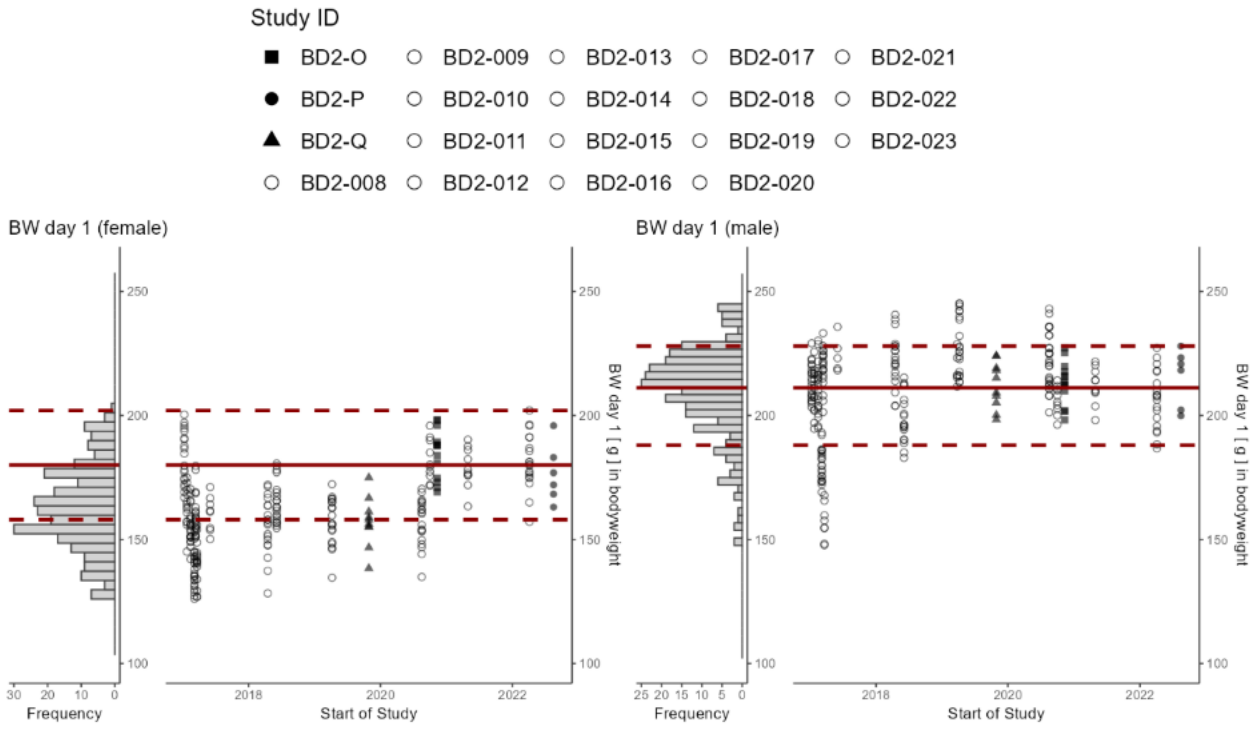


Figure 16: Body weight of control group animals referenced against study BD2-O. *Body weight of the HCD animals at day 1, before start of treatment, arranged by the start date of study. Red lines correspond to the study BD2-O and indicate the mean (solid line), min and max (dashed line) body weight of BD2-O control animals on day 1, i.e., before start of treatment.*

