



Food for Thought ...

Introducing the Concept of Virtual Control Groups into Preclinical Toxicology Animal Testing

Thomas Steger-Hartmann¹, Annika Kreuchwig¹, Lea Vaas¹, Jörg Wichard¹, Frank Bringezu², Alexander Amberg³, Wolfgang Muster⁴, Francois Pognan⁵ and Chris Barber⁶

¹Bayer AG, Pharmaceuticals, Investigational Toxicology, Berlin, Germany; ²Merck Healthcare KGaA, Chemical & Preclinical Safety, Darmstadt, Germany; ³Sanofi, Preclinical Safety, Frankfurt, Germany; ⁴Roche Pharmaceutical Research & Early Development, Pharmaceutical Sciences, Roche Innovation Center, Basel, Switzerland; ⁵Novartis Institute for Biomedical Research, Basel Switzerland; ⁶Lhasa Ltd. Leeds, UK

Abstract

Sharing legacy data from *in vivo* toxicity studies offers the opportunity to analyze the variability of control groups stratified for strain, age, duration of study, vehicle, and other experimental conditions. Historical animal control group data collected in a repository could be used to construct virtual control groups (VCGs) for toxicity studies. VCGs are an established concept in clinical trials, but the idea of replacing living beings with virtual data sets has so far not been introduced into the design of regulatory animal studies. The use of VCGs has the potential to reduce animal use by 25% by replacing control group animals with existing randomized data sets. Prerequisites for such an approach are the availability of large and well-structured control data sets as well as thorough statistical evaluations. The foundation of data-sharing has been laid within the Innovative Medicines Initiatives projects eTOX and eTRANSafe. To establish proof of principle, participating companies have started to collect control group data for subacute (4-week) GLP studies with Wistar rats (the strain preferentially used in Europe) and are characterizing these data for its variability. In a second step, the control group data will be shared among the companies, and cross-company variability will be investigated. In a third step, a set of studies will be analyzed to assess whether the use of VCG data would have influenced the outcome of the study compared to the real control group.

1 Introduction

Virtual control groups (VCGs), also called synthetic control arms, represent a concept that is well established for randomized clinical trials (Berry et al., 2017). VCGs are used to compare an experimental treatment with the standard of care. The main purpose of using virtual participants or patients instead of true patients is to increase the speed of the trials while avoiding costly recruitment of probands, thus allowing all recruited patients to receive the new treatment (ethical considerations). The concept is particularly used in clinical cancer studies and assumes that if the range of variables is kept sufficiently constant between the patients foreseen for a new experimental treatment and those who have received the standard of care, the outcome of the clinical trial can be assessed based on comparison between the virtual control arm and the patients treated with the new experimental design (Switchenko et al., 2019).

In the clinical setting, such virtual control arms can be constructed in several ways. The simplest approach is the use of historical control data for the respective endpoint under examination (Berry et al., 2017); this is also described in the International Council of Harmonisation (ICH) guideline E10 (ICH, 2001). However, differences in patient selection can easily confound the findings in the comparison of new treatments with historical controls. To overcome this problem, VCGs also can be constructed based on the clinical characteristics of the chosen patients using statistical models such as Bayesian approaches (Spiegelhalter, 2004) or nomograms that, for example, estimate progression-free survival under standard of care in oncology trials (Jia et al., 2014). A third approach to constructing VCGs is based on the use of electronic health records (EHR), i.e., the systematized collection of electronically-stored health information on patients and populations in a digital format (Eichler et al., 2016).

Received December 4, 2019; Accepted February 19, 2020;
Epub February 25, 2020; © The Authors, 2020.

ALTEX 37(3), 343-349. doi:10.14573/altex.2001311

Correspondence: Thomas Steger-Hartmann, PhD
Bayer AG, Pharmaceuticals, Investigational Toxicology,
Müllerstraße 178, 13353 Berlin, Germany
(thomas.steger-hartmann@bayer.com)

This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 International license (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is appropriately cited.



