

<https://doi.org/10.47183/mes.2026-503>



METHODOLOGICAL ASPECTS OF STUDY DESIGN IN PRECLINICAL DRUG DEVELOPMENT: A REVIEW

Olga V. Shreder[✉], Dinara M. Ivashova

Scientific Centre for Expert Evaluation of Medicinal Products, Moscow, Russia

Introduction. Recent advances in molecular biology and computational science necessitate a revision of preclinical methodologies. Initiatives such as Tox21 program integrate traditional approaches with innovations, enabling new drug research strategies and transforming toxicology into a predictive science. However, the development of safe drugs through modeling requires the standardization of experimental and statistical approaches, while accounting for the characteristics of target populations.

Objective. To systematize methodological approaches toward unified statistical and experimental planning of preclinical drug studies.

Discussion. Analysis of the reviewed approaches shows that methodological rigor at the preclinical stage is a critical factor in the successful introduction of drugs into clinical practice. This paper proposes a tiered framework of study designs: ranging from exploratory and pilot studies that generate hypotheses to rigorous translational and pivotal trials intended for confirmation. Robust statistical planning is essential for ensuring both internal and external validity, thereby enhancing the reliability and translatability of research findings. Pilot and exploratory studies focus on effect size estimation and descriptive statistics, whereas confirmatory (pivotal) trials require rigorous a priori power analysis alongside strict randomization and blinding protocols. The presented classification of experimental units — whether single subjects, littermates, or repeated measures — is critical for ensuring statistical validity and preventing pseudoreplication. Unifying statistical designs for translational preclinical trials by analogy with clinical studies is the definitive strategy for ensuring reliable safety and efficacy predictions for pharmacologically active compounds prior to human trials.

Conclusions. The unification of statistical and experimental plans at the preclinical stage is necessary for obtaining reliable data on drug safety and efficacy. Adherence to unified methodological principles in exploratory, translational, and pivotal studies ensures data continuity and reduces the risk of erroneous predictions when extrapolating findings to human populations. The presented approaches can serve as a methodological basis for harmonizing preclinical and clinical strategies in medical and veterinary pharmacology.

Keywords: pivotal studies; translational studies; exploratory studies; preclinical studies; statistical designs; experimental design; experimental unit; validity; reproducibility

For citation: Shreder O.V., Ivashova D.M. Methodological aspects of study design in preclinical drug development: A review. *Extreme Medicine*. 2026. <https://doi.org/10.47183/mes.2026-503>

Funding. This work was carried out within the framework of the State Assignment of the Scientific Centre for Expert Evaluation of Medicinal Products (NCEMP), No. 056-00061-26-00, for applied scientific research (State Research Registration No. 124022300127-0).

Potential conflict of interest. The authors declare no conflict of interest.

AI tools use. The authors declare no AI (artificial intelligence) tools were used in the preparation of this article.

✉ Olga V. Shreder olga_shreder@list.ru

Received: 6 Apr 2026 **Revised:** 15 June 2026 **Accepted:** 2 July 2026 **Online first:** 5 Oct. 2026

УДК 615.038.8:615.03

МЕТОДОЛОГИЧЕСКИЕ АСПЕКТЫ В ДИЗАЙНАХ ДОКЛИНИЧЕСКИХ ИССЛЕДОВАНИЙ ЛЕКАРСТВЕННЫХ СРЕДСТВ: ОБЗОР

О.В. Шредер[✉], Д.М. Ивашова

Научный центр экспертизы средств медицинского применения, Москва, Россия

Введение. Современное развитие молекулярной биологии и компьютерных наук обуславливает необходимость пересмотра методологии доклинических исследований. Программа Tox21 обеспечивает интеграцию традиционных подходов с инновациями, что позволяет создавать новые стратегии исследования лекарственных средств и трансформировать токсикологию в прогностическую науку. Для разработки безопасных лекарственных препаратов с помощью моделирования требуется стандартизация экспериментальных и статистических подходов с учетом особенностей целевых популяций.

Цель. Систематизация методологических подходов к статистическому и экспериментальному унифицированному планированию доклинических исследований лекарственных средств.

Обсуждение. Анализ рассмотренных подходов показывает, что методологическая строгость на доклиническом этапе является критическим фактором успешного внедрения лекарственных средств в клинику. В работе обоснована четкая иерархия исследовательских дизайнов: от поисковых и пилотных эмпирических схем, предназначенных для генерации гипотез, до трансляционных и опорных исследований, направленных на их подтверждение. Установлено, что ключевым результатом корректного статистического планирования является обеспечение внутренней и внешней валидности получаемых данных. В частности, показано, что в пилотных исследованиях приоритет отдается точечным оценкам и описательным статистикам, тогда как в подтверждающих (опорных) дизайнах критически важны априорный расчет мощности, рандомизация и ослепление. Представленная в работе классификация экспериментальных единиц («отдельное животное», «гнездовая выборка», «повторные наблюдения») напрямую определяет достоверность статистических выводов и минимизацию риска псевдорепликации. Унификация статистических планов трансляционных доклинических испытаний по аналогии с клиническими исследованиями является результирующей стратегией,

© O.V. Shreder, D.M. Ivashova, 2026

обеспечивающей надежность прогнозов безопасности и эффективности фармакологически активных соединений перед началом их изучения на человеке.

Выводы. Унификация статистических и экспериментальных планов на доклиническом этапе — необходимое условие получения надежных данных о безопасности и эффективности лекарственных средств. Соблюдение единых методологических принципов в поисковых, трансляционных и опорных исследованиях обеспечивает преемственность результатов и снижает риск ошибочных прогнозов при экстраполяции на человека. Представленные подходы могут служить методической основой для гармонизации доклинических и клинических стратегий в медицинской и ветеринарной фармакологии.

Ключевые слова: опорные исследования; трансляционные исследования; поисковые исследования; доклинические исследования; статистические планы; дизайн эксперимента; экспериментальная единица; валидность; воспроизводимость

Для цитирования: Шредер О.В., Ивашова Д.М. Методологические аспекты в дизайнах доклинических исследований лекарственных средств: обзор. *Экстремальная биомедицина*. 2026. <https://doi.org/10.47183/mes.2026-503>

Финансирование. Работа выполнена в рамках государственного задания ФГБУ «НЦЭСМП» Минздрава России № 056-00061-26-00 на проведение прикладных научных исследований (номер государственного учета НИР 124022300127-0).

Потенциальный конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Использование инструментов ИИ. Авторы подтверждают, что при подготовке рукописи не использовались инструменты искусственного интеллекта.

✉ Шредер Ольга Васильевна olga_shreder@list.ru

Статья поступила: 06.04.2026 **После доработки:** 15.06.2026 **Принята к публикации:** 02.07.2026 **Online first:** 05.10.2026

INTRODUCTION

Breakthroughs in molecular biology and computational science are driving a paradigm shift in pre-clinical toxicology, demanding more sophisticated methodological frameworks¹. Evolving within the “21st Century Toxicology” concept, initiatives like the Tox21 program integrate conventional methodologies with technological innovations to design new drug research strategies [1]. A core component of this program is a publicly accessible library covering 10,000 structurally diverse compounds, enabling high-throughput toxicity profiling of new substances and post-market surveillance of approved drugs. This approach, incorporating computational technologies and alternative research methods, complements traditional *in vitro* and *in vivo* experimental models, expanding the predictive capabilities of toxicology as a science. Furthermore, this new concept relies on a comprehensive analysis of biological data to uncover the mechanisms of action on molecular targets through simulation methods and computational modeling². In the future, the transition to predictive toxicology based on mathematical modeling will allow the preclinical drug development

program to take a leading role in providing and interpreting scientific data essential for informed public health decision-making [1, 2].

Developing effective and safe drugs via mathematical modeling requires standardized research approaches tailored to specific designs, goals, and objectives, alongside consideration of the target patient populations for which the product is being developed. These principles apply not only to human medicinal products but are also relevant to the study of veterinary drugs³.