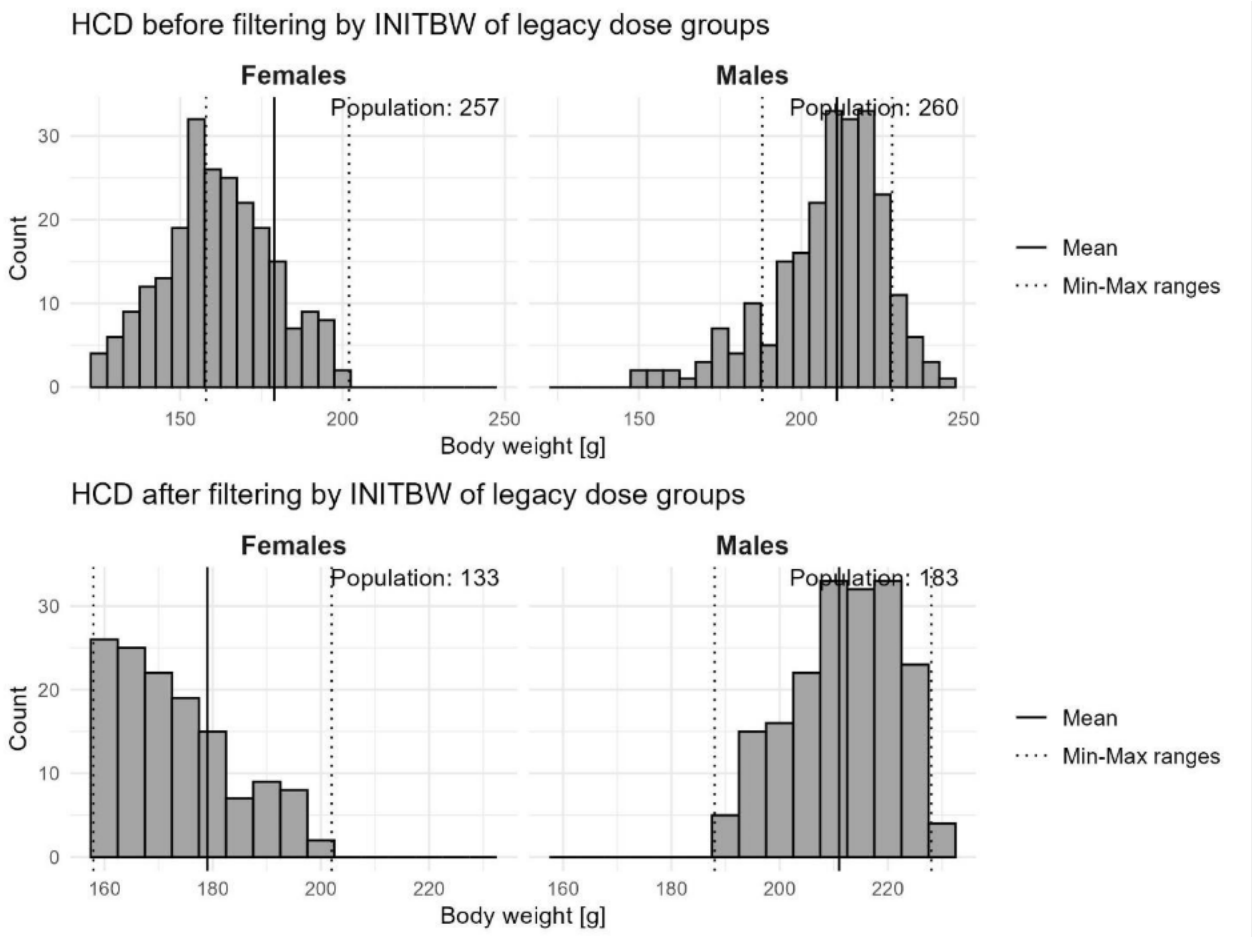


Figure 17: Body weight of HCD animals referenced against controls of study BD2-O. *Body weight of the HCD animals at day 1, before start of treatment, arranged by the start date of study. Red lines correspond to the study BD2-O and indicate the mean (solid line), min and max (dashed line) body weight of BD2-O control animals on day 1, i.e., before start of treatment.*

Figure 18 shows the distribution of initial body weights for all animals over the course of the selected time period. The specific distribution of body weight distribution for animals of study BD2-P is indicated by the red lines.

Figure 19 shows the distribution of the body weights in the control data before and after matching with the initial body weight of the treatment animals for study BD2-P.

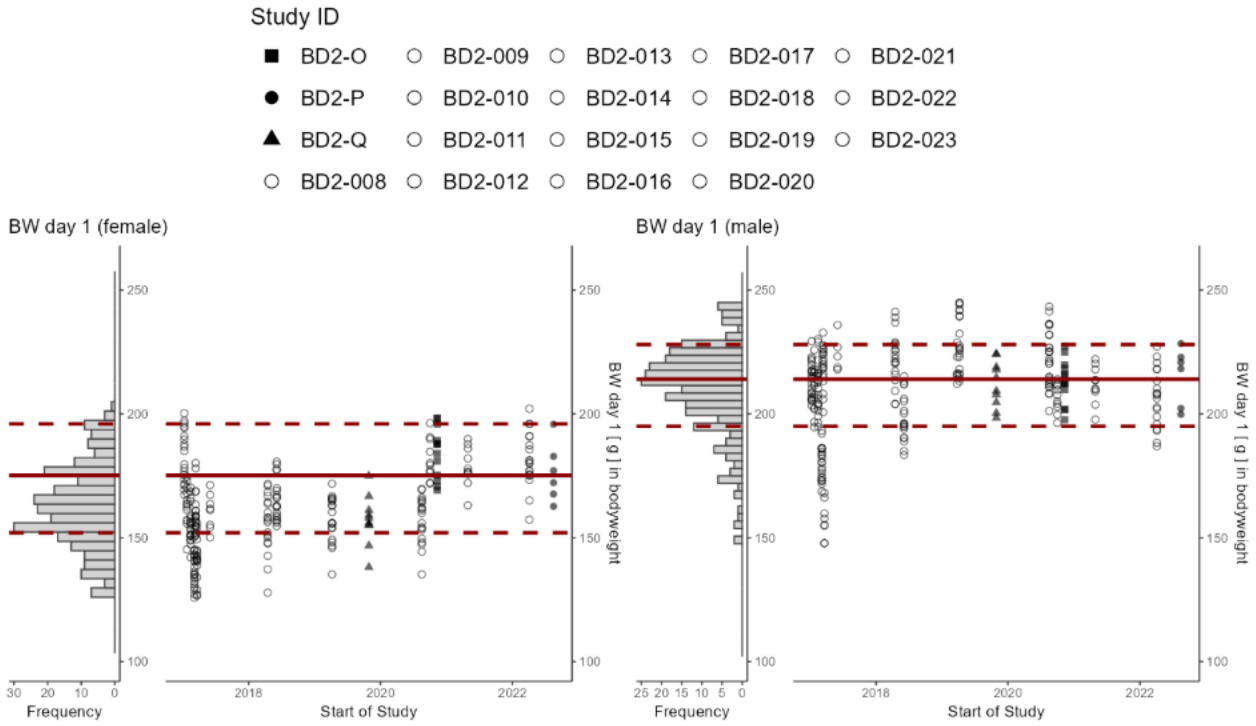


Figure 18: Body weight of control group animals referenced against study BD2-P. *Body weight of the HCD animals at day 1, before start of treatment, arranged by the start date of study. Red lines correspond to the study BD2-P and indicate the mean (solid line), min and max (dashed line) body weight of BD2-P control animals on day 1, i.e., before start of treatment.*

