101172693 – VICT3R

Developing and implementing Virtual Control Groups to reduce animal use in Toxicology Research

WP3 – Extension to other study types

D3.1 Study designs overview

Lead contributor:	Marco Dilger (18-BASF SE)
Other contributors:	Olga Valverde (1-UPF), Javier Orihuel (1-UPF), Julia Anna Matyasiak (13-VUB), Annika Kreuchwig (14-Bayer), Frank Bringezu (15-MKDG), Sandrine d'Argouges (16-Amgen), Eleanor McRae (17-AZ), Bernd Baier (19-BII), Marc Ellison (22-Instem), Anthony DeLise (25-Novartis), Mary Carsillo (31-Takeda)
Due date:	28 February 2025
Delivery date:	12 May 2025
Dissemination level:	Public

Contents

Document History

Version	Date	Description
V1.0	07 March 2025	Version for internal review
V1.1	08 April 2025	Version after internal review
V1.2	09 April 2025	Version for consortium review
V1.3	06 May 2025	Implemented changes after consortium review
V2.0	08 May 2025	Final version

Definitions

Participants of the VICT3R Consortium are referred to herein according to the following codes:

1. UNIVERSIDAD POMPEU FABRA (UPF)
2. SYNAPSE RESEARCH MANAGEMENT PARTNERS SL (SYNAPSE)
3. FRAUNHOFER GESELLSCHAFT ZUR FORDERUNG DER ANGEWANDTEN FORSCHUNG EV (ITEM)
4. GRITSYSTEMS AS (grit42)
5. TECHNISCHE UNIVERSITAT DORTMUND (TUDO)
6. PHUSE CUIDEACHTA FAOI THEORAINN RATHAIOCHTA (PHUSE)
7. MEDBIOINFORMATICS SOLUTIONS SL (MBIS)
8. BARCELONA SUPERCOMPUTING CENTER CENTRO NACIONAL DE SUPERCOMPUTACION (BSC CNS)
9. UNIVERSITAT KONSTANZ (UKON)
10. DECIPHEX LIMITED (Deciphex)
11. ORGANON SRL (Organon)
12. SYNCWORK AKTIENGESELLSCHAFT (Syncwork)
13. VRIJE UNIVERSITEIT BRUSSEL (VUB)
14. BAYER AKTIENGESELLSCHAFT (Bayer)
15. MERCK KOMMANDITGESELLSCHAFT AUF AKTIEN (MKDG)
16. AMGEN RESEARCH (MUNICH) GMBH (Amgen)
17. ASTRAZENECA UK LIMITED (AZ)
18. BASF SE (BASF SE)
19. BOEHRINGER INGELHEIM INTERNATIONALGMBH (BII)
20. BRISTOL-MYERS SQUIBB COMPANY CORP (BMS)
21. INCYTE BIOSCIENCES DISTRIBUTION B.V (Incyte BD)
22. INSTEM SCIENTIFIC LIMITED (Instem)
23. IPSEN INNOVATION SAS (IPSEN)
24. JANSSEN PHARMACEUTICA NV (Janssen)
25. NOVARTIS PHARMA AG (Novartis)
26. NOVO NORDISK A/S (Novo)
27. ORION OYJ (Orion)
28. PFIZER INC (Pfizer)
29. F. HOFFMANN-LA ROCHE AG (ROCHE)
30. SANOFI-AVENTIS DEUTSCHLAND GMBH (Sanofi)
31. TAKEDA PHARMACEUTICALS INTERNATIONAL AG (TPIZ)
32. UCB BIOPHARMA (UCB)
33. ABBVIE INC (AbbVie)

Grant Agreement	The agreement signed between the beneficiaries and the IHI for the undertaking of the VICT3R project (101172693).
Project	The sum of all activities carried out in the framework of the Grant Agreement.
Consortium	The VICT3R Consortium, comprising the above-mentioned legal entities.
Consortium Agreement	Agreement concluded amongst VICT3R beneficiaries for the implementation of the Grant Agreement (GA). Such an agreement shall not affect the parties' obligations to the Community and/or to one another arising from the GA.

Executive Summary

This report evaluates the feasibility of applying the VCG concept to study types not handled by WP2, such as Developmental and Reproductive Toxicology (DART), carcinogenicity, and certain pharmacological/mechanistic studies. The evaluation considers the compatibility of study designs with the VCG concept, taking into consideration aspects like industry-specific guidelines and differences in study design between different testing facilities or sponsors.

A survey was conducted among consortium partners to gather insights from subject matter experts regarding the compatibility of these study types with the VCG approach. The survey results were analyzed to prioritize study types for further exploration in subsequent tasks. Additionally, data from the Animal Use Reporting EU System (ALURES) was used to determine areas where the VCG concept could have the greatest benefit in terms of reducing animal usage.

We conclude that DART studies, particularly Embryo-Fetal Developmental (EFD) studies, are well-suited for VCG implementation due to their standardized study designs and comprehensive coverage by the SEND standard. Fertility studies, another subtype of DART studies, also are a strong candidate for VCG application. Application of VCG to carcinogenicity studies is considered to be the easiest achievable amongst the studies in scope, however they are conducted less frequently. While carcinogenicity studies are considered the most straightforward to adapt to the VCG approach among the evaluated study types, their lower frequency of conduct limits the impact of VCGs. Safety Pharmacology (SP) studies also present a promising opportunity for VCGs, given the high level of data standardization and willingness among consortium members to share data.