Ensuring the quality of such studies requires adherence to both formal regulatory guidelines⁴ — which outline fundamental principles — and study-specific protocols tailored to each project’s objectives for streamlined drug development. Such documents must comply with international standards, primarily regarding statistically robust experimental designs. This requirement is the foundation of the design not only of computational experiments (computer modeling) but also of classical experimental studies involving laboratory animals [1, 2].

Thus, in addition to statistical aspects, successful drug development needs to address the key elements

¹ Decision of the Board of the Eurasian Economic Commission No. 202 of Nov. 26, 2019 “On approval of the Guidelines for preclinical safety studies for the purpose of conducting clinical trials and registration of medicinal products”; GOST R 56701-2015. Medicinal products for medical use. Guidelines for the planning of preclinical safety studies with the aim of subsequent clinical trials and registration of medicinal products; GOST R 57689-2017. Medicinal products for medical use. Detection of toxic effects on reproductive function and male reproductive function; ICH M3(R2) Guidance on nonclinical safety studies for the conduct of human clinical trials and marketing authorization for pharmaceuticals, 2020 (access date Oct. 4, 2021).

² ICH M15 Guideline on general principles for model-informed drug development. Nov. 20, 2024, EMA/CHMP/ICH/496426/2024 Committee for Human Medicinal Products; Considerations with respect to future MIDD related guidelines: output from ICH model-informed drug development (MIDD) discussion group (DG) 2021. Members of ICH MIDD Discussion Group January 2021 to March 2022.

³ Order of the Ministry of Agriculture No. 101 of Mar. 6, 2018 “On approval of the rules for conducting preclinical studies of a medicinal product for veterinary use, clinical studies of a medicinal product for veterinary use, and bioequivalence studies of a medicinal product for veterinary use” (as amended on June 5, 2020, No. 309).

⁴ Decision of the Board of the Eurasian Economic Commission No. 202 of 26.11.2019 “On approval of the Guidelines for preclinical safety studies for the purpose of conducting clinical trials and registration of medicinal products”; GOST R 56701-2015. Medicinal products for medical use. Guidelines for the planning of preclinical safety studies with the aim of subsequent clinical trials and registration of medicinal products; GOST R 57689-2017. Medicinal products for medical use. Detection of toxic effects on reproductive function and male reproductive function; ICH M3(R2) Guidance on nonclinical safety studies for the conduct of human clinical trials and marketing authorization for pharmaceuticals, 2020 (access date Oct. 4, 2021); Order of the Ministry of Agriculture No. 101 of Mar. 6, 2018 “On approval of the rules for conducting preclinical studies of a medicinal product for veterinary use, clinical studies of a medicinal product for veterinary use, and bioequivalence studies of a medicinal product for veterinary use” (as amended on June 5, 2020, No. 309).

of organizing and planning animal experiments⁵. These factors encompass the inherent complexity of preclinical workflows, logistical constraints regarding model selection and execution, and the necessity for tailored statistical designs across complementary toxicological studies [2–6].

This study aims to systematize methodological approaches to unified statistical and experimental planning of preclinical drug development.

MAIN BODY

Preclinical drug development comprises two main phases: the exploratory, aimed at identifying, developing, and evaluating new drugs, and the pivotal, conducted *in vivo* and *in vitro* to assess the safety and efficacy of a drug or other medical product before initiating human clinical trials [7]. This classification is defined by distinct objectives, as well as the role of each phase in the overall drug development strategy.

Exploratory preclinical investigations serve to identify and study the primary pharmacological effects of original molecules, as well as to establish the optimal dose levels and routes of administration. In contrast, pivotal preclinical testing validates acceptable toxicological profiles, potential therapeutic effects, and adverse reactions that determine the risk for medical use. By investigating and confirming the effects of active pharmaceutical substances, preclinical studies elucidate drug action mechanisms, advancing our understanding of their interaction with biological systems and expanding their potential applications in human and veterinary medicine [6, 7]. Standardized statistical methods are employed to ensure the correct interpretation, comparability, and extrapolation of research results⁶.

The modern approach to data quality is defined by the principles of evidence-based medicine, which relies on statistical methods of assessment and prediction [8–11]. Statistical aspects of experimental research encompass the processing and interpretation of obtained information to identify patterns and relationships between the intervention (therapy, experimental conditions) and the outcome. These considerations are important at all stages of drug development, during which a candidate drug undergoes evaluation across various experimental platforms: computer modeling (*in silico*), cell culture experiments (*in vitro*), and animal testing (*in vivo*) [12–15].

Notably, decisions regarding the feasibility of drug development are based on primary assessments of the toxicity, efficacy, and safety of newly synthesized

chemical compounds. This process involves the use of a statistical research framework selected according to the specific experimental design.

This review outlines the main statistical principles and frameworks used in planning experiments throughout preclinical drug development, encompassing exploratory, pivotal, and translational study designs.

Exploratory studies in the preclinical development of novel medicinal products

Exploratory studies represent the initial stage of development, where researchers formulate hypotheses and objectives to evaluate novel medicinal products [16]; the main task here is to identify promising candidate compounds for further analysis. These investigations rely on three primary approaches to evaluate the activity, selectivity, and potential toxicity of compounds: high-throughput screening (computerized robotic screening for the analysis of compound libraries, genomes, proteins, and peptides), molecular modeling (an approach that uses computational methods to create, analyze, and visualize three-dimensional molecular structures, enabling the study of molecular properties and prediction of their behavior under various conditions), and biological target analysis (monitoring of molecules that interact with drugs or other biologically active substances and cause certain biological effects). These approaches employ *in vitro*, *in vivo*, and *in silico* methods, alongside the evaluation of the pharmacokinetics and pharmacodynamics of drug candidates [2, 12].

After identifying the best candidates, their physicochemical properties (such as solubility, stability, and bioavailability) are optimized to increase efficacy and reduce side effects; this is followed by the assessment of safety and efficacy within the framework of pivotal preclinical studies and clinical trials to confirm the drug's potential for medical or veterinary use [2–4, 17].

Overall, exploratory preclinical investigations aim to identify promising candidates for subsequent clinical development. The selection process is often driven by scientific data obtained, among other sources, from experimental studies, their summaries, and general theories, which may not always be included in the registration dossier. Exploratory preclinical studies are critically important for selecting the most promising candidate compounds and their further development within the framework of pilot empirical, translational, and pivotal studies [7, 18, 19].

⁵ Principles of Good Laboratory Practice (GLP) regarding medicinal products. PharmAdvisor, translation version of Oct. 1, 2021. URL: <https://pharmadvisor.ru/document/tr3899/> (access date Jan. 26, 2017)

⁶ ICH Harmonised tripartite guideline. Statistical principles for clinical trials E9, 1998; Federal Law of the Russian Federation No. 61-FZ of Apr. 12, 2010 “On the Circulation of Medicinal Products”; Recommendations of the Board of the Eurasian Economic Commission No. 19 of Nov. 3, 2020 “On the Guidelines for the application of biostatistics principles in clinical trials of medicinal products”.

Pilot studies in the preclinical development of novel medicinal products

Pilot investigations are typically stratified into three methodological categories⁷: operational (procedural), empirical (exploratory), and translational (focused on data translation) [16, 20].

The choice of the pilot study type depends on:

- feasibility of implementation: assessed during the planning stage of the operational pilot study (logistics, process optimization); can be conducted without animal involvement;
- effectiveness of the approach: determined during the empirical exploratory experiment (correspondence of results to the tasks);
- prospects for implementation: assessed in the translational study (generalizability, reliability, reproducibility, replication in different models/conditions); confirms the concept (hypothesis) based on the results of empirical pilot studies.

Important planning principles for all types of pilot studies are:

- the 3R principles (replacement, reduction, refinement), which require justification of the use of animals in experiments and consideration of alternatives such as *in vitro* studies and systematic reviews;
- defining the goal and benefit ensures obtaining meaningful information for planning large-scale studies; it is important to clearly define what information is planned to be obtained and how it will be used;
- modeling: the most relevant animal and disease experimental models should be used to achieve high informativeness of the results;
- investment in pilot projects should not exceed 20% of the total resources planned for the main study;
- justification of pilot trials. Pilot trials play a key role in biomedical research: they help identify problems and reduce risks, saving time and resources [7, 21–23].