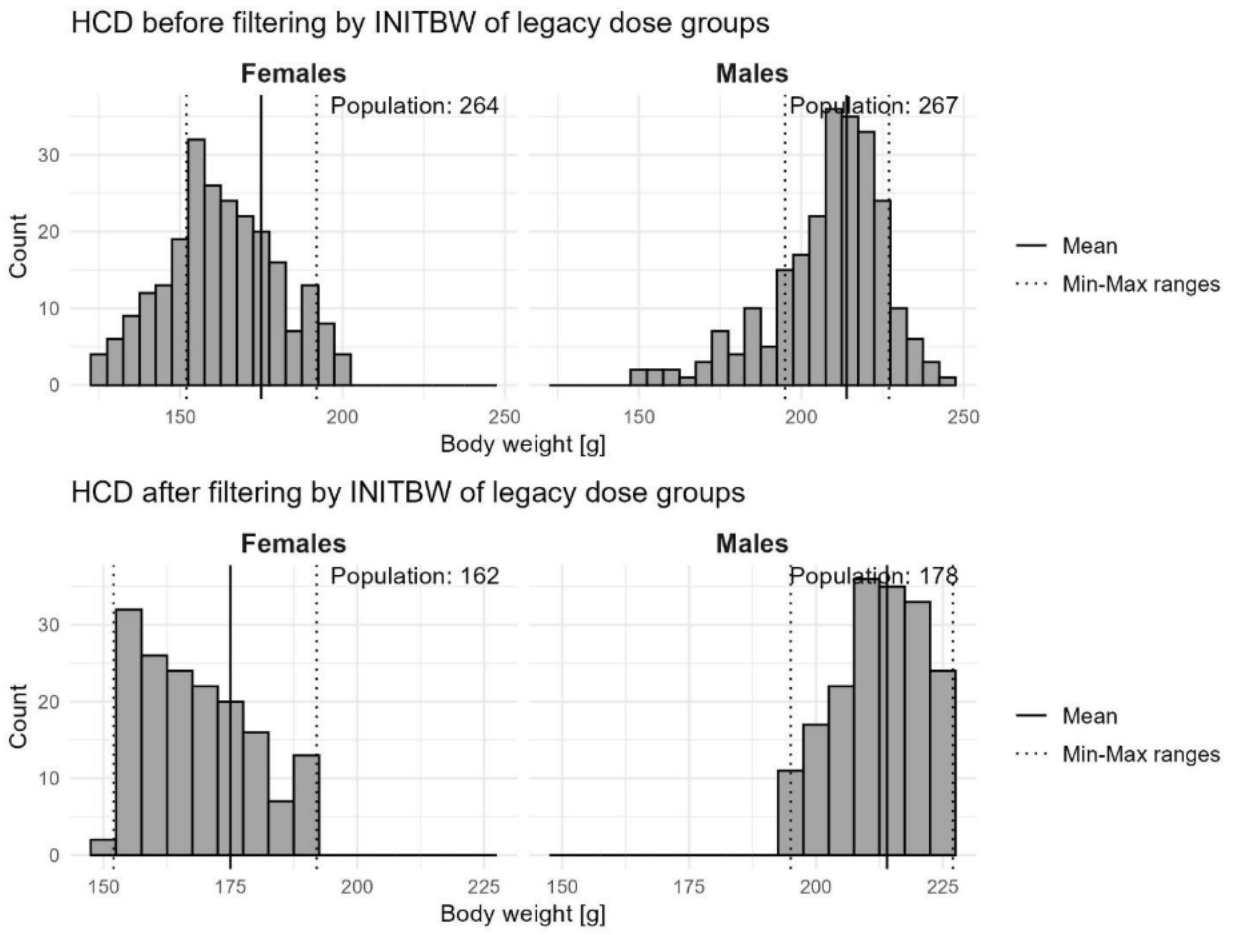


Figure 19: Body weight of the HCD and VCG-pool selected for study BD2-P. *Population of HCD animals and the selected VCG pool for study BD2-P at start of study. Upper row: all HCD animals; lower row: remaining HCD animals after filtering by the initial body weight (INITBW) of legacy dose groups, i.e. the VCG pool from which VCG group animals are sampled.*

Figure 20 shows the distribution of initial body weights for all animals over the course of the selected time period. The specific distribution of body weight distribution for animals of study BD2-Q is indicated by the red lines.

Figure 21 shows the distribution of the body weight in the control data before and after matching with the initial body weight of the treatment animals for study BD2-Q.