Introduction

This report evaluates the suitability of applying the Virtual Control Group (VCG) concept to study types not handled by WP2, such as Developmental and Reproductive Toxicology (DART), carcinogenicity, and certain pharmacological/mechanistic studies. The evaluation considers the compatibility of study designs with the VCG concept, industry-specific guidelines, variations between testing facilities, and intra-facility differences. Subject matter experts from several consortium partners participated in a survey and provided information based on their studies and a feasibility assessment. This evaluation is essential to prioritize study types for further tasks in WP3 and interactions with other work packages (WP 6, 7, and 8).

Survey among beneficiaries

A survey was designed and distributed amongst all beneficiaries of the action. The survey aimed to evaluate the feasibility of applying the Virtual Control Group (VCG) concept to study types not covered by WP2, such as Developmental and Reproductive Toxicology (DART), carcinogenicity, and certain pharmacological/mechanistic studies. The survey was designed to gather insights from subject matter experts regarding the compatibility of these study types with the VCG approach considering variations in industry-specific guidelines, differences between testing facilities, and intra-facility discrepancies (Table 1).

Table 1: Survey questions and intended purpose

Company Information	We began by collecting basic information such as the respondent's company name.
Study Type Frequency and Data Availability	Respondents were asked to indicate the frequency of various study types conducted within their organization (e.g., DART, carcinogenicity, juvenile studies, etc.) and the availability of data for each. This information was vital for assessing the potential volume of data that could support the VCG concept
Compatibility with VCG Concept	The survey included questions on the perceived compatibility of each study type with the VCG approach. Respondents were asked to rank the suitability of study types for VCG application, considering aspects like study design and industry guidelines. This helped identify which study types were most feasible for VCG integration.
Data Format and Accessibility	To understand data accessibility, we asked about the formats in which study data were available (e.g., PDF, Excel/CSV, SEND format). This question aimed to identify potential challenges in data harmonization across different companies.

Additional Comments and Insights	Respondents were encouraged to provide additional comments or insights regarding the VCG approach. This open-ended question allowed for the capture of nuanced perspectives and suggestions that could inform future work.

The responses were collected and analyzed within WP3, where they were discussed and finally consolidated into this deliverable. This process involved evaluating the frequency and compatibility of the data to prioritize study types for further exploration in the subsequent tasks. The insights gained from the survey provided a foundational understanding of the current landscape and will inform strategic decisions regarding the application of the VCG concept for different study types.

Numbers of animals used in regulatory studies

Beyond assessing VCG feasibility, it is essential to identify areas where the VCG concept could provide the greatest benefit in terms of reducing animal usage. While research applications account for a higher overall use of animals, the regulatory sector remains a more viable candidate for VCG implementation due to its higher degree of standardization. In addition to the survey performed amongst consortium members, data on reported animal uses in the EU was used as a basis for the rationale to guide the development of the VCG concept, particularly in terms of its potential impact on reducing animal use. The Animal Use Reporting EU System (ALURES) is a statistical database commissioned by the EU under directive 2010/63/EU on the protection of animals used for scientific purposes. It collates data submitted by the member states to provide a holistic overview of the scientific use of animals in Europe. It is important to note that the US, where animal use is estimated by some analyses to be much higher (Taylor 2024), does not regulate systematic accounting for data per sector and legislation as in ALURES. Therefore, we assume that the available data from the EU is useful for prioritizing types of studies for VCG implementation. Data was retrieved from the ALURES online portal (https://webgate.ec.europa.eu/envdataportal/content/alures/section2_number-of-uses.html, accessed 06/02/2025). The analysis focused on data from the latest reported year (2022) and on species present in the VICT3R database at the time of writing. In ALURES, the term “use” refers to both the first use and any subsequent reuse of animals in research and testing. The 15 most frequently reported use categories and those mentioned in the report are shown in Figure 1, with other categories grouped into “Other studies”. The Sankey plot demonstrates the number of animals used by species (left), their assignment to key regulatory uses (middle), and their allocation to study types (right). Most animals used in the regulatory sector were associated with medical products for human use (500,106), followed by veterinary medicinal products (128,302) and industrial chemicals (123,898). Plant protection products (19,719) made up for a smaller share, while use under other legislative frameworks was considerably lower.