Although the use of historical data is well established in pre-clinical studies for comparative purposes (Yanagawa and Hoel, 1985; Haseman, 1992), the concept of VCGs actually replacing control group animals has not yet entered into regulatory toxicity testing because, compared to the recruitment of patients, the availability of control animals is not a technical issue and, in comparison to clinical studies, the cost and time factor do not play as important a role in the conduct of animal studies. However, ethical considerations, i.e., the drive to reduce animal numbers in preclinical research, have triggered consideration of reducing or replacing control groups. In the area of *in vivo* genotoxicity studies, Pfuhler et al. (2009) investigated the possibilities and acceptance of reducing the size of positive control groups or completely omitting concurrent positive control groups with reference to historical positive control data. For pharmacological studies, Kramer and Font (2016) have proposed strategies for reducing control group sizes in animal studies by incorporating historical control group data. Despite these efforts, replacing control groups in systemic toxicity studies has not yet been approached.

The conventional setting of a regulatory toxicology study uses 25% of the animals as controls: OECD TG 407 “Repeated Dose 28-Day Oral Toxicity Study in Rodents” (OECD, 2008a) may be cited here as an example for subacute studies, though the pertinent guidelines for chronic studies or carcinogenicity studies essentially follow the same scheme (OECD, 2008b), with the exception that the animal numbers per group increase with the duration of the study. TG 407 requires at least three dose groups and one control group. Animal numbers in these four-week rodent studies are normally 10 per dose group or control (five females and five males) and may be greater if interim euthanasia or recovery groups with satellite animals are planned. Four-week rat studies for food ingredients or pharmaceutical compounds usually apply higher numbers of animals, i.e., 10 per sex and dose (FDA, 2003a). Numbers may be even higher if recovery groups, which then usually also have parallel control groups, are included in the study design. For non-rodents, three to four animals per sex and dose group are used in four-week studies (FDA, 2003b). In chronic studies, animal numbers for rodents might increase up to 50 rats per sex and dose group and 10 dogs or primates per sex and dose group (Gad, 1995).

2 Current use of historical control data in preclinical animal studies

None of the above-mentioned guidelines foresee the use of VCGs, though it is advised that “*historical control data are collected and that for numerical data, coefficients of variation are calculated*” (OECD, 2008a). The main purpose of this data collection is the performance control of the study and the assessment of outliers, which may occur in individual studies for various reasons. Legacy data from control animals are used to determine the range of parameters of untreated animals, its changes over time or the influences of changes in analytical methods.

Petterino and Argentino-Storino (2006) list three applications of historical control data in the area of clinical pathology and hematology:

- 1) Evaluation of the clinical and hematological pathology data
- 2) Evaluation of background pathology in animal populations
- 3) Comparison of analytical methods

If a statistically significant difference between a dose group and the control group is observed in a study for a specific parameter but the changes in the treated group lie within historical control ranges, then it is questionable whether the observation actually represents a compound-related effect.

Historical control data are of particular importance for the evaluation of carcinogenicity studies with respect to incidences of spontaneous tumors observed, which depend on the species and the strain used (Morawietz et al., 1992). In addition to the direct comparison with the control group, such a comparison with historical control data allows the assessment whether the occurrence of a rare tumor or a marginally increased tumor incidence is of biological relevance, i.e., caused by the chemical under investigation (Yanagawa and Hoel, 1985; Haseman, 1992; Greim et al., 2003). For the assessment of developmental toxicity studies, the situation regarding historical control data is similar compared to carcinogenicity studies. The data collection on fetal findings, including incidences of spontaneous external, visceral, and skeletal anomalies, contributes to the differentiation between spontaneous findings and true toxic effects, particularly for rare findings (Kuwagata et al., 2019).

In summary, the value of historical control data has been referenced for decades, however, a replacement of control group animals with virtual animals based on existing data has been proposed only recently for animal disease models with surgical interventions (Kramer and Font, 2016, 2017). The concept has, however, not found broad repercussions in regulatory safety testing yet, probably also because of the lack of appropriate large and easily accessible control animal data sets.

3 Challenges for the use of VCGs

There are several reasons why existing data from control group animals cannot be easily used for constructing VCGs. The first reason, which was broadly discussed in a publication on historical data use for assessing carcinogenicity data (Greim et al., 2003), is the heterogeneity of the data sets. The physiology of experimental animals and, thus, the data acquired during animal studies are influenced by numerous factors such as genetic disposition, strain, animal housing conditions, diet, stress during handling and administration, age of the animals, and infections.