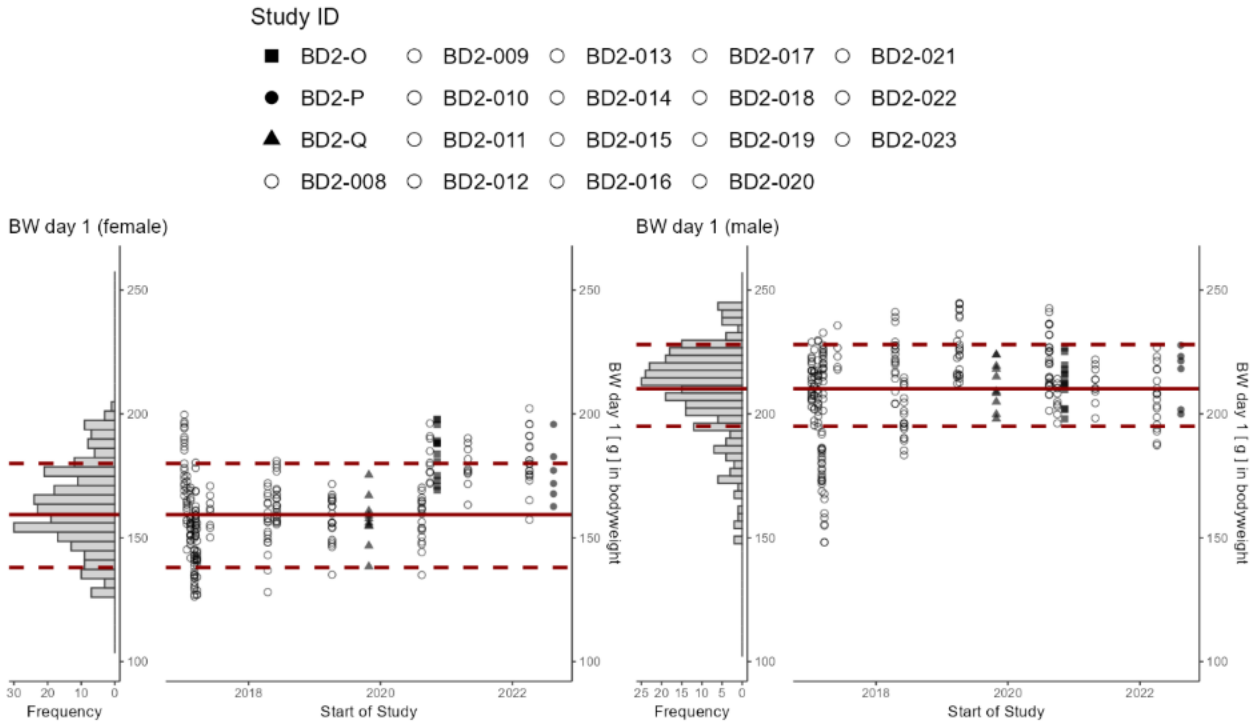


Figure 20: Body weight of HCD animals referenced against controls of study BD2-Q. *Body weight of the HCD animals at day 1, before start of treatment, arranged by the start date of study. Red solid lines correspond to the study BD2-Q and indicate the mean (solid line), min and max (dashed line) body weight of BD2-Q control animals on day 1, i.e., before start of treatment.*