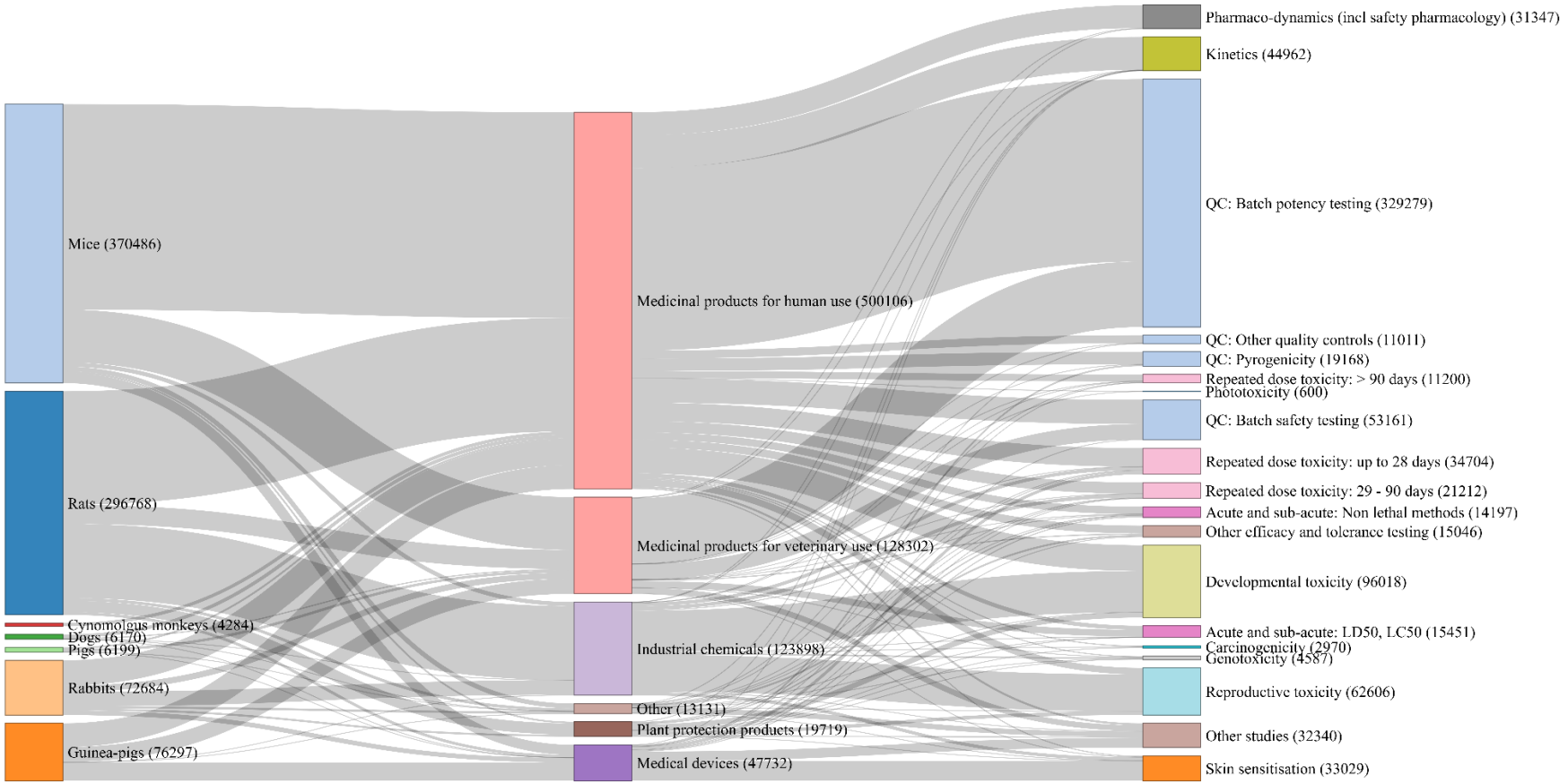

Figure 1: Volumes of regulatory uses of animals across the regulatory sector in the EU in 2022, stratified by legislation and type of study. Data comprises animal species which have been considered relevant for VCG scoping in VICT3R: dogs, guinea pigs, mice, non-human primates (NHPs), pigs, rabbits, and rats (minipigs are not reported). Legislative areas food, feed, biocides and other categories are pooled as Other. Studies with lower yearly uses, with the exception of carcinogenicity, genotoxicity and phototoxicity are pooled as "Other studies".

An analysis of animal use by study type reveals that a substantial number of test animals are used for quality control (QC) purposes. More specifically, batch potency testing accounts for the most extensive QC-related animal use (329,279 animals), followed by animals used in batch safety testing (53,161), pyrogenicity testing (19,168) and other QC procedures (11,011). As explained in more detail below, the suitability of constructing VCGs for these study types remains questionable, also since these studies are currently under discussion to be replaced by in vitro methods and subsequently deleted in the pharmacopeia.

DART studies represent the next largest category, with 96,018 animals used for developmental toxicity testing and 62,606 for reproductive toxicity studies. This is followed by repeated-dose studies, including short-term (34,704), subchronic (21,212), and chronic studies (11,200). Since repeated-dose studies are already the main focus of the project, they fall outside the scope of this deliverable. However, the other study types mentioned here will be examined in the following sections. Though the data offer insights into the animal use numbers, which can help assessing the relevance and feasibility of VCG implementation, they do not provide detailed information on specific study types or designs. For instance, a variety of pharmacology safety studies are grouped together into the "Pharmacodynamics" category, making it difficult to identify the distribution of specific study types there). In addition, it is not known how many uses pertained to control animals. Some study types, such as acute toxicity and kinetic studies, involve large numbers of test animals but are not explored further in this report, as they typically do not use a significant amount of control animals. Finally, while the ALURES database highlights areas of high animal use, feasibility assessments must also consider data availability and potential for standardization, rather than focusing solely on numbers of animal used.