Some of these factors are under the influence of the investigator and thus can be controlled, and are also defined by the guidelines or related good practice documents. Such controllable factors are strain, age, diet, and housing conditions (temperature, lighting, humidity, bedding, single or group housing). Others are less amenable to standardization, such as stress during animal handling or infections occurring during the course of the study.

An important aspect, which can only be controlled by stringent breeding programs (Low-Marchelli, 2017), is genetic drift in the inbred animals used in toxicity studies. Genetic drift may remain silent and go undetected until a new phenotype is identified by the analytics used in the studies, including histopathology. However, it represents a cause for irreproducibility of experimental results (Brekke et al., 2018). As long as control animals are used, these underlie the same level of genetic drift. A replacement by virtual controls using historical data might result in differences between the treatment groups and the VCGs. These might falsely be attributed to a test compound effect though they were rather caused by genetic differences between the previously used animals and those in the current study.

A further reason for using control animals is the detection of spontaneous infections, which may influence hematological and histopathological parameters of the affected animals (GV-SOLAS, 1999). Without a control group, such infections might erroneously be attributed to effects elicited by a test compound. In a more general sense, the use of control groups in this context also allows assessment of the interaction between animal behavior and the laboratory environment, which was identified to represent a major contribution to variability of results between test sites (Crabbe et al., 1999).

The third reason is the interlaboratory and time-dependent variability of analytical methods used for clinical chemistry evaluations, which constantly undergo changes due to improvements of the applied analytical technologies (e.g., reduced sample volumes, improved detection limits, changed assay composition). These are usually tracked within the laboratory by using quality control samples and are also evaluated by ring tests. However, such changes limit the applicability of control group data to certain time periods for which identical methods and biochemical assays (“kits”) have been used.

A fourth reason relates to non-numerical, descriptive findings gathered during animal studies such as gross pathology and histopathology. In order to make these findings interoperable, comparable and compatible between studies, there is a need for controlled vocabularies and ontologies (Hardy et al., 2012). Despite tremendous efforts by the Society of Toxicologic Pathology to standardize the vocabularies and terminologies through the International Harmonization of Nomenclature and Diagnostic Criteria (INHAND) (Creasy et al., 2012), the heterogeneity of terms for describing similar or identical findings is still an issue. This was also addressed in the IMI eTOX project by setting up ontologies and controlled terminologies for histopathology (Briggs et al., 2015). Standardization and harmonization of terminologies is of particular importance since, in contrast to numerical parameters, the original observation in an animal study, i.e., the microscopic slide, is usually not stored as an image in databases, thus preventing a rapid automated re-assessment. Complex guidelines have been developed for histopathology peer review of toxicology studies (OECD, 2014) to assure the quality and the accuracy of interpretation, but these regulations do not solve the issue of varying terminologies. The implementation of the Standard for Exchange of Nonclinical Studies (SEND) developed by Consortium of Standards in the Clinical

Research Process (CDISC, 2016) has triggered further harmonization of terminologies for toxicology studies in the pharmaceutical industry. This standard is mandatory for repeated-dose toxicity studies used in new drug applications (NDAs) at FDA since December 2016 (FDA, 2014) and is being gradually extended to other types of toxicity studies.

The fifth and probably most important aspect is the lack of appropriate large data sets and repositories for control animal data. Study centers, including contract research organizations, usually set up their own inventories for the collection of historical data, but these data sets are not broadly shared, except for above-mentioned purposes in the assessment of incidences of spontaneous neoplasms (Registry of Industrial Toxicology Animal data [RITA] or the North American Control Animal Database [NACAD]) (Morawietz et al., 1992; Greim et al., 2003) or developmental findings. Most of the reference data sets for physiological parameters are published in the literature (Loeb and Quimby, 1989) or in disparate databases with rather limited query functionalities, thus limiting their usability.