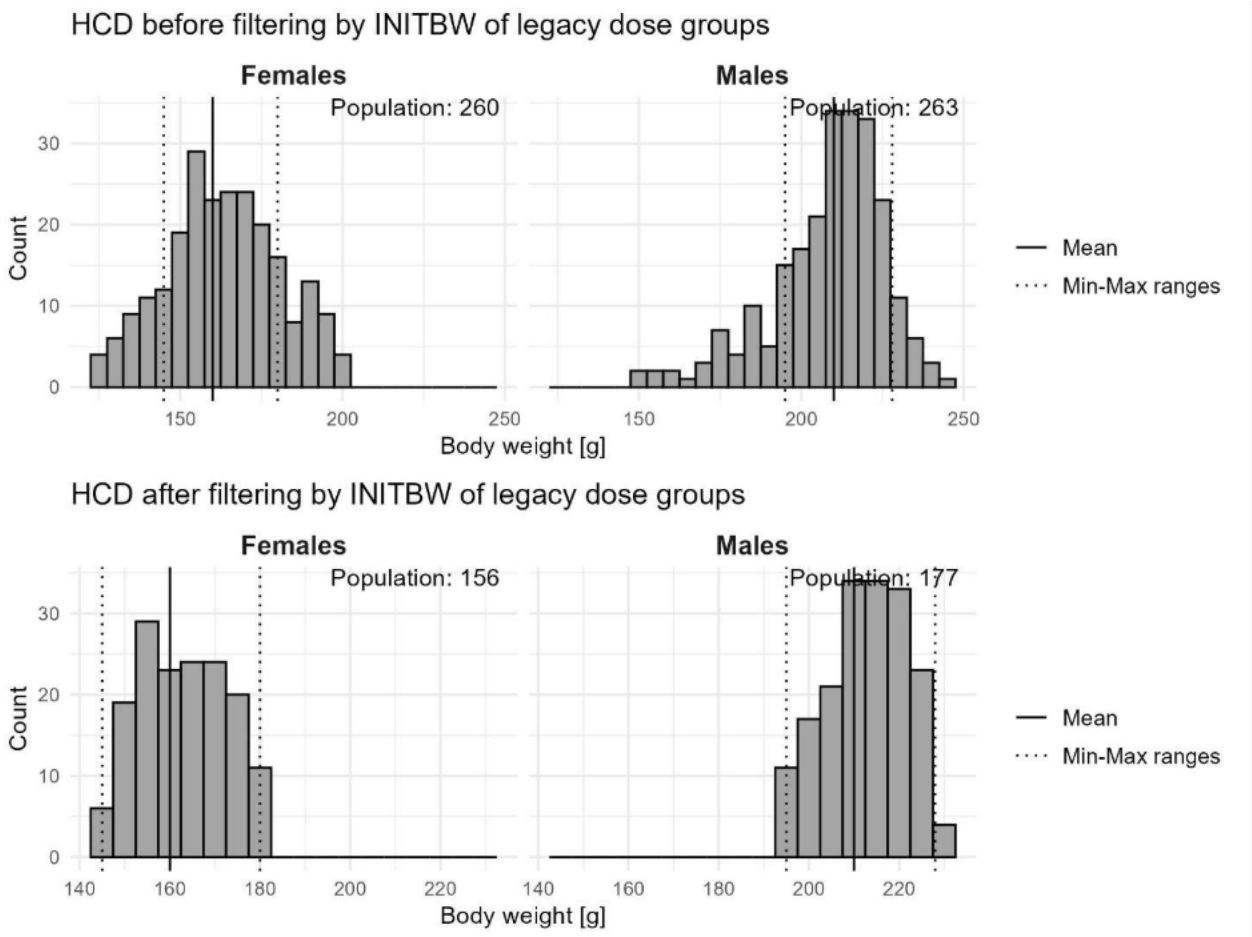


Figure 21: Body weight of the VCG-pool selected for study BD2-Q. *Population of HCD animals in study BD2-Q at start of study. Upper row: all HCD animals; lower row: remaining HCD animals after filtering by the initial body weight (INITBW) of legacy dose groups, i.e. the VCG pool from which VCG group animals are sampled.*

3.10. Results of study reanalysis H

3.10.1. Study description and original results (study H)

In this GLP 4-week toxicity study with Hsd:Sprague Dawley (SD) rats the test item was administered twice daily via oral gavage for 28 days.

Test item-related microscopic findings consisted of hypertrophy of the centrilobular hepatocytes in the liver of animals in the HD, which correlated with increased liver organ weights in animals in the MD; and erosion/ulcer and/or inflammation in the non-glandular stomach of males starting in the LD and females in the MD. 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. Test item-related clinical observations were observed in females in the HD 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 MD.

3.10.2. Results 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 sponsor's VCG pool. E.g., for the liver microscopic findings, the findings observed in the treatment groups were neither seen in male nor in female control groups (see CCG and the example of a VCG in Figure 22). Consequently, interpretation

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	1	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

Figure 22). 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 test item-related.

The same was true for finding of the nonglandular stomach: there were no changes after replacing the CCGs by VCG Random from the sponsor's 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 HD animals were reported in the VCGs whereas they were seen in the CCG. But when using the VCG min-max and incidence values, the interpretation and assessment come to the same conclusion that these kidney findings are not test item-related because of the relative high incidences and maximal occurrence of these findings in control groups per study which was in the same range as observed for the HD.

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.

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	1	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

Figure 22: 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.*

Accordingly, changes for the other clinical observations, clinical pathology and other parameters were analysed in the same way. Table 11 shows examples for which statistical analysis differs between the CCG and a randomly selected VCG from the sponsor's VCG pool as example, ordered by the most prominent changes for male and female rats. An additional graphical analysis is shown in Figure 23 for a few examples. Differences were seen in the statistical results, e.g., for ALP in male HD animals (see also Table 11: ALP was statistically significant in CCG HD with p=0.00022, but not significant in the VCG with p=0.23). However, since microscopic changes in the liver of HD 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 MD, and first adverse effects were seen in the HD.

Table 11: *Overview of results for legacy study H, where the statistical analyses differ comparing CCGs with VCGs for male and female animals.*