Research objectives dictate the execution strategy, requiring either an iterative approach to problem-solving with incremental improvements at each stage or separate pilot studies for each objective [20]. Consequently, a comprehensive understanding of the drug development strategy requires clear differentiation between the designs of pilot exploratory and translational preclinical investigations [24, 25].

The statistical design of a study is determined by its primary objective, distinguishing between hypothesis-generating (empirical or exploratory) and hypothesis-testing studies. The latter include translational studies and pivotal (confirmatory) trials. Hypothesis-generating studies provide the theoretical foundation from which testable hypotheses can be

derived regarding disease pathophysiology and potential drug targets [25].

To assess the validity of the proposed hypotheses and justify further clinical study of therapeutic targets, research designs must incorporate standardized requirements (governing the number, conditions, and sequence of experiments that are necessary and sufficient to solve the problem with the required precision) and statistical methods. These designs must include a protocol corresponding to the study type, with a detailed description of the technique and rules for conducting experiments; statistical analysis methods that ensure hypothesis testing; and a procedure for presenting results to validate the proposed hypotheses [26–28].

In order to obtain reliable and valid results in studies on animal models, it is necessary to ensure the reproducibility of intervention effects. This requires conducting studies with maximum rigor and transparency, adhering to methodological standards and procedural sequences at each stage [16, 20]:

1. Hypothesis formulation: specification of a testable assumption based on scientific evidence.
2. Selection of methodology and comparison group: definition of the experimental protocol, analytical approach, and data comparison.
3. Experiment: assurance of identical housing conditions and manipulations for all groups except for the studied intervention (single difference principle).
4. Data collection and analysis: recording and statistical processing of the obtained data.
5. Result interpretation: conclusions based on data analysis and hypothesis testing.
6. Documentation of results: a clear and detailed description of outcomes for further use.

The quality of pilot data, which determines the reliability of the study, is directly related to its internal validity (confirmation that the evaluated factor causes the observed effect; it ensures accurate effect measurement and minimal error) and external validity (confirmation that results can be generalized). Internal validity assesses the risk of random or systematic errors, while external validity assesses the representativeness (reflecting the key characteristics of the modeled condition, i.e., the validity of the model) and reliability of the results (consistency of results in different models and conditions) for a broader population [20, 24]. Although bias compromises both types of validity, the absence of bias does not guarantee external validity if the sample is not representative (Table 1).

Even with high internal validity, conclusions may be unreliable for a broad population due to a non-representative model or clinical insignificance. Careful use of *p*-values is permissible only in large, repeated, clinically

⁷ Decision of the Board of the Eurasian Economic Commission No. 202 of Nov. 26, 2019 "On approval of the Guidelines for preclinical safety studies for the purpose of conducting clinical trials and registration of medicinal products"; Order of the Ministry of Agriculture No. 101 of Mar. 6, 2018 "On approval of the rules for conducting preclinical studies of a medicinal product for veterinary use, clinical studies of a medicinal product for veterinary use, and bioequivalence studies of a medicinal product for veterinary use" (as amended on June 5, 2020, No. 309).

oriented pilot studies. For reliable conclusions, internal and external validity are important, as is the minimization of bias.

Exploratory pilot studies

Exploratory pilot studies aim to assess feasibility and provide descriptive information rather than test hypotheses. While animal sample sizes are not regulated at this stage, they should be minimized without compromising the required informative value. For a series of exploratory experiments within pilot investigations, it is recommended to use a sample of three (*in vitro*) and six (*in vivo*) animals per group [25, 29–31]. Statistical analysis should focus on describing the results using point estimates, frequencies, and ranges (minimum and maximum values) to reflect the spectrum of observed effects. Consequently, calculating confidence intervals and testing statistical hypotheses are omitted at this stage due to limited sample sizes and the absence of efficacy-confirmation objectives. Instead, modeling and sensitivity analysis play an important role in assessing the uncertainties associated with the predicted sample size [11, 32]. Such exploratory studies are essential to analyze best- and worst-case scenarios for statistical models, various covariates (numerical characteristics associated with the studied process), and analytical methods planned for the main study. A key element of exploratory experiment design consists in clearly defining criteria by which its success and potential for further application are assessed [11, 20–26].

Pilot empirical and translational studies

Empirical and translational studies serve to provide a preliminary assessment of efficacy and the translational potential of results. They help determine the feasibility

of further experiments, allowing researchers to optimize sample sizes based on the information power principle, defined as maximizing useful data yield while minimizing animal use [7, 11, 16].

Empirical studies demonstrate potential efficacy under optimal conditions, which entail conducting an experiment where all subjects are in identical environments, while validating the problem and hypothesis formulation, the adequacy of research methods, subject selection, and mathematical analysis⁸ [32]. The evaluation of efficacy is based on the quality of the outcome variables (responses), which for a correct assessment must meet the following criteria:

- effect size: the magnitude of the observed effect must be sufficiently pronounced for reliable measurement;
- accuracy and precision: assurance of stability and reproducibility across repeated measurements;
- relevance: correspondence to the predictions formulated in the study hypothesis [10, 14].

The main objective of an empirical study is to derive comprehensive information from the data [7, 16]. At this stage, investigations should be conducted under controlled conditions, without attempting to create a clinically relevant model. To achieve reliability and validity, researchers need to ensure the homogeneity of groups to exclude the impact of other factors on the course and outcomes of the experiment, control all factors except the independent variable to precisely determine its influence on the study logic and outcomes, and repeat the procedures several times to check the reliability and reproducibility of the results. This approach provides additional verification of the obtained data and validates the findings.

Notably, animal studies conducted at the initial stage of a project are exploratory in nature and are performed using a small number of animals. Rather than aiming to achieve statistical significance, the sample size at

Table 1. Causes of bias (systematic error) and methods for its minimization

Cause of bias	Bias minimization method	Notes on application in the study
Incorrect selection of animals for experimental studies	Adequate control groups or comparators	Ensures comparability of groups by baseline characteristics
Incorrect allocation of animals to intervention groups	Randomization procedure	Eliminates systematic bias when allocating animals to intervention groups
Breach of the order of performing procedures with animals	Blinding (allocation concealment)	Prevents the influence of the investigator on the outcome of procedures
Measurement and result evaluation errors	Clear inclusion and exclusion criteria for experimental animals	Standardization of assessment, reduction of variability
Exclusion of animals during the experiment and data loss	Clearly defined outcome metrics (primary/secondary)	Allows objective analysis of excluded data without introducing bias

Table prepared by the authors based on data from [20, 24]

⁸ Titova AA, Bilyalov AI, Kiyasov AP, Titova MA. Educational and methodological guide "Laboratory animals for scientific research." Kazan: Kazan University; 2021.

this point is determined by the need to obtain adequate information for subsequent analysis.

Translational studies represent a modern scientific framework designed to translate fundamental research findings into practical medical technologies and treatment methods [7, 16, 20]. This paradigm encompasses *in silico* (computer modeling), *in vitro* (in a test tube), *ex vivo* (on cell cultures), and *in vivo* (animal models) experiments. Its ultimate goal is to transform scientific achievements into clinical solutions for medical practice [1, 31]. The process involves a sequential transition from fundamental and mechanistic studies to clinically relevant animal models based on the Predicted Probability of Prediction (pPoP; an assessment generated by statistical models indicating the probability of a specific outcome based on predictor variables ranging from 0 to 1, where 1 indicates complete confidence and 0.5 represents random guessing). Investigations validate the preclinical proof of concept (pPoC; demonstrating that the studied substance or method possesses the claimed biological action) by assessing the intervention effects on clinically relevant animal models via parameters associated with a specific disease. This is followed by the transition to early phases of clinical trials to establish the clinical proof of concept (cPoC; data on the efficacy and safety of a new drug or therapy in a small patient cohort) and, ultimately, to randomized trials (Table 2) [16].

In translational studies, predictive accuracy is important for assessing clinical significance and the potential of extrapolating results from animal models to humans. Therefore, the representativeness of

any animal model is assessed based on three main criteria [20, 24, 25, 31]:

- phenotype or phylogenetic similarity;
- construct homology: confirmation of construct validity (experimental design, physiological and pathophysiological characteristics, administration routes, compound doses, treatment conditions, biomarkers of the modeled pathological process, and the methodological alignment and interpretation of preclinical and clinical endpoints);
- predictive accuracy: the ability of the model to determine the probability of specific clinical outcomes.