4 Steps towards the implementation of VCGs

4.1 Data and data repository requirements

Some of the hurdles mentioned above already have been addressed in the context of historical data comparisons. Greim et al. (2003) lists requirements that need to be fulfilled by historical data sets in order for them to be eligible for comparative re-use in the context of carcinogenicity assessments:

- 1) The historical data were obtained with the same species and strain of experimental animal. The animals were acquired from one supplier only.
- 2) The historical control data were produced in the same laboratory as the experimental data.
- 3) The study design, experimental methods and assessment criteria were the same, i.e., parameters such as age of the animals at the beginning of the experiment, animal housing conditions, methods of obtaining samples (e.g., number, size, localization and orientation of the tissue sections), diagnostic criteria (e.g., standardized terminology, peer review of critical findings by expert pathologists) need to be assessed and controlled for.
- 4) The studies used as historical controls were carried out during a defined time window to limit the variability of the parameters.

With exception of requirement no. 2, these criteria can also be applied to the VCG approach. The idea of VCGs is based on large, shared data sets from toxicology studies for regulatory dossiers across industry. The shared data should have the highest possible level of granularity (individual animal data for all parameters acquired during a study plus meta-information on how the data was measured and assessed) to allow analyses on whether certain data sets are comparable across studies, laboratories and over specific time periods.

Such a control animal data collection could thus not only be used to reduce animal numbers in control groups, but also to in-



vert some of Greim's criteria by applying rigorous statistical procedures: If a data set from an individual animal falls within pre-defined statistical limits, it can be used for a VCG, independent of whether the data was produced in the same laboratory or using the same procedure, hereby partly overcoming the described problem of statistically "under-powered" animal studies (Hartung, 2007).

Prerequisite for the concept of VCGs is the availability of a repository for shared data, which is accessible not only to the companies providing data, but also to institutions assessing the data, such as regulatory authorities, or the scientific community. The infrastructure and the legal framework for sharing toxicity data already has been established in the recent IMI project eTOX (Sanz et al., 2017). Since the SEND implementation with its standardized terminologies came into force towards the end of the project, the follow-up IMI project eTRANSafe¹, which started in 2017, now focusses on data-sharing using these new standards.

eTRANSafe is a five-year project, funded by the IMI-2 Joint Undertaking together with the pharmaceutical industry. It aims to develop an advanced data integration infrastructure together with innovative computational methods to improve the safety in drug development processes. Key component of the data sharing activities is the so-called honest broker, who receives and protects the data from the participants and generates as well as manages the eTRANSafe data warehouse. The honest broker also administers the access rights to the data base and ensures the separation of confidential from non-confidential shared data. This separation is of importance since the complete shared study may contain data that is considered sensitive information by the providing company, particularly with regard to compound information (e.g., structure, target or indication), whereas the control group animal data requires no intellectual property protection, i.e., can be shared freely among all participating partners and potentially also with future users not belonging to the project, including regulators or the scientific community.

It is important to keep in mind that for the ultimate implementation of the VCG concept, the data repository needs to fulfill certain criteria for computer system validation. Regulatory animal studies are performed under Good Laboratory Practices (GLP). Therefore, the use of historical control data to construct the VCG underlies the same principles, namely the requirements set forth in the advisory document "Application of GLP Principles to Computerised Systems" (OECD, 2016).

4.2 Statistical procedures

The feasibility of the proposed concept will be assessed by the participants of the eTRANSafe project in a stepwise approach. Initially, the companies will analyze their internal historical control repositories and subject these to internal statistical analyses. Key components are descriptive statistics on data distribution and general characteristics like those that have previously been published for the inclusion of historical control data for the mi-

Tab. 1: List of minimum parameters to be collected and analyzed in the context of the VCG concept

Parameters
– Strain, supplier
– Age
– Sex
– Weight
– Study start
– Study duration
– Route of administration
– Group size
– Vehicle information
– Food & water consumption
– Body weight & weight gain
– Clinical observations and behavior (in-life observations)
– Clinical chemistry
– Hematology
– Coagulation
– Organ weights (absolute and relative)
– Gross pathology
– Histopathology

cronucleus assay (Igl et al., 2019). The purpose of these statistical considerations and the stratification of data is to gain insight into general features such as variability of parameters and how these change over time. This will be visualized with the help of control charts. Key factors of experimental conditions that influence the variability shall be identified and numerically described. Such factors could be, for example, changes in analytical procedures for measuring clinical chemistry parameters or modifications in the housing conditions. A list of the minimum parameters to be collected and analyzed is given in Table 1.