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

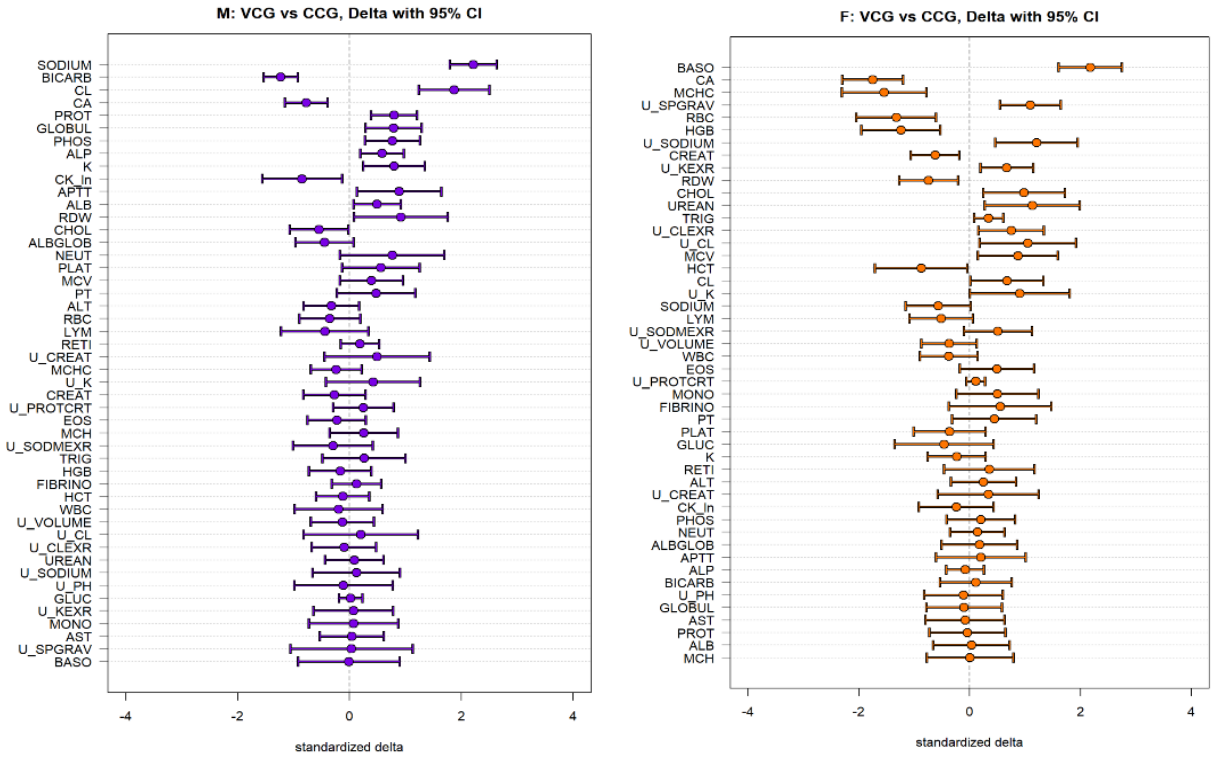


Figure 23: Overview of main clinical pathology results from the legacy rat study H at study end after 28-day treatment.

4. Overall performance of VCGs regarding study conclusions

4.1. Differences and commonalities between matching procedures

An SOP was developed during the qualification procedure (see Annex) that describes the recommended processes for generation, application and documentation of VCGs in preclinical studies. Although the use of an SOP is not required for CoU I given the non-GLP nature of DRF studies, their matching procedures were aligned with the recommendations of the SOP to the maximum extent possible.

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 studies of the Summary Qualification Document the following criteria were applied in all eight studies (see Table 2):

- 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

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 (see Table 2).

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

4.2. 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 and is described in 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)

4.2.1. 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.

4.2.2. 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.

4.2.3. 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.

4.2.4. 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 (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.

4.3. Replicability results

Table 12 provides a summary of the replicability of the four key study objectives described.

Table 12: *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".

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.

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
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 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.

4.4. Additional replicability observations

With respect to the performance of VCGs, additional 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 preclinical studies and are therefore unavoidable in a qualification process (Steger-Hartmann *et al.* 2025). We therefore consider such discrepancies rather 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 expert, 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.

4.5. Overall conclusions

The high rate of replicability of findings from original studies leads to the conclusion that the qualification of VCGs yielded satisfactory results within CoU I for rats.

5. List of references

Golden, E. *et al.* (2024) 'Toward implementing virtual control groups in nonclinical safety studies', *ALTEX*, 41(2):282–301. Available at: https://doi.org/10.14573/altex.2310041.

Gordon, I.R. (2018) 'CONFIDENCE INTERVALS AND REPLICATION: THE REALITY', In: Sorto IMA, White A, (Eds.) & LG (eds) Proceedings of the Tenth International Conference on Teaching Statistics (ICOTS10). Available at: https://iase-web.org/icots/10/proceedings/pdfs/ICOTS10_8A3.pdf?1531364291.

Gurjanov, A. *et al.* (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', *Frontiers in Pharmacology*, 14. Available at: https://doi.org/10.3389/fphar.2023.1142534.

Gurjanov, A. *et al.* (2024a) 'Replacing concurrent controls with virtual control groups in rat toxicity studies', *Regulatory Toxicology and Pharmacology: RTP*, 2024; 148, p. 105592. Available at: https://doi.org/10.1016/j.yrtph.2024.105592.

Gurjanov, A. *et al.* (2024b) 'The road to virtual control groups and the importance of proper body-weight selection', *ALTEX.* 2024;41(4):660-665. Available at: https://doi: 10.14573/altex.2403141.

de Kort, M. *et al.* (2020) 'Historical control data for hematology parameters obtained from toxicity studies performed on different Wistar rat strains: Acceptable value ranges, definition of severity degrees, and vehicle effects', *Toxicology Research and Application*, 4: 2397847320931484. Available at: https://doi.org/10.1177/2397847320931484.

Kluxen, F. *et al.* (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., *et al.* (2010). Predictive toxicology approaches for small molecule oncology drugs. Toxicologic Pathology 38(1)**,** 148-164. https://doi.org/10.1177/0192623309356448.