The criteria for external validity are often subjective, requiring a balance between different validation approaches. The predictive validity of an animal model is determined by its similarity to the target population and the specificity of the therapeutic response. This metric demonstrates how well the results observed in animals translate to humans, thereby establishing the clinical significance of the disease model [20, 24, 25]. Maximizing external validity requires compromises, especially for diseases that are difficult to replicate due to complex and variable clinical manifestations. To ensure a successful translational application, the model must be informative and predictable, even if its complete correspondence is impossible [11, 24].

Pivotal studies in the preclinical development of new medicinal products

Pivotal studies focus on confirming drug safety to obtain permission for conducting clinical trials. The results of these investigations form the basis of the registration

Table 2. Definition of the principles of proof of concept in preclinical and clinical studies

Criterion	Predicted Probability of Prediction (pPoP)	Preclinical proof of concept (pPoC)	Clinical proof of concept (cPoC)
Description	Statistical model-generated probability of a specific outcome based on predictor variables; confirmation of the effect on the target pathophysiological mechanism or disease process	Confirmation that the investigated substance or method possesses the claimed biological properties within relevant, disease-specific preclinical animal models	Preliminary efficacy and safety evaluation of a new medicinal product or therapy in a small patient cohort
Role (functions)	Quantitative assessment of predictive confidence; differentiation between positive and negative classes in binary classification; results are presented as surrogate parameters reflecting the disease process	Confirmation of the feasibility of achieving the desired effect prior to large-scale and costly studies; generation of results directly correlated to clinical disease manifestations	Testing of the clinical applicability hypothesis; demonstration of the medical product's prospects; decision-making on the feasibility of conducting large-scale trials (Phase III) and defining their design
Example	Efficacy evaluation of a new antitumor drug via surrogate biomarkers of cell proliferation and apoptosis	Efficacy evaluation of a new antitumor drug via tumor regression in patient-derived xenograft mouse models	Efficacy and safety validation of the new drug within a randomized study using a small patient cohort

Table prepared by the authors based on data from [20, 24]

dossier, which is necessary for initiating randomized controlled trials and subsequently registering the new pharmaceutical.

To assess drug toxicity within the framework of pivotal studies, various designs and methods are used (Fig. 1–5). They differ in structure, specificity of data unit definition, and approaches to their statistical analysis⁹ [7, 16–19, 32]. These designs include the following study types:

- acute toxicity¹⁰: assessment of toxic reactions after a single drug administration, where the observational unit is the individual animal;
- repeated-dose toxicity (subchronic and chronic): assessment of the body's reactions to multiple drug administrations and identification of the dose-response relationship; the observational unit is the individual animal under repeated observations;
- genotoxicity: assessment of DNA damage (*in vitro*, *in vivo*) to predict carcinogenicity or hereditary disorders during embryonic cell analysis, where the observational unit is the individual culture (*in vitro*) or tissue samples of an individual animal (*in vivo*);
- ontogenetic and reproductive toxicity: evaluation of drug effects on the reproductive system and offspring development; the observational unit is a group of animals for analyzing effects on the reproductive function of males and females with repeated drug administration or animal clusters (litter sample) for assessing embryotoxicity;
- carcinogenicity: assessment of the drug's ability to induce tumors in animals to predict cancer risk in humans, where the observational unit is the individual animal under repeated observations.

Figure 1 illustrates the qualitative and quantitative assessment of toxic reactions after a single drug administration. The design includes a stepwise dose escalation to identify threshold and lethal exposure levels. The statistical sample size depends on the species of laboratory animals: for rodents, it is recommended to use at least 5–6 animals of each sex per group; for non-rodents (e.g., rabbits, minipigs, dogs) — 3–5 animals of each sex per group. The study should include an acclimatization phase: a mandatory period of housing animals in the vivarium prior to the experiment to adapt them to the conditions, personnel, and schedules, which is necessary to account for group and individual response variability.

When assessing general toxicity under repeated exposure conditions (Fig. 2), the design involves administering the drug to two animal species (rodents and non-rodents) with statistically justified group sizes (rodents: at least 10 animals of each sex per group; non-rodents: at least 6 animals of each sex per group). The study duration is determined by the experimental goals and conditions, including the dosing regimen (daily or at intervals), administration route (oral, intravenous, intraperitoneal, etc.), course duration, and the need to evaluate the recovery period. Depending on these variables, observation periods can range from 2 weeks to 1 month (intravenous; intraperitoneal) or extend to 3 and 6 months (subchronic and chronic studies) with an acceptable route of administration (oral). This design enables the identification of dose-dependent toxic effects¹¹, including cumulative and delayed reactions¹².

Drug genotoxicity assessment¹³ involves the use of two animal species with sample sizes of five (rodents) and three animals (non-rodents) for each assessment period (day 2 and day 14). This approach guarantees sufficient statistical power to detect expected genotoxic effects (Fig. 3). This design is also applicable for pilot (exploratory) assessment of genotoxicity in somatic and germ cells of males and females, as well as in embryos. Moreover, the observation duration (at least 14 days) allows the identification of risks for impaired offspring development.

Ontogenetic and reproductive toxicity upon repeated administration represent toxic effects of substances that may manifest upon repeated exposure over a relatively short period of time (Fig. 4). They are associated with disorders of reproductive function and offspring development. The experiment involves repeated administration of the test substance, taking into account the periods of spermatogenesis (in males) and oogenesis (in females) in rats. To ensure sufficient statistical power, the number of animals for assessing fertility should be at least 10 subjects of each sex per group. The experiment involves obtaining F1 offspring (12 pups from 10 litters) to assess the impacts on reproductive function and embryotoxicity. This design enables the detection of any (including delayed) effects of the drug on the reproductive system and offspring development.

The carcinogenicity study design included a comparative analysis of two dose levels — the maximum

⁹ Decision of the Board of the Eurasian Economic Commission No. 202 of Nov. 26, 2019 “On approval of the Guidelines for preclinical safety studies for the purpose of conducting clinical trials and registration of medicinal products”; Order of the Ministry of Agriculture No. 101 of Mar. 6, 2018 “On approval of the rules for conducting preclinical studies of a medicinal product for veterinary use, clinical studies of a medicinal product for veterinary use, and bioequivalence studies of a medicinal product for veterinary use” (as amended on June 5, 2020, No. 309); Recommendations of the Board of the Eurasian Economic Commission No. 19 of Nov. 3, 2020 “On the Guidelines for the application of biostatistics principles in clinical trials of medicinal products”.

¹⁰ OECD (2001) OECD Guideline for testing of chemicals. Acute Oral Toxicity — Acute Toxic Class Method No. 423; OECD (2001) OECD Guideline for testing of chemicals. Acute Oral Toxicity — Fixed Dose Procedure No. 420.

¹¹ Guidance Document No. 116 on the Conduct and Design of Chronic Toxicity and Carcinogenicity Studies, Supporting Test Guidelines 451, 452 and 453. OECD Environment, Health and Safety Publications. Series on Testing and Assessment. No. 116. Paris: OECD; 2011. ENV/JM/MONO(2011)47.

¹² Decision of the Board of the Eurasian Economic Commission No. 202 of Nov. 26, 2019 “On approval of the Guidelines for preclinical safety studies for the purpose of conducting clinical trials and registration of medicinal products”.

¹³ OECD (2016) OECD Guideline for testing of chemicals. *In vivo* mammalian alkaline comet Assay No. 489.