When these factors are sufficiently characterized, the proof of principle would be to use the collected control group data to construct VCGs for studies that have been performed in the past. Two approaches are possible: 1) In the "resampling approach", the VCG data is drawn in a randomized way out of a pool of animal data sets to match the actual control group of the study regarding the above-mentioned key factors. 2) In the "simulation approach", the gained information about the distribution of endpoints is used to simulate the necessary amount of virtual control data and thus provide virtual control data. In both cases, the studies are then re-analyzed using the VCG instead of the actually

¹ www.etransafe.eu

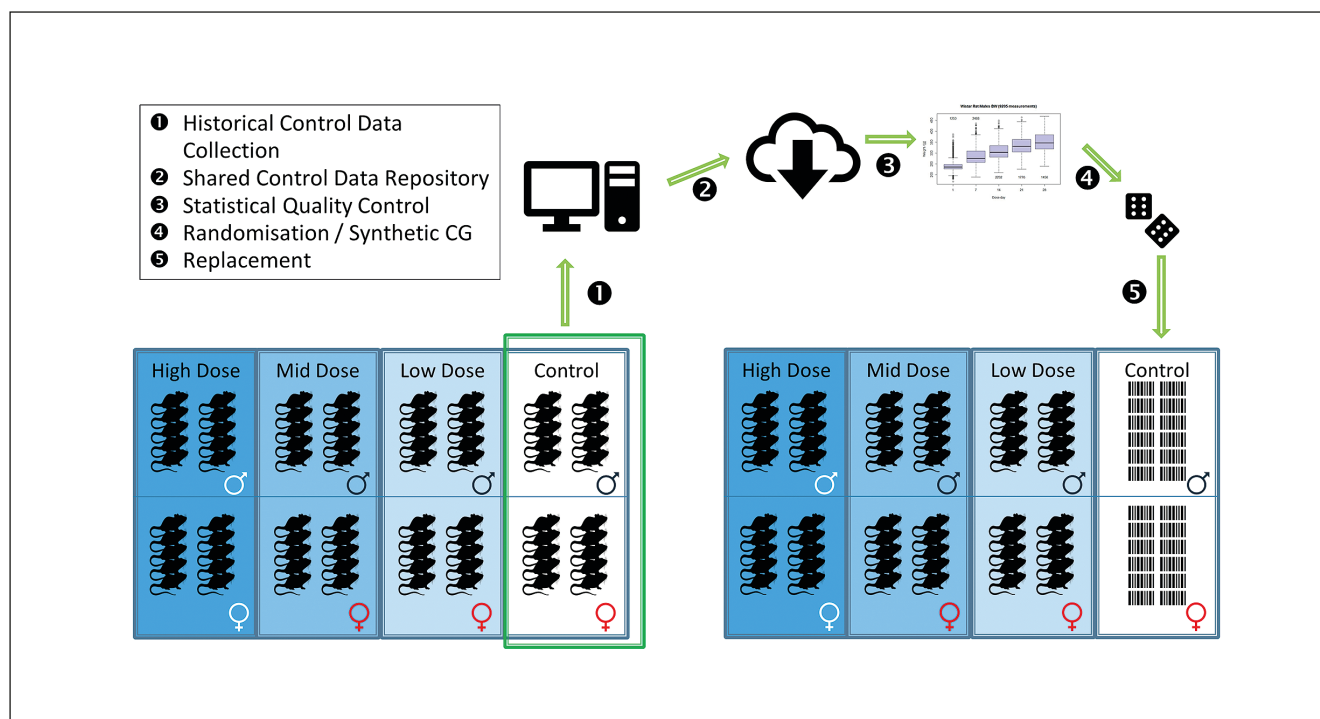


Fig. 1: Graphical illustration of the five steps in the replacement of animals in control groups by VCGs

used control animals, applying the accepted procedures of choice. Figure 1 illustrates the described steps for replacing control animals by VCGs.