National Academies of Sciences, Engineering, and Medicine. (2019) *Reproducibility and Replicability in Science*. Washington (DC): National Academies Press (US). Available at: https://doi.org/10.17226/253ordPalazzi, X. *et al.* (2024) 'Points to consider regarding the use and implementation of virtual controls in nonclinical general toxicology studies', *Regulatory toxicology and pharmacology: RTP*, 150, p. 105632. Available at: https://doi.org/10.1016/j.yrtph.2024.105632.

OECD (2002). Guidance Notes for Analysis and Evaluation of Chronic Toxicity and Carcinogenicity Studies. OECD Environment, Health and Safety Publications Series on Testing and Assessment No. 35 And Series on Pesticides No. 14. https://one.oecd.org/document/ENV/JM/MONO(2002)19/en/pdf

Palazzi, X., et al. (2024). Points to consider regarding the use and implementation of virtual controls in nonclinical general toxicology studies. Regulatory Toxicology and Pharmacology, 2024;150, 105632. Available at: https://doi.org/10.1016/j.yrtph.2024.105632.

Sato, G. *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. Available at: https://doi.org/10.33611/trs.2023-009.

Steger-Hartmann *et al.* (2020). 'Introducing the concept of virtual control groups into preclinical toxicology testing'. *ALTEX*, 2020;37(3):343-349. Available at: https://doi: 10.14573/altex.2001311.

Steger-Hartmann, T. and Clark, M. (2023). 'Can Historical Control Group Data Be Used to Replace Concurrent Controls in Animal Studies?' *Toxicologic Pathology*, 2023;51, 6. Available at: https://doi:10.1177/01926233231208987.

Steger-Hartmann, T. *et al.* (2025) 'Virtual control groups in non-clinical toxicology - A replicability challenge', *ALTEX*, *2025;* 42(3):538-542. Available at: https://doi.org/10.14573/altex.2503061.

Weber, K. *et al.* (2011) 'Differences in rat models used in routine toxicity studies', *International Journal of Toxicology*, 30(2):162–173. Available at: https://doi.org/10.1177/1091581810391818.

Wolford, S.T. *et al.* (1987) 'Age-related changes in serum chemistry and haematology values in normal Sprague-Dawley rats', *Fundamental and Applied Toxicology*, 8(1):80–88. Available at: https://doi.org/10.1016/0272-0590(87)90102-3.

6. Annex - SOP

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

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

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

Version: 1.1

Table of Content:

1. Introduction

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

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

Current European and international regulations require animal studies for the nonclinical safety assessment prior to conducting clinical trials and market authorization for pharmaceuticals (8). The conventional setting of a regulatory toxicology study uses 25 % of the animals as controls (e.g. OECD

TG 407 "Repeated Dose 28-Day Oral Toxicity Study in Rodents" (12)). The relevant guidelines for chronic studies or carcinogenicity studies follow a similar scheme (11, 13), with the exception that the animal numbers per group increase with the duration of the study. Applying this SOP to other or less conventional study designs may require additional consideration not addressed here.

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

2. Scope

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

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

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

3. Responsibilities

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

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

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

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

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

4. Definitions

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

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

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

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

5.1. Data Collection

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

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

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

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

5.2. Data Standard and Curation

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

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

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

6. Procedure for Generation of VCGs

6.1. List of matching criteria

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

These include the following criteria:

a) Test facility

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

This list of criteria represents a subset of criteria established by other authors (14, 15) that may impact on the variability of parameters measured in animal studies, chosen based on the biological plausibility of a criteria to elicit a high influence on variability. These criteria can be pre-specified during the planning phase of the study as soon as the final study protocol is available. One exception is the initial body weight for rodents (criterion k) that becomes available only at study start upon delivery of the animals to the test facility or during the acclimation phase. Therefore, initial body weight will be the last criterion applied to the HCD before constructing VCGs for rodents (6). When applying this criterion, the initial body weight of rodents in the selected HCD should reflect the initial body weight range of the animals assigned to a new study. Statistical tests or visual inspection should be performed to ensure that the initial body weight does not differ substantially from the initial weights of the treatment group animals (7). At a minimum, the initial body weight of rodents in the selected HCD must fall within the minimum and maximum initial body weight of treatment group animals.

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

In addition to the above-described criteria, study-specific parameters measured during the acclimation phase particularly for larger species can serve as additional matching criteria or as criterion for deselecting animals due to extreme pre-values, especially if the pre-values for animals in the selected HCD show a different distribution compared to treatment groups animals (see 6.4) and specified in the protocol.

6.2. Potential modifications of the matching criteria

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

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

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

6.3. Required animal number in VCGs

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

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

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

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

6.4. Matching based on pre-values for larger species

Pre-values, also called baseline values or pretreatment values, are measurements or observations taken prior to the first administration of the test item, usually also before the allocation of animals into treatment and control groups. Recommendations on how to perform these measurements are not explicitly mentioned in the OECD guideline for dog repeated dose toxicity studies (11) but are mentioned in other guidelines (1), (3). However, a full list of parameters measured is not set forth in these guidelines and no explicit recommendation on how to use the pre-values before, during and after the conduct of a study is given.