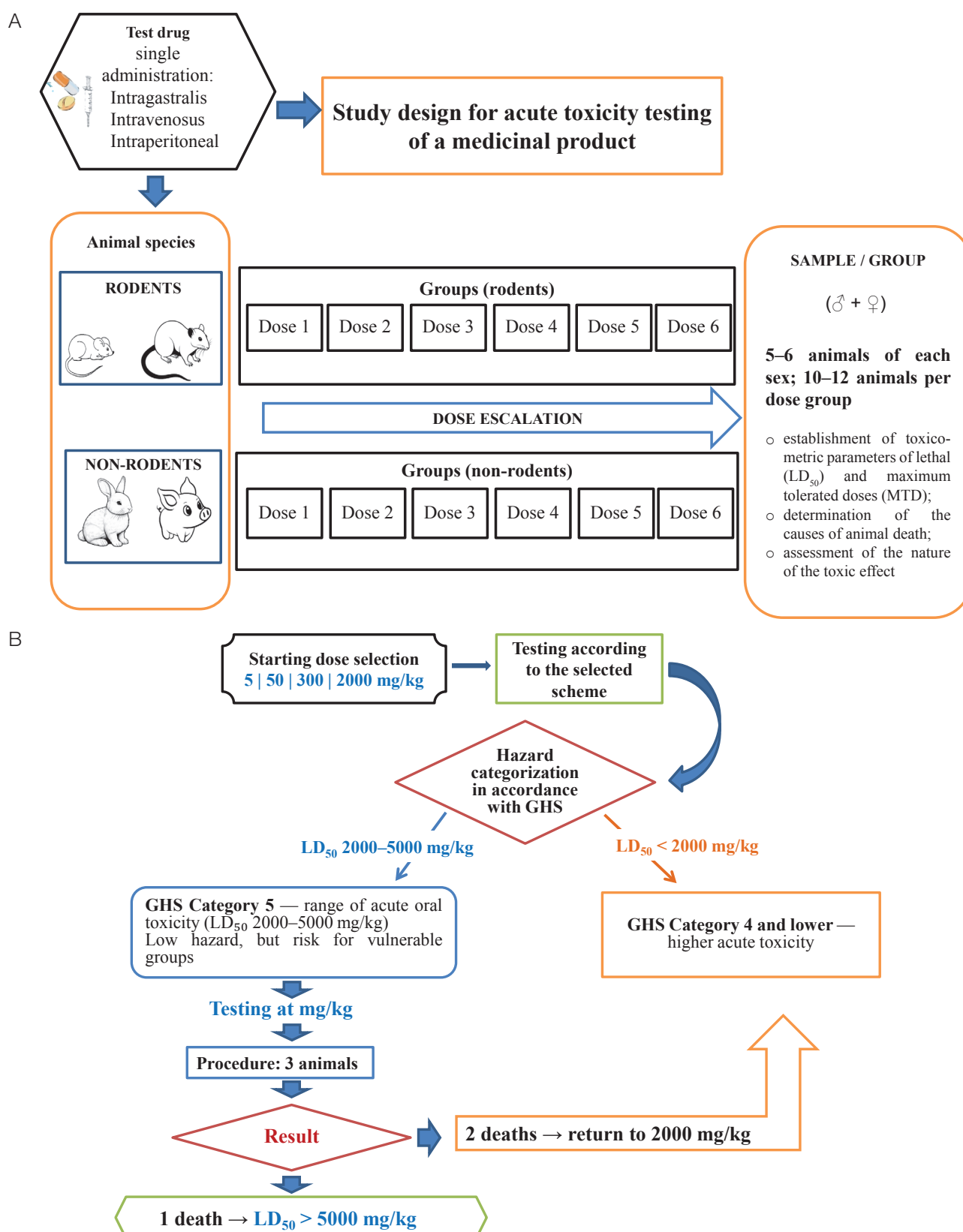


Figure prepared by the authors based on data from [16]

Fig. 1. Acute toxicity assessment. A — experimental designs of a pivotal preclinical study; B — general principles of dose level selection; *Intragastralis* — intragastric administration (via gavage); *Intravenous* — intravenous administration; *Intraperitonealis* — intraperitoneal administration; GHS — Globally Harmonized System of classification and labeling of chemicals based on LD_{50} threshold values: Category 1 (> 0–5 mg/kg), Category 2 (> 5–50 mg/kg), Category 3 (> 50–300 mg/kg), Category 4 (> 300–2000 mg/kg), Category 5 (> 2000–5000 mg/kg), or not classified

tolerated dose (MTD) and the therapeutic dose (ED) — across control and treated cohorts using relatively short and long observation periods (Fig. 5). Assessments were conducted at multiple time points (t_1, t_2, t_3, t_4, t_n) corresponding to key experimental phases: initiation of dosing, intermediate intervals, cessation of treatment, and the post-treatment observation period. The duration of the study varies: standard terms for carcinogenicity analysis in small animals (mice, rats, hamsters, rabbits) are 2 years for chronic administration, and may also reach 7–10 years for large animals (dogs, primates) depending on the objectives and regulatory requirements. Dose selection relies on preliminary studies: the maximum tolerated dose is defined as the highest dose that does not cause lethality and significant toxicity, while the therapeutic dose reflects the level required to produce the desired pharmacological effect. This design enables investigators to detect tumor formation, assess the dose-dependence of the carcinogenic effect, compare MTD and ED to determine the therapeutic index, and extrapolate risks to humans.

The experimental unit is a crucial element of any research design. Its selection determines the level at which observations are made and results are analyzed [16–19]. The validity, interpretability, and applicability of the resulting data directly depend on the correctness of this choice.

Table 3 outlines the primary experimental units used in toxicological studies and their specific applicability in various exposure assessment scenarios [7, 16]. These design elements form the basis for data collection

principles and the selection of statistical analysis methods [32–35].

An individual animal as the experimental unit

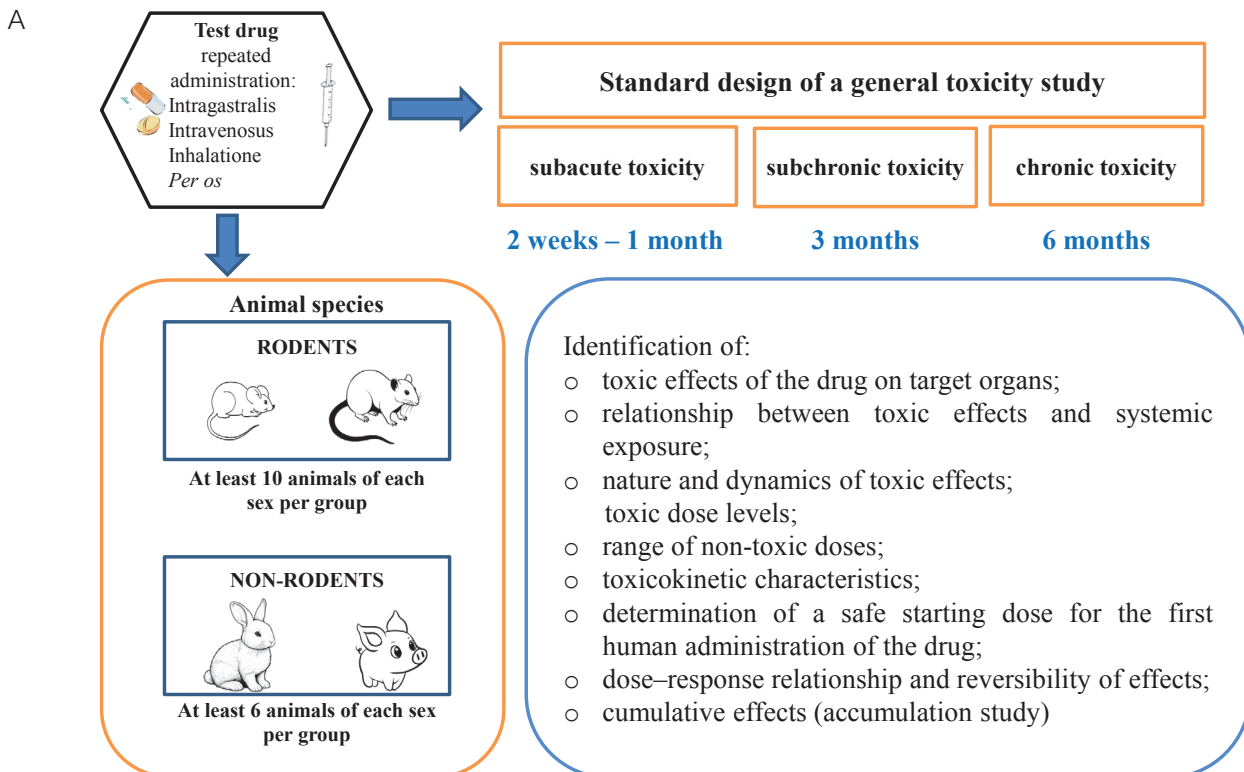
This design defines the experimental unit as the individual animal, which simultaneously serves as the statistical unit (Table 3).