In cases where differences in the assessment of compound-related toxicities are detected due to the use of the VCGs, a root cause analysis will be performed to identify which variation of parameters caused the deviating assessment. These analyses will be performed first within individual companies, but subsequently also across companies, in order to investigate the level of variability across companies or test facilities. Assessing the shared and pooled data is important since with larger data sets from different sources the distribution of normal ranges will naturally broaden. This needs to be strictly controlled in order to be able to delineate normal findings from treatment-related findings in these larger control animal data sets.

4.3 Steps towards scientific and regulatory acceptance

The eTRANSafe project has a scientific advisory board consisting of members with considerable regulatory background. In a first step, these members will assess the feasibility of the VCG concept. If the described analyses support a proof of concept, it is intended to approach the European Medicines Agency (EMA) within its Innovation Task Force (EMA, 2014) to present the approach and explore its regulatory acceptance.

We expect that the acceptance of the VCG concept will be achieved step-wise, where initially the size of control groups will be reduced and the omitted animals complemented with VCG data. In a subsequent step, after gaining sufficient experience and

confidence to allow a complete replacement of control animals, so-called sentinel animals can be housed within the animal facilities (Lipman and Homberger, 2003). These sentinel animals could be used for several studies and would only need to be investigated in case of suspected infections or other expected laboratory environment influences, thus overcoming the respective concerns.

5 Conclusion

Even though regulatory acceptance of VCGs across the different regions (EMA, FDA, PDMA (Japanese Pharmaceuticals and Medical Devices Agency)) will probably take years, if not decades, the data collection is an important asset and advancement for the 3Rs concept. The EMA guideline on repeated toxicity studies (EMA, 2008) states that the size of treatment groups is *inter alia* dependent on the “background knowledge concerning the ranges of variables to be studied in the species and strains”.

Assessing this variability of control group data through the availability of large data sets offers the chance to significantly reduce group sizes of control groups, hopefully even omitting them. Improving the background knowledge will also enhance the assessment of whether an observation is treatment- or compound-related or rather a spontaneous finding. An improved understanding of within- and between-animal variation will furthermore be helpful in reducing uncertainty in the determination of no (adverse) effect levels (NO(A)EL) (Paparella et al., 2013). It can also contribute to selecting the most appropriate strain



of species for specific studies, e.g., for carcinogenicity studies, where spontaneous tumor rates differ between strains (Greaves and Rabemampianina, 1982; Morawietz et al., 1992).

In conclusion, the success of the VCG concept will enable a significant reduction of animal use in repeated dose toxicity studies, thus substantially contributing to the 3R concept.