Study designs

1. Developmental and Reproductive Toxicology (DART)

In pharmaceutical drug development, Developmental and Reproductive Toxicology (DART) studies are needed to support clinical trials including WOCBP (Women of Childbearing Potential) for most non-oncology indications. These DART studies assess the effects of a drug on fertility and early embryonic development (FEED), embryo-fetal development (EFD, teratology) and prenatal and post-natal development (PPND, previously referred to as segment I, II and III). For biologics, non-human primates (NHPs) may be required for DART studies because they may be the only pharmacologically relevant species. In such cases, EFD and PPND studies may be combined in enhanced pre-and post-natal development studies (ePPND). The compound attrition rate is relatively high in pharmaceutical drug development due to a multitude of reasons including safety, poor pharmacokinetics, and lack of efficacy to highlight a few. Hence, only a small number of drug candidates reach clinical development phases that require the conduct of DART studies. As a result, companies often take a stage-gated approach to prioritize EFD studies earlier in clinical development and therefore more data may be

available. Most DART studies are conducted only after a compound demonstrates promise in earlier safety and efficacy clinical trials, as these studies are expensive, require specialty expertise and are resource intensive. Since only a few compounds make it to these later clinical stages each year, the number of DART (especially PPND) studies is correspondingly limited even in larger pharmaceutical companies, when compared to repeated-dose general toxicity studies (e.g., 28-day duration studies).

For industrial chemicals, DART testing requirements in the EU adhere to OECD Guidelines and are based on the volume of a substance placed on the market. Starting with a yearly volume of 100 tons, a screening study for reproductive and developmental toxicity is required and hence is the most frequently performed DART study for chemicals. As production volume increases, more DART studies, such as EFD studies in up to two species, are required. Since EU testing requirements usually meet the needs of other jurisdictions, DART studies conducted for EU registrations can serve as proxy for global data availability. For agrochemicals, DART studies are always required and typically comprise EFD studies in two species and a two-generation study addressing fertility and (post-natal) development.

The different requirements regarding study conduct are likely the reason why, relative to other study types, fewer DART studies are carried out in pharmaceutical drug development than for (agro-) chemicals (Figure 1). In 2016, the Scientific and Regulatory Policy Committee of the Society of Toxicologic Pathology published the results of a survey sent to pathologists and reproductive toxicologists to understand member practices for the inclusion of reproductive end points in general toxicity studies and pathology endpoints in reproductive and juvenile toxicity studies during drug development (Halpern et al., 2016). There were 100 respondents representing industry, CROs, government, academia and consultants. The findings revealed a high degree of variability regarding the inclusion of specific reproductive end points, as these endpoints are often approached on a case-by-case basis. Despite this variability, the creation of a VCG for DART studies remains valuable, given the high number of animals used per study. Nonetheless, the availability of sufficient numbers of studies with a core set of standardized data is a key factor that needs to be considered when assessing the feasibility of developing a robust VCG.

1.1 Fertility and Early Embryonic Development (FEED)

FEED studies are usually conducted only in a single species (in the large majority the rat). The relevant ICH S5 (R3) guideline describes diverse possible modifications of the study design, but the most relevant differences would be the inclusion of both sexes or testing only one sex in the dedicated study and the duration of the premating treatment duration leading to differences in the age of the animals at start of mating and during gestation. FEED studies in species other than mice and rats are not easily available, making the application of VCG highly challenging due to limited data availability.

In safety testing for industrial chemicals and agrochemicals, pure FEED studies are of limited relevance. Instead, the studies required by regulations in these sectors combine fertility and postnatal development. Such studies typically involve an administration phase lasting between 2 to 10 weeks prior to mating, followed by ongoing treatment of the mother animals until at least postnatal day 14, and in some study designs extending it into adulthood or a second mating. The rat is virtually the only species used. Two key variable parameters concerning the application of VCG methodologies for these types of studies are the varying starting ages of the animals, which can depend on the duration of the

pre-treatment, and the increasing emphasis on endocrine endpoints. This heightened focus is likely to introduce new evaluated parameters that may limit the availability of historical control data. These fundamental differences in prevalent study designs between drug development and (agro-)chemicals likely impedes cross-usability of control animal data between these sectors. In addition to study design parameters of general relevance for VCG construction (e.g., vehicle, strain, age), the following parameters are examples for being considered relevant especially for FEED studies: pre-mating treatment duration, duration and method of estrous cycle determination, sperm evaluations.

At the moment no SEND standard for the special endpoints of FEED studies is available (estrus cycle, mating) but CDISC is actively developing an update for the SEND DART IG that will include fertility studies and also allow for modelling intermediate study designs.

Considering availability of FEED studies, 9/14 entities answered that <5 studies per year would be available, but suitability to use VCG was ranked above all other study types.

1.2 Embryo-fetal developmental studies (EFD)

In the pharmaceutical industry, EFD studies are usually needed to enter Phase 2 clinical trials, as otherwise inclusion of WOCBP would not be possible. This means that these studies again are less frequently conducted compared to short-term general toxicity studies, explaining the relatively low number of available studies by company.