Different species of laboratory animals (mice, rats, rabbits, minipigs) may be used in the study. For each species selected for the experiment, groups of several individuals are formed (conventionally numbered from 1 to 6). Following random group assignment, every animal is exposed to the evaluated factor (drug dose vs. control) according to a specified administration regimen, yielding individual results that constitute the primary data of the study.

This approach ensures independence of observations and statistical power, i.e., the probability of detecting a true effect if it exists. The size of the experimental groups (number of animals per group) is determined based on an a priori power analysis, taking into account: the expected effect size, the variability of the parameter, and the accepted error levels ($\alpha = 0.05$, $\beta = 0.20$).

A cluster of animals (cluster/nest sampling) as the experimental unit

In this design, the experimental (and statistical) unit is not the individual animal, but a cluster of animals, i.e., a litter or brood sample (Table 3).



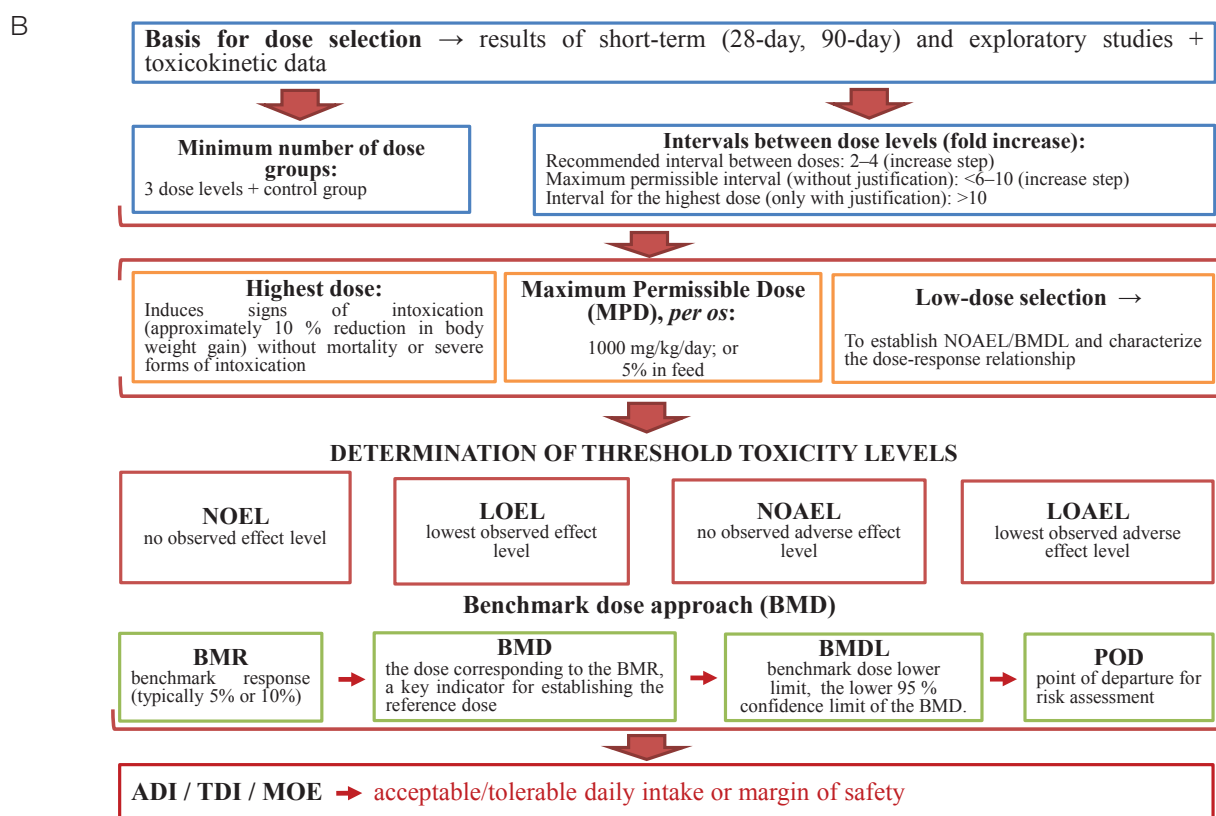


Figure prepared by the authors based on data from [16, 35]

Fig. 2. Repeated-dose general toxicity assessment. A — experimental designs of pivotal preclinical studies; B — general principles of dose level selection; MTD — maximum tolerated dose; *Intragastralis* — intragastric administration (via gavage); *Intravenosus* — intravenous administration; NOEL (no observed effect level) — the highest dose without any statistically significant effects; LOEL (lowest observed effect level) — the lowest dose that causes an observed effect; NOAEL (no observed adverse effect level) — the dose that does not cause visible adverse effects; LOAEL (lowest observed adverse effect level) — the lowest dose that causes an observed adverse effect; benchmark dose (BMD) approach — a statistical method for determining a toxicological point of departure by modeling the entire dose–response curve and calculating the dose corresponding to a specified response level (baseline mortality response, BMR), followed by an estimate of its lower confidence limit (benchmark dose lower confidence limit, BMDL); BMR — a pre-specified change in response (e.g., 5% or 10%) used to calculate the BMD; BMD — the dose corresponding to the specified change in response (BMR) compared to the control; BMDL — the lower 95% confidence bound for the BMD (used as the point of departure, POD); POD — the dose from which extrapolation to acceptable exposure levels for humans begins (usually NOAEL or BMDL); ADI (acceptable daily intake) — the amount of a substance that can be consumed daily over a lifetime without significant health risks; TDI (tolerable daily intake) — the acceptable daily intake of a contaminant that can be consumed daily without harm; MOE — the ratio of the POD to the estimated human exposure level (margin of exposure)

This approach is used in studies of reproductive and ontogenetic toxicity, where the male and/or female (parental pair or one of the parents) receives treatment, while subsequent analyses focus on the offspring (F1 generation).

Each experimental group comprises several litters, conventionally designated as Litters 1 through 6 (by dam number). Since each litter serves as a single experimental unit, individual littermates are not considered independent statistical units: their data are averaged or analyzed with clustering accounted for. This allows for an accurate assessment of effects associated with within-litter correlation (genetic similarity,

common gestation and nursing conditions). The analysis also takes into account the sex of the offspring. The size of the experimental group is determined by the number of litters (clusters) rather than by the total number of animals, which is critical for calculating statistical power and avoiding artificial sample size inflation (pseudoreplication).

Repeated observations on the same animal

In this design, the experimental (and statistical) unit is the individual animal; however, repeated measurements are performed for each unit at successive time