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

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

6.5. Assessing the relevance of the matching criteria

The animals selected for VCGs should relate to the treatment group population in the same way as the CCG. In particular, the CCG, treatment group and the VCG animals should represent samples from the

same animal population, for this purpose treatment group animal data before the start of the treatment should be used. Based on additional data collection and analyses, the VICT3R consortium will recommend in the future which matching criteria will have to be considered and which can be disregarded or grouped together.

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

VCGs cannot be applied if important characteristics of the planned study (e.g. a special type of vehicle, a new biomarker or a previously not determined endpoint) are not represented in the pool of controls used to build VCGs. In such cases a CCG would need to be used for this specific study. The data of the CCG should be captured in the VCG database.

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

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

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

6.6. Documentation and Reporting of VCGs

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

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

f. VCG data must be reported in Appendices I and II of the final report together with statistical results identical to the procedures established for CCGs documentation.
g. The report should contain a general statement on the validity of the generated VCGs.

7. Conclusions

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

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

8. References

(1) EMA (2010). Guideline on repeated dose toxicity. 18 March 2010 CPMP/SWP/1042/99 Rev 1 Corr* Committee for Human Medicinal Products (CHMP). https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-repeated-dose-toxicity-revision-1_en.pdf
(2) EMA (2024). Reflection paper on establishing efficacy based on single-arm trials submitted as pivotal evidence in a marketing authorisation application: Considerations on evidence from single-arm trials. (EMA/CPMP/458061/2024). https://www.ema.europa.eu/en/documents/scientific-guideline/reflection-paperestablishing-efficacy-based-single-arm-trials-submitted-pivotal-evidence-marketingauthorisation-application_en.pdf.
(3) 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.
(4) Golden E, Allen D, Amberg A, Anger LT, Baker E, Baran SW, Bringezu F, Clark M, Duchateau-Nguyen G, Escher SE, Giri V, Grevot A, Hartung T, Li D, Lotfi L, Muster W, Snyder K, Wange R, Steger-Hartmann T. Toward implementing virtual control groups in nonclinical safety studies. ALTEX. 2024;41(2):282-301. doi: 10.14573/altex.2310041.
(5) Gurjanov A, Kreuchwig A, Steger-Hartmann T, Vaas LAI. Hurdles and signposts on the road to virtual control groups-A case study illustrating the influence of anesthesia protocols on electrolyte levels in rats. Front Pharmacol. 2023 Apr 20;14:1142534. doi: 10.3389/fphar.2023.
(6) Gurjanov, A, Vieira-Vieira, C, Vienenkoetter, J, Vaas, LAI, Steger-Hartmann, T (2024). 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
(7) Gurjanov A, Vaas LAI, Steger-Hartmann T. The road to virtual control groups and the importance of proper body-weight selection. ALTEX. 2024;41(4):660-665. doi:10.14573/altex.2403141.
(8) ICH M3 (2008). Guideline M3(R2) on non-clinical safety studies for the conduct of human clinical trials and marketing authorisation for pharmaceuticals. https://www.ema.europa.eu/en/ich-m3-r2-nonclinical-safety-studies-conducthuman-clinical-trials-pharmaceuticals-scientific.
(9) ICH E10 (2001). Topic E 10 Choice of Control Group in Clinical Trials. Note or Guidance on Choice of Control Group in Clinical Trials. (CPMP/ICH/364/96). doi.org/10.1787/9789264070684-en.
(10) National Cancer Institute (2025): Enterprise Vocabulary Services: https://evs.nci.nih.gov/evs-download/thesaurus-downloads.
(11) OECD (1998), Test No. 409: Repeated Dose 90-Day Oral Toxicity Study in Non-Rodents, OECD-Guidelines for the Testing of Chemicals, Section 4, OECD Publishing, Paris, doi.org/10.1787/9789264070721-en.
(12) OECD (2008), Test No. 407: Repeated Dose 28-day Oral Toxicity Study in Rodents, OECD Guidelines for the Testing of Chemicals, Section 4, OECD Publishing, Paris, doi.org/10.1787/9789264070684-en.

(13) OECD (2008b). Test No. 452: Chronic Toxicity Studies. OECD Guidelines for the Testing of Chemicals, Section 4. OECD Publishing, Paris. doi:10.1787/9789264071209-en.
(14) Palazzi, X., Anger, L.T., Boulineau, T., Grevot, A., Guffroy, M., et al. (2024). Points to consider regarding the use and implementation of virtual controls in nonclinical general toxicology studies. Regulatory Toxicology and Pharmacology, 150, 105632. doi.org/10.1177/01926233241309905.
(15) Sato, G., Nakajima, M., Sakai, K., Togashi, Y., Yamamoto, M., et al. (2024). Potential issues associated with the introduction of virtual control groups into non-clinical toxicology studies. Translational and Regulatory Sciences, 6(1), 1-9. doi.org/10.33611/trs.2023-009.
(16) Standard for Exchange of Nonclinical Data (SEND) Implementation Guide. Version v3.1.1. https://www.cdisc.org/standards/foundational/send/sendig-v3-1-1.
(17) Steger-Hartmann, T., Kreuchwig, A., Vaas, L., Wichard, J., Bringezu, F., Amberg, A., et al. (2020). Introducing the concept of virtual control groups into preclinical toxicology testing. ALTEX, 37(3), 343-349. doi.org/10.14573/altex.2001311.