EFD studies for pharmaceuticals usually are conducted in a rodent (mostly rat) and a non-rodent (mostly rabbit) animal model. In these studies, only mated females are included, which might have been mated either at the breeder or on-site at the testing facility. Another relevant difference between laboratories is the duration of treatment during organogenesis, which also differs from practices in chemical industry, where treatment extends until hysterectomy. It is also necessary to differentiate between dose range finding studies in pregnant animals, preliminary EFD (pEFD) studies and the “pivotal” EFD study. While not all endpoints are included in every design, many endpoints are captured in almost every study, like body weight development of the dam during pregnancy and number and status at implants at hysterectomy. Pivotal EFD studies usually include at least 16 litters in the control group, resulting in a substantial amount of data, with at least 200 fetuses available per study. However, it is not standard practice to photograph fetuses, and under current preservation approaches limit their usable lifespan. As such, the data available to the VGC related to the subjective detailed macroscopic observation of the fetus realistically would need to rely on the original evaluation (which corresponds to historical control data) instead of re-evaluating preserved specimens.

EFD studies in species other than rodents and rabbits are relatively rare. Nonhuman primate studies mostly are conducted as enhanced pre- and postnatal development studies (ePPND) reducing the availability of pure embryo-fetal data for monkeys. Minipigs are also mentioned as alternative species in ICH S5, but their use remains limited, and differences in study design (especially early or late termination in the fetal phase) may pose challenges for cross-laboratory comparability.

Since March 2023, EFD studies must be submitted with a mandatory SEND package, so all requirements to model the data in SEND are available. Yet, some additional metadata parameters may need to be considered for construction of VCGs. For example, the location of mating, which was already suggested to CDISC for inclusion in future revisions of the SEND implementation guide (SENDIG).

Treatment duration and the need to conduct a double staining for skeletal examination are the main differences for embryo-fetal studies conducted in the chemical compared to the pharmaceutical industry. In general, high consistency between studies for many endpoints can be expected. Implementing VCGs in EFD studies has the potential to significantly reduce control animals, thereby lowering the number of animals used. An added value of focusing on EFD studies is that the evaluated parameters in the definitive study often fully include the parameters of the DRF, making it feasible to establish a larger database for VCG for DRFs. Additionally, these DRF studies usually do not take place under Good Laboratory Practice (GLP), which greatly facilitates implementation. But even if VCG implementation in EFD studies is ultimately deemed not feasible, collecting data from EFD studies for different laboratories and building a single data base with a clear presentation of results would be a great leap towards increased reliability of these highly relevant endpoints (Kluxen and Hothorn, 2021; Wise, 2025).

Survey data on the availability of EFD studies revealed that 2 entities reported having no studies of this type, while 6 reported <5 studies per year, 4 reported 5-10 studies and 2 reported 10-50 studies. This translates to at least 46 available studies per year amongst survey respondents. A very conservative estimation suggests this corresponds to at least 300 dams and more than 2000 fetuses in the control groups alone. In the survey, EFD studies were ranked second in terms of suitability for VCG implementation.

1.3 Prenatal and Postnatal Development (PPND)

Prenatal and postnatal developmental studies are usually conducted in rodents (mostly rats) and for pharmaceuticals are only needed to support the marketing application of a new drug. The studies involve treating pregnant females (F0) during gestation, allowing them to produce litter and continue treatment until the end of lactation. As a result, F1 offsprings are exposed to the test compound both in the uterus and via milk after birth. For the F1 animals, key developmental markers, such as eye opening, are recorded, along with examinations of memory and sensory functions. Part of the F1 animals is raised to adulthood, mated and examined in a similar fashion as in a FEED study. The design and execution of PPND studies are complicated and have not yet been covered by a SEND implementation guide. In regulatory safety testing for industrial chemicals and agrochemicals, standalone PPND studies are of limited relevance, as these studies are typically incorporated into a design that also addresses fertility.

Survey data demonstrated that 9/13 entities reported conducting less than 5 PPND studies per year. In terms of suitability for VCG implementation, PPND studies were ranked at the 4th position.

2. Juvenile Studies

Juvenile studies are generally not mandatory unless a drug is being developed for a pediatric population. In general, these studies are conducted only if relevant effects have been observed that need further clarification in a dedicated developmental study in young animals. Study designs are therefore very variable and may include biomarker assessments, histopathology or serial behavioral testing. For (agro-)chemicals generally juvenile studies are not performed. Considering that juvenile studies are defined by the developmental stage of the animals at start of dosing in a study (Kim et al., 2017), a relevant proportion of "general toxicity" studies already available in eTRANSAFE might also be used as basis for comparison of juvenile studies to adolescent animals. Rodent studies often use animals in around puberty, while in dog and non-human primate (NHP) studies, sexual maturity is also often not reached at study start. For earlier treatment start ages, such as pre-weaning, the situation is different, and litter effects have to be included in the assessment.

A publicly available database collecting developmental endpoint data from different species and stains would be beneficial for design and interpretation of these often-complex studies.

The survey responses indicated that 10/14 entities proposed availability of <5 studies per year, while the remaining entities apparently do not perform these studies. In terms of suitability for VCG development, juvenile studies were ranked at position 6th.