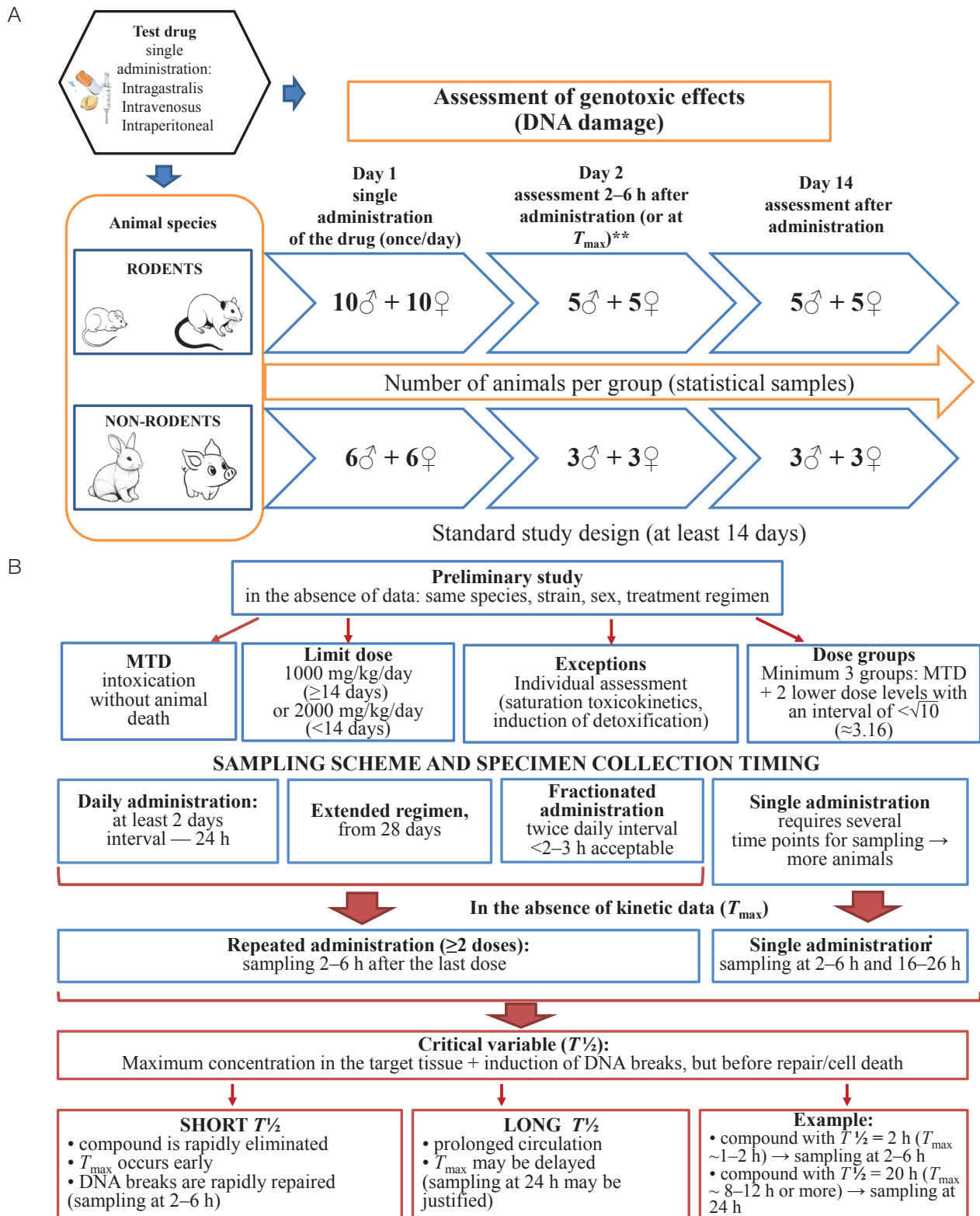


Figure prepared by the authors based on data from [16]

Fig. 3. Single-dose genotoxicity (comet assay) and developmental outcomes. A — experimental designs of pivotal preclinical studies; B — general principles of dose level selection; *Intragastralis* — intragastric administration (via gavage); *Intravenosus* — intravenous administration; *Intraperitonealis* — intraperitoneal administration; $T_{1/2}$ — half-life of the test substance; T_{max} — time to reach the maximum concentration of the test substance and/or its active metabolites in target tissue or plasma. Sampling is recommended at T_{max} , or 2–6 h post-administration if T_{max} is unknown, to capture the maximum number of induced DNA breaks prior to repair

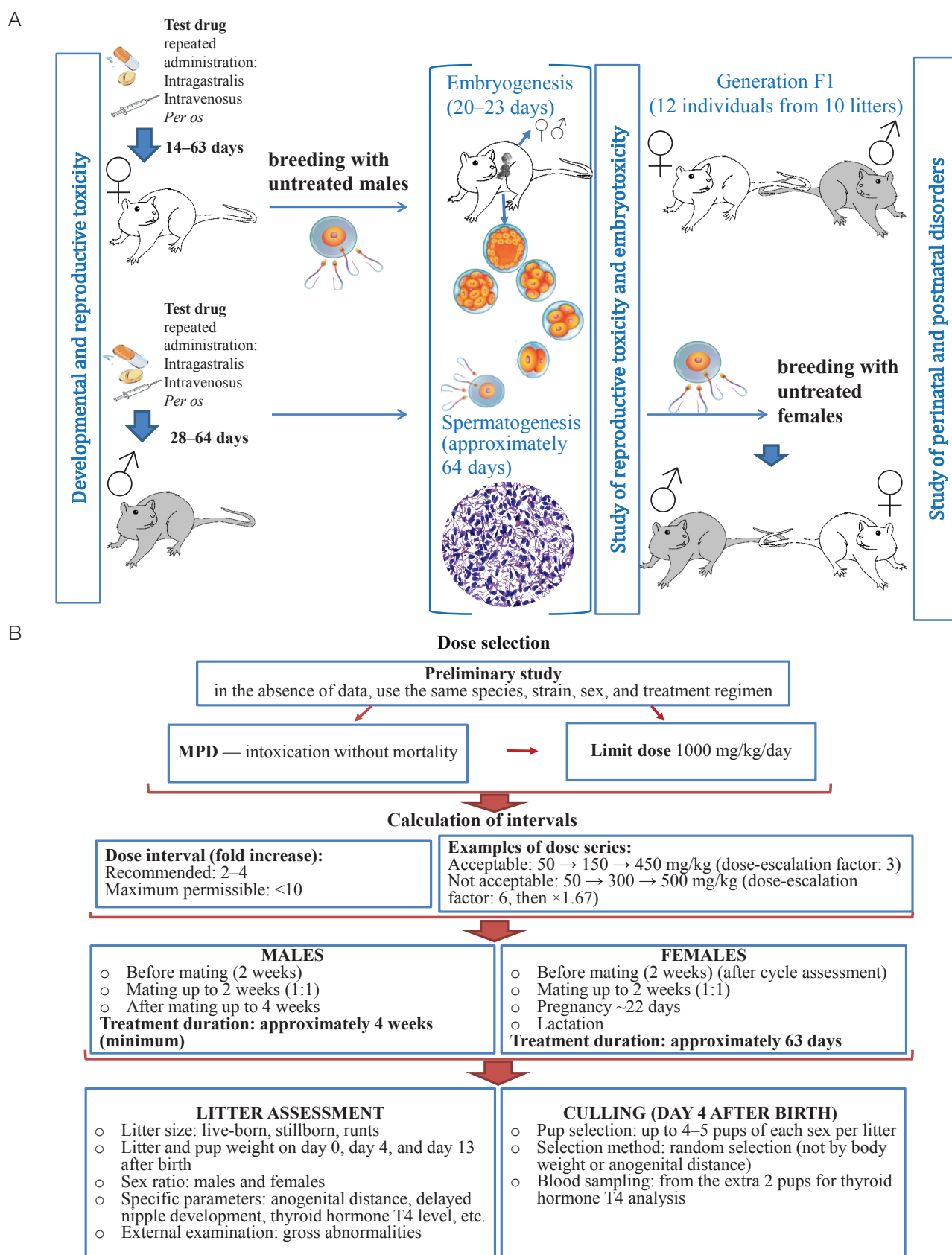


Figure prepared by the authors based on data from [16]

Fig. 4. Repeated-dose assessment of ontogenetic and reproductive toxicity. A — experimental designs of pivotal preclinical studies; B — general principles of dose level selection; MTD — maximum tolerated dose; *Intragastralis* — intragastric administration (via gavage); *Intravenosus* — intravenous administration; *per os* — oral administration; T4 — thyroxine