References

- Berry, D. A., Elashoff, M., Blotner, S. et al. (2017). Creating a synthetic control arm from previous clinical trials: Application to establishing early end points as indicators of overall survival in acute myeloid leukemia (AML). *J Clin Oncol* 35, Suppl, 7021. doi:10.1200/jco.2017.35.15_suppl.7021
- Brekke, T. D., Steele, K. A. and Mulley, J. F. (2018). Inbred or outbred? Genetic diversity in laboratory rodent colonies. *G3 (Bethesda)* 8, 679-686. doi:10.1534/g3.117.300495
- Briggs, K., Barber, C., Cases, M. et al. (2015). Value of shared preclinical safety studies – The eTOX database. *Toxicol Rep* 2, 210-221. doi:10.1016/j.toxrep.2014.12.004
- CDISC – Consortium of Standards in the Clinical Research Process (2016). Nonclinical Studies Version 3.1. Standard for Exchange of Nonclinical Data Implementation Guide. CDISC Standard for Exchange of Nonclinical Data Team. doi:10.32388/kix3u3
- Crabbe, J. C., Wahlsten, D. and Dudek, B. C. (1999). Genetics of mouse behavior: Interactions with laboratory environment. *Science* 284, 1670-1672. doi:10.1126/science.284.5420.1670
- Creasy, D., Bube, A., de Rijk, E. et al. (2012). Proliferative and nonproliferative lesions of the rat and mouse male reproductive system. *Toxicol Pathol* 40, Suppl, 40S-121S. doi:10.1177/0192623312454337
- Eichler, H.-G., Bloechl-Daum, B., Bauer, P. et al. (2016). “Threshold-crossing”: A useful way to establish the counterfactual in clinical trials? *Clin Pharmacol Ther* 100, 699-712. doi:10.1002/cpt.515
- EMA – European Medicines Agency (2008). Guideline on Repeated Dose Toxicity. Doc. Ref. EMEA/CHMP/SWP/488313/2007. https://www.ema.europa.eu/en/documents/scientific-guideline/draft-guideline-repeated-dose-toxicity_en.pdf
- EMA (2014). Mandate of the EMA Innovation Task Force (ITF). Human Medicines Research and Development Support Division. EMA/484400/2014. https://www.ema.europa.eu/en/documents/other/mandate-european-medicines-agency-innovation-task-force-itf_en.pdf
- FDA – US Food and Drug Administration (2003a). Toxicological Principles for the Safety Assessment of Food Ingredients. Redbook 2000. Chapter IV.C.3.a. Short-Term Toxicity Studies with Rodents. <https://www.fda.gov/files/food/published/Toxicological-Principles-for-the-Safety-Assessment-of-Food-Ingredients.pdf>
- FDA (2003b). Toxicological Principles for the Safety Assessment of Food Ingredients. Short-Term Toxicity Studies with Non-Rodents. Redbook 2000. Chapter IV.C.3.b. <https://www.fda.gov/files/food/published/Toxicological-Principles-for-the-Safety-Assessment-of-Food-Ingredients.pdf>
- FDA (2014). Guidance for Industry: Providing Regulatory Submissions in Electronic Format – Standardized Study Data. <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/providing-regulatory-submissions-electronic-format-standardized-study-data>
- Gad, S. C. (1995). Subchronic and chronic toxicity studies. In S. C. Gad (ed.), *Safety Assessment of Pharmaceuticals* (111-128). New York, USA: Van Nostrand Reinhold, Thompson Publishing.
- Greaves P. and Rabemampianina Y. (1982). Choice of rat strain: A comparison of the general pathology and the tumour incidence in 2-year old sprague-dawley and long-evans rats. In C. M. Chambers and P. L. Chambers (eds.), *New Toxicology for Old*. Archives of Toxicology, Supplement. Volume 5. Berlin, Heidelberg, Germany: Springer. doi:10.1007/978-3-642-68511-8_53
- Greim, H., Gelbke, H.-P., Reuter, H. et al. (2003). Evaluation of historical control data in carcinogenicity studies. *Hum Exp Toxicol* 22, 541-549. doi:10.1191/0960327103ht3940a
- GV-SOLAS (1999). Working Party Report: Implications of infectious agents on results of animal experiments. *Lab Anim* 33, Suppl. 1, 39-87. doi:10.1258/002367799780639987
- Hardy, B., Apic, G., Carthew, P. et al. (2012). Food for thought ... a toxicology ontology roadmap. *ALTEX* 29, 129-137. doi:10.14573/altex.2012.2.129
- Hartung, T. (2007). Food for thought ... on validation. *ALTEX* 24, 67-73. doi:10.14573/altex.2007.2.67
- Haseman, J. K. (1992). Value of historical controls in the interpretation of rodent tumor data. *Drug Inf J* 26, 191-200. doi:10.1177/009286159202600210
- ICH – International Council of Harmonisation (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). https://www.ema.europa.eu/en/documents/scientific-guideline/ich-e-10-choice-control-group-clinical-trials-step-5_en.pdf
- Igl, B.-W., Birsch, A., Bringezu, F. et al. (2019). The rat bone marrow micronucleus test: Statistical consideration on historical negative control data. *Regul Toxicol Pharmacol* 102, 13-22. doi:10.1016/j.yrtph.2018.12.009
- Jia, Z., Lilly, M. B., Koziol, J. A. et al. (2014). Generation of “virtual” control groups for single arm prostate cancer adjuvant trials. *PLoS One* 9, e85010. doi:10.1371/journal.pone.0085010
- Kramer, M. and Font, E. (2016). Strategies for reducing control group size in experiments using live animals. *Conference on Applied Statistics in Agriculture*, 100-120. doi:10.4148/2475-7772.1479
- Kramer, M. and Font, E. (2017). Reducing sample size in experiments with animals: Historical controls and related strategies. *Biol Rev Camb Philos Soc* 92, 431-445. doi:10.1111/brev.12237
- Kuwagata, M., Sakai, Y., Tanaka, S. et al. (2019). Historical control data on developmental toxicity studies in rats. *Congenit Anom* 59, 125-131. doi:10.1111/cga.12305
- Lipman, N. S. and Homberger, F. R. (2003). Rodent quality as-