3. Carcinogenicity

Overall, only a small number of compounds that reach clinical development phases require carcinogenicity studies. However, mouse and rat carcinogenicity studies may be good candidates for inclusion in a virtual control database. For industrial chemicals this study is rarely performed, while it remains a requirement for registration of pesticidal active ingredients in major economic areas. For pharmaceuticals, the ICH S1B(R1) addendum provides the possibility of a waiver for the 2-year rat carcinogenicity study in rats, where a weight-of-evidence approach demonstrates that it would not add value to human risk assessment. Whilst this approach is not always applicable, implementing a virtual control group may contribute to a reduction in the overall number animals in 2-year rat carcinogenicity studies performed, compared to previous years. Carcinogenicity studies often follow standardized protocols such as those recommended by regulatory bodies. These protocols usually involve large numbers of animals to ensure robust historical control data on tumor incidence and progression. This property makes them well-suited for integration into a virtual control database. Carcinogenicity studies may further appear an appealing study type in this project, as procedures for data conversion and curation can largely be adopted from the repeated dose studies for which the VCG concept is already in a later phase of development. Yet, the large number of animals used in carcinogenicity studies could provide a challenge for the creation of a VCG repository of digital whole slide images (WSI), which are considered particularly important here due to the high significance of the histopathological evaluations. Currently, primary microscopic reads and peer reviews of carcinogenicity studies digitally are not widely available due to the sheer size of the WSI image files and infrastructure challenges that impede a quick and smooth user experience for the pathologist.

The similarity between carcinogenicity and repeated dose studies likely explains why this study type received a high ranking in the survey in terms of their suitability for VCG development. However, when compared to other study types, the relatively low number of yearly (control) animals used suggests a lower priority of exploring VCGs in carcinogenicity studies.

4. *In vivo* genotoxicity (e.g., micronucleus or comet assay)

According to the survey, *in vivo* genotoxicity studies are conducted in relevant amounts in the pharmaceutical industry. The basic design of such a study is very similar to that of a repeated dose study but frequently performed with a shorter exposure time. The main difference is that after sacrifice more specific investigations regarding genetic damage are carried out in organs believed to be susceptible for insult by the test item, most commonly the liver and bone marrow. This leads to the situation that this type of study is often conducted as a side arm or even just as an additional investigation in a repeated dose study. And even though some survey respondents indicate that they consider *in vivo* genotoxicity studies to be a worthwhile and feasible target for VCG, this characteristic of the study design likely results in a small database of remaining control animals to construct VCGs, as soon as influencing factors are matched. In safety testing of (agro-) chemicals, this study type plays a minor role, as it is not a standard data requirement. For those cases where an *in vivo* genotoxicity study is performed, the study design is highly variable with different combinations of investigated organs and detection methods. Hence, application of the VCG concept faces the same challenges as in the pharmaceutical sector. The survey results for *in vivo* genotoxicity studies showed varied rankings regarding their suitability for the VCG approach. While some respondents found these studies highly feasible for VCG integration, others were uncertain or considered them only “somewhat feasible”. This variability likely stems from differences in study design and data format, which can impact VCG compatibility. The responses regarding the frequency of study conduct were heterogeneous. Most respondents conduct none or fewer than 5 studies per year, while 2 participants reported more than 5 studies annually, and even 4 survey participants conduct between 10 to 50 studies each year. Although genotoxicity studies have potential for VCGs, their prioritization may be lower compared to more standardized study types due to these challenges and the moderate frequency of such studies within organizations.

5. Safety Pharmacology Studies (CNS, Resp, CV or Neuropsychopharmacology studies)

For the pharmaceutical industry, Safety Pharmacology (SP) studies are essential during the early stages of drug development typically required before progressing to First-in-Human or Phase 1 clinical trials. Still, SP studies are generally less frequent than routine general toxicity studies due to their specific focus on core physiological systems like the cardiovascular, central nervous, and respiratory systems. SP studies commonly employ both rodent and non-rodent animal models. Rats and mice are the most frequently used rodent models, while dogs and NHPs are often chosen as the non-rodent species depending on the pharmacological target and intended therapeutic use. The design of SP studies and the endpoints vary by system under evaluation. For example, telemetry studies for cardiovascular assessment or behavioral and reflex tests for central nervous system (CNS) evaluation. Endpoints in SP studies typically include metrics like QT interval prolongation (cardiovascular), respiratory rate and

tidal volume (respiratory), or motor activity and coordination (CNS). Relevantly, SENDIG 4.0, expected in 2026, is expected to introduce standardized terminology and data structures for reporting findings related to the nervous system in nonclinical studies. The new Nervous System (NV) domain will require the inclusion of neurobehavioral endpoints, which are expected to cover both central and peripheral nervous systems, including a broader range of behavioural tests beyond just motor activity and coordination. Currently, an NV domain can be created, but it would need to be a custom domain. The CDISC team is open to share updates on this domain, which will help guiding the curation of datasets in alignment with the upcoming SEND guidelines and the future development of VCGs.

According to the survey, and as anticipated, SP studies are among the most frequently conducted new study types considered. Similarly, the availability of data reported by companies in the VICT3R consortium is relatively high, with most respondents indicating they have between 10 and 50 or more studies available. Additionally, most respondents expressed a willingness to share their control data.