- surance testing: Use of sentinel animal systems. *Lab Anim* 32, 36-43. doi:10.1038/labani0503-36
- Loeb, W. and Quimby, F. (1989). *The Clinical Chemistry of Laboratory Animals*. New York, USA: Pergamon Press.
- Low-Marchelli, J. (2017). Strategies to Minimize Genetic Drift and Maximize Experimental Reproducibility in Mouse Research. Charles River – The Jackson Laboratory. http://animalab.eu/sites/all/pliki/white_paper_eng.pdf
- Morawietz, G., Rittinghausen, S. and Mohr, U. (1992). RITA – Registry of industrial toxicology animal-data: Progress of the working group. *Exp Toxicol Pathol* 44, 301-309. doi:10.1016/s0940-2993(11)80216-2
- OECD – Organisation for Economic Co-operation and Development (2008a). Repeated Dose 28-Day Oral Toxicity Study in Rodents (OECD TG 407), in Revised Guidance Document 150 on Standardised Test Guidelines for Evaluating Chemicals for Endocrine Disruption. OECD Publishing, Paris. doi:10.1787/9789264304741-22-en
- OECD (2008b). Test No. 452: Chronic Toxicity Studies. *OECD Guidelines for the Testing of Chemicals, Section 4*. OECD Publishing, Paris. doi:10.1787/9789264071209-en
- OECD (2014). Guidance on the GLP Requirements for Peer Review of Histopathology. *OECD Series on Principles of Good Laboratory Practice and Compliance Monitoring, No. 16*. OECD Publishing, Paris. doi:10.1787/9789264228306-en
- OECD (2016). Application of GLP Principles to Computerised Systems. *Series on Principles of Good Laboratory Practice and Compliance Monitoring, No. 17*. <https://bit.ly/35r8Gqx>
- Paparella, M., Daneshian, M., Hornek-Gausterer, R. et al. (2013). Uncertainty of testing methods – What do we (want to) know? *ALTEX* 30, 131-144. doi:10.14573/altex.2013.2.131
- Petterino, C. and Argentino-Storino, A. (2006). Clinical chemistry and haematology historical data in control Sprague-Dawley rats from pre-clinical toxicity studies. *Exp Toxicol Pathol* 57, 213-219. doi:10.1016/j.etp.2005.10.002
- Pfuhler, S., Kirkland, D., Kasper, P. et al. (2009). Reduction of use of animals in regulatory genotoxicity testing: Identification and implementation opportunities – Report from an ECVAM workshop. *Mutat Res* 680, 31-42. doi:10.1016/j.mrgentox.2009.09.002
- Sanz, F., Pognan, F., Steger-Hartmann, T. et al. (2017). Legacy data sharing to improve drug safety assessment: The eTOX project. *Nat Rev Drug Discov* 16, 811-812. doi:10.1038/nrd.2017.177
- Spiegelhalter, D. J. (2004). Incorporating Bayesian ideas into health-care evaluation. *Statist Sci* 19, 156-174. doi:10.1214/088342304000000080
- Switchenko, J. M., Heeke, A. L., Pan, T. C. et al. (2019). The use of a predictive statistical model to make a virtual control arm for a clinical trial. *PLoS One* 14, e0221336. doi:10.1371/journal.pone.0221336
- Yanagawa, T. and Hoel, D. G. (1985). Use of historical controls for animal experiments. *Environ Health Perspect* 63, 217-224. doi:10.1289/ehp.8563217

Conflict of interest

The authors of this publication are employees of the participating companies or organizations. The contribution to this publication was part of their work and is paid by their salaries.

Acknowledgements

The described initiative is performed under the IMI eTRANSAFE project. eTRANSAFE has received support from IMI2 Joint Undertaking under Grant Agreement No. 777365. This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation program and the European Federation of Pharmaceutical Industries and Associations (EFPIA).