Notably, this study type had the highest number of respondents reporting that their data is already available in SEND or SEND-like formats, with a smaller proportion using Excel/CSV formats, and only a few reporting data stored in PDF format. This demonstrates a significant degree of standardization and readiness for data integration in SP studies compared to other study types. Yet, it is also common to use a Latin Square design for dosing (i.e. each animal receives each of the dose levels throughout a study), which is difficult to model in SEND.

The perceived potential of VCGs for SP Studies is promising, ranking 5th among the study types evaluated in the survey. It is exceeded only by "Fertility and Early Embryonic Development," "Developmental Toxicity/Teratology," and "Carcinogenicity Studies" but remains strong and comparable in perceived value relative to these higher-ranked study types. Within the consortium, expert opinions on the feasibility of implementing VCGs for SP studies are predominantly positive. Most experts describe the approach as “somewhat feasible” or “highly feasible,” with only a small fraction rating it as “not feasible.” A minority of respondents remain “unsure,” with this group being larger than those deeming the approach "not feasible" but smaller than those expressing high confidence in its feasibility.

The diversity of experimental designs in SP studies presents both opportunities and challenges for creating effective control groups. Some respondents highlighted that SP studies typically do not include separate control animals, suggesting that the use of virtual controls in these studies may not be appropriate or necessary. Implementing VCGs in SP studies would likely require analysing experimental design variabilities, which could further subdivide this study type into subclasses with differing potentials for VCG generation. Identifying and addressing these variations will be crucial for tailoring virtual control strategies to the specific needs and designs of SP studies. This must be complemented by standardizing data collection and reporting across laboratories. In the case of CNS related measurements metadata is extremely important and should be meticulously standardized and reported to capture the expected variability in relevant neurobehavioral endpoints. This will ensure accurate data integration, comparability across studies, and alignment with regulatory guidelines, facilitating the effective use of VCGs in CNS studies.

6. Xenograft studies

Xenograft tumor model studies were defined as the implantation of tumors, most commonly cell line-derived and patient-derived, into immunodeficient mice and for use in cancer research to test therapy efficacy. These study types are not ideal candidates for a VCG due to several factors. A significant challenge to experimental standardization for VCG generation is tumor engraftment variability which arises from a combination of factors related to the tumor cells themselves, the overall experimental conditions and the health and immune status of the mouse model. Different tumor types (e.g., breast cancer, melanoma, glioblastoma) and their genetic characteristics can impact their growth in xenograft models. Additionally, the establishment of subclones during the passage of cell lines *in vitro* can occur due to several factors related to the growth environment, selection pressures, and inherent genetic heterogeneity of the cell population. As a result, the same cell line originally obtained from a commercial source will develop differing tumor growth characteristics in different labs and even with extended passage within a single laboratory.

Preparation of the cells, the number of cells implanted or the size of the tumor fragment, the implantation site (subcutaneous, orthotopic (in the tissue of origin), or intravenous), use of Matrigel or other matrices can influence how well the tumor takes and grows in the host.

The strain, age and specific pathogen free status of the immunocompromised mice used can significantly affect engraftment and tumor growth. Different strains of immunodeficient mice (e.g., SCID, NOD-SCID, athymic nude mice) have varying levels of immune suppression, which impacts tumor engraftment and growth kinetics. The age, general health/microbiome, and physiologic stress status of mice can influence the immune response to tumor cells.

These factors, just to name a few, make it difficult to standardize xenograft model studies and create challenges in capturing the totality of the metadata relevant to xenograft studies to generate a robust virtual control database.

As a result, despite the frequent conduct and easy accessibility of these studies, the survey results indicated that they are not perceived as suitable for implementing VCGs (ranking last of all predefined study types in the survey).

7. Other study types

Study types in this section were not included in the design of the survey but were mentioned by individual participants in open text fields or encountered during the research for the preparation of this report. They are briefly described here but cannot be addressed in as much depth as the other study types that were explicitly within the scope of the survey and the discussions held during the work package meetings.

7.1 *In vivo* phototoxicity

For pharmaceuticals, where a molecule has previously tested positive for phototoxicity in the *in vitro* 3T3 Neutral Red Uptake phototoxicity study, a follow-up *in vivo* phototoxicity study may be required. There is currently no defined standardized approach to study design for *in vivo* phototoxicity studies, and a best practice approach is taken where this study is warranted. A vehicle control group may be included in the study design to aid study interpretation. A non-irradiated control group may also be included, in which animals will receive the test item but will not be exposed to UV light. This latter control group would not be suitable to the VCG concept. Based on the non-standardized study design and the low number of animals used annually for this study type, this study type is considered a lower priority item.

7.2 Inflammation efficacy

Efficacy studies using mouse models of inflammatory disease may be a good candidate for creating a virtual control group database. Inflammation-based mouse models often have standardized protocols and often reproduce similar disease outcomes, making them more predictable and suitable for a virtual control database. Some commonly used examples of these models include

- Dextran Sulfate Sodium (DSS)-Induced Colitis is used as a model of ulcerative colitis, a form of inflammatory bowel disease.
- Collagen-Induced Arthritis (CIA) is used to study autoimmune arthritis, resembling rheumatoid arthritis in humans. Mice are immunized with type II collagen, leading to the production of autoantibodies and subsequent joint inflammation.
- Bleomycin-Induced Pulmonary inflammation is used to study lung fibrosis and inflammation, particularly in the context of idiopathic pulmonary fibrosis (IPF).
- Oxazolone-induced atopic dermatitis (AD) model is a topical hapten mouse model that mimics human AD symptoms.
- Choline-deficient, high-fat diet (CDAHFD) in mice can cause liver disease, including fatty liver, non-alcoholic steatohepatitis (NASH) and fibrosis.

Though there is a wealth of data available from various studies using these mouse models, including baseline inflammation markers, gene expression profiles, and histopathological features, these endpoints are variably applied across studies. Building such a database would require standardization of a minimum set of metadata related to induction of the model, ensuring that VCG data is robust in terms of experimental variables, methods, and outcomes.

7.3 Quality control studies (QC)

QC studies were not explicitly mentioned as part of the survey and were also not named by any survey respondent. However, this study type is addressed here due to the sheer number of animals used (Fig 1). Even though this category is not very well-defined, we assume that a large part of these studies can be attributed to the efficacy and safety tests for vaccines. These studies usually include a relevant

number of control animals. However, they still do not seem to be a good target for VCG implementation, as there is a consensus that these animal studies will be phased out and replaced by animal-free alternatives in the near future. The project team members in this work package have only limited experience with this study type, which has resulted in a brief evaluation at this stage that may not provide the most accurate representation.

7.4 Ecotoxicological studies

The focus of our activities lies on mammalian studies. However, a significant number of vertebrates are also used in ecotoxicological studies. The endpoints examined in these studies are generally less complex than those considered in mammalian studies. This makes ecotoxicological studies a worthwhile target for VCG. Furthermore, these studies are expected to play a larger role in the future for the registration of pharmaceuticals. According to our information, the data are generally available in less harmonized formats, and it is more common to collect the data outside of LIMS systems. This presents a challenge for the evaluation and generation of VCG. Additionally, there is currently a lack of experts in the consortium. Nevertheless, we believe it is worthwhile to examine these types of studies more closely and plan to onboard experts from beneficiaries to support us in this work.

Conclusion

Based on the survey responses, animal usage statistics, and discussions among the members of WP3, we have identified varying levels of suitability for the study types we examined.

DART studies generally received high rankings across all evaluated criteria. Within the DART category, EFD studies are considered the most suitable due to their relatively frequent execution, significant animal usage, highly standardized study designs, and comprehensive coverage by the SEND standard. Survey respondents also indicated the highest willingness to donate data for this study type. FEED studies were rated even slightly higher in terms of their potential for VCG by survey participants along with a promising interest for data sharing. However, this particular study type exhibits less uniformity in design across different industries, and essential parameters are not yet incorporated into the SEND guidelines, although updates are anticipated. Conversely, PPND studies appear to be the least suitable among the DART studies evaluated.

Carcinogenicity studies represent a unique situation, as the data gathered from these studies aligns well in terms of type and terminology with the SEND standard for repeated dose studies. As a result, the SEND format is widely utilized in this context, leading to a high ranking of carcinogenicity studies within the consortium in terms of both suitability and feasibility. However, carcinogenicity studies are conducted relatively infrequently, which is why available data is limited, and, above all, the potential for reducing animal usage in these studies is lower compared to other study types.

SP is also regarded as a promising area for the implementation of the VCG approach. A significant number of SP studies are conducted, and there is a strong willingness among consortium members to share data. Additionally, SP studies demonstrate a high level of data standardization, as most datasets

are already formatted in SEND or SEND-like formats. Expert opinions within the consortium further support this view, with the majority of members categorizing the VCG approach as either "somewhat feasible" or "highly feasible" for this study type. Overall, there is considerable confidence in the potential for integrating VCGs into SP studies.

References

Halpern WG, Ameri M, Bowman CJ, et al. (2017). Scientific and Regulatory Policy Committee Points to Consider Review: Inclusion of Reproductive and Pathology End Points for Assessment of Reproductive and Developmental Toxicity in Pharmaceutical Drug Development. Toxicologic Pathology. 2016;44(6):789-809. doi:10.1177/0192623316650052

Kim NN, Parker RM, Weinbauer GF, Remick AK, Steinbach T (2017). Points to Consider in Designing and Conducting Juvenile Toxicology Studies. International Journal of Toxicology. 36(4):325-339. doi:10.1177/1091581817699975

Kluxen FM, Weber K, Strupp C, Jensen SM, Hothorn LA, Garcin JC, Hofmann T (2021). Using historical control data in bioassays for regulatory toxicology. Regul Toxicol Pharmacol. 125:105024. doi: 10.1016/j.yrtph.2021.105024

Taylor K (2024). Trends in the use of animals and non-animal methods over the last 20 years. ALTEX. 41(4):503-524. doi: 10.14573/altex.2410111

Wise, LD (2025), The 6Rs of EFDT Studies. Birth Defects Research, 117: e2429. https://doi.org/10.1002/bdr2.2429