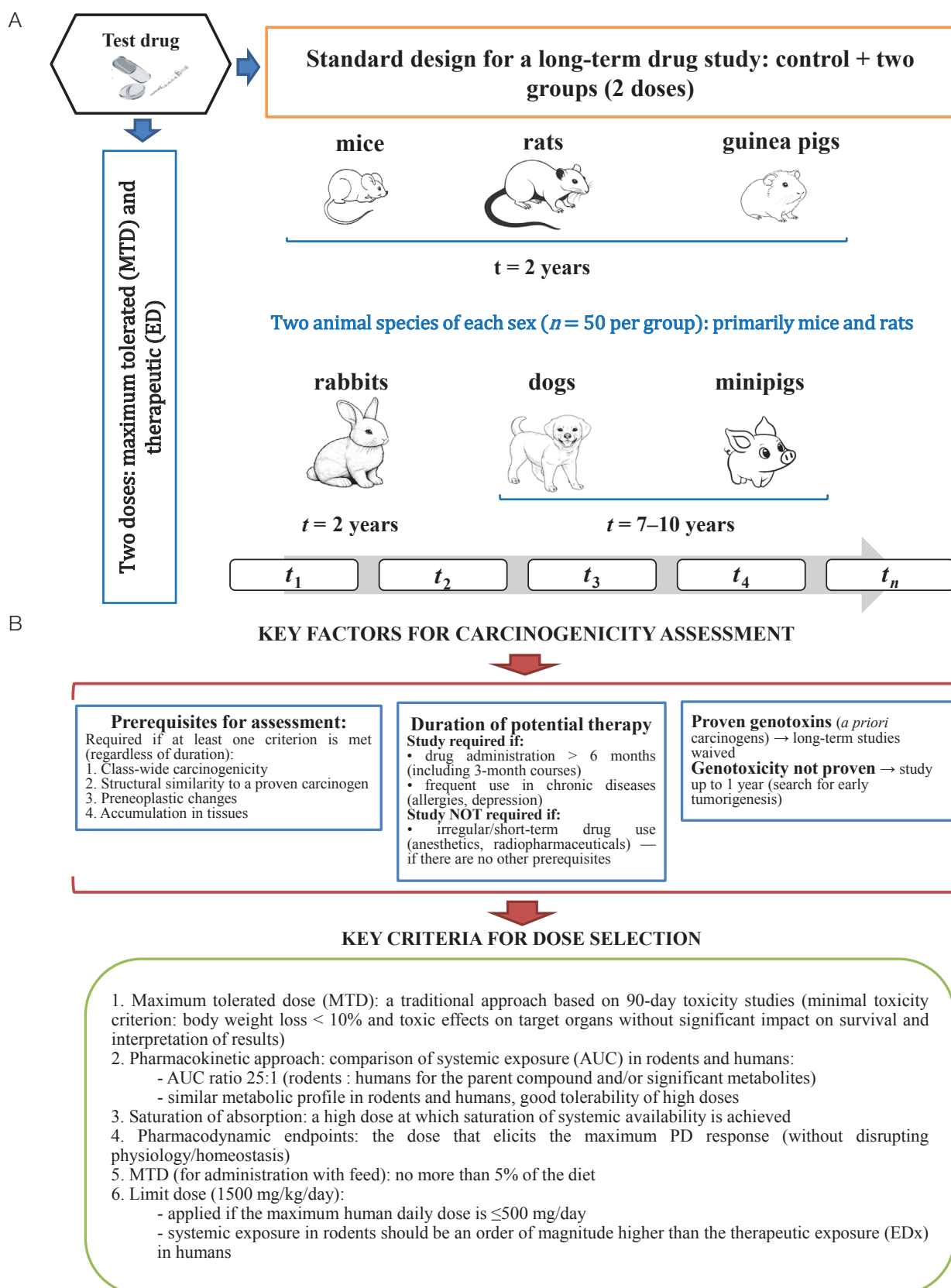


Figure prepared by the authors based on data from [16]

Fig. 5. Assessment of the carcinogenic potential of a medicinal product. A — experimental designs of pivotal preclinical carcinogenicity studies; B — general principles for carcinogenicity studies; t — observation time (years); ED_x (effective dose) — the calculated dose of the test substance that under a given administration regimen, causes a statistically significant or biologically significant pharmacological effect in a predetermined proportion of experimental animals (x%), determined relative to the control

Table 3. Methodological characteristics of experimental units in preclinical studies

№	Experimental unit	Definition	Application conditions	Design-consistency principle	Implementation example
1	Individual animal	One animal, randomly allocated to an exposure group	Assessment of acute and chronic toxicity; independent action of the substance	Each animal represents one data source; observations are independent	5–10 rodents per group; individual recording of body weight, blood biochemical parameters, etc.
2	Cluster (litter, brood)	Group of offspring from one exposed parent pair	Assessment of embryotoxicity, reproductive toxicity, and ontogenetic disorders	Parental exposure; offspring analysis; unit: litter/brood	10 females per group; assessment of 12 F1 offspring from 10 litters; averaging of indicators within the litter
3	Individual animal with repeated measurements	One individual with recording of parameters at several time points	Assessment of chronic toxicity, pharmacokinetics, and recovery period	The animal serves as its own control; measurements are not independent	Rodents (≥ 10 /group) and non-rodents (≥ 6 /group); blood collection on days 7, 14, 28, and 90 for parameter analysis
4	Part of an animal (anatomical region)	An isolated body area receiving local exposure	Dermal toxicity, phototoxicity, and local irritant action	Areas are anatomically isolated; systemic action is excluded	Application of the test substance to the right flank and the control substance to the left flank of the same animal

Table prepared by the authors based on data from [16]

points (t_1, t_2, t_3, t_4, t_n) (Table 3). This approach is used in studies where it is necessary to assess the parameter changes over time, particularly when evaluating chronic toxicity, delayed effects, or time-dependent responses. In this case, a time series of measurements is generated for each animal (numbered 1 through 6). Crucially, the statistical unit remains the individual animal, and repeated measurements are not considered independent units: they exhibit within-subject correlation. This design enhances the sensitivity of the study (since each animal serves as its own control) and reduces the required number of animals, which is consistent with the 3R principles.

A part of an animal as the experimental unit

In this design, the experimental unit is not the whole animal but a specific part (Table 3). This approach is viable when a region of the animal’s body is subjected to independent, local exposure that is isolated from other parts of the same animal. Examples include the application of different substances to different skin areas of the same animal (the left side of the back as a control vs. the right side for the test substance) and injections into different limbs. In this case, each treated site (anatomical region) is an independent experimental unit since exposure to one area does not affect (or minimally affects) the other. This design controls for inter-individual variability (each

animal serves as its own control) and reduces the total number of animals. A critical condition is that the selected areas must be sufficiently anatomically isolated, and the substance must not exert systemic effects.

Statistical analysis must account for the possible correlation between parts of the same animal by using models with repeated measures or mixed models with the animal as a random factor.

Objective safety assessment of pharmacological effects in preclinical studies requires not only data collection but also a pre-developed statistical analysis plan. This plan should regulate the sample size, collection, processing, analysis, and interpretation of results, enabling researchers to identify patterns, make informed conclusions, and generate reliable predictions [10–12].

Detailed statistical planning of the experiment is essential for a more comprehensive assessment of test drug safety and predicted efficacy. This structured approach helps researchers draw reliable conclusions about the benefits and risks of test drug use, while providing a foundation for decision-making on advancing to the clinical trial phase.

Statistical approaches in medicinal product studies

The statistical framework of drug development encompasses design planning and methods for

analyzing research results. Proper planning is the basis for obtaining reliable results, and the quality of evidence determines overall research value. The most developed area for applying statistical planning methods in the life cycle of a drug is the stage of its clinical development [7, 9, 16, 28]. The main statistical plans for registrational clinical studies include a summary of the design in the form of a hypothesis statement, a formula for sample size calculation, evaluation criteria, and analytical methods. This serves as a “blueprint” for the statistical design of pre-registrational and registrational clinical studies [33]. However, successful design of clinical studies relies heavily on preliminary data regarding drug pharmacodynamics and toxicity obtained in animal models. The key problem here is the correct extrapolation of data to humans. This can be attributed to the significant gaps persisting at the preclinical stage of drug development due to the absence of unified approaches to the statistical planning of experiments. This, in turn, leads to variability in operational procedures and methods of data analysis used to solve specific research tasks.

In this regard, practicing statisticians emphasize the need to standardize statistical plans applied in both preclinical and clinical studies [7, 16]. The purpose of such standardization is to reduce the risk of misrepresentations and erroneous predictions regarding the safety and efficacy of drugs.

Standardization of statistical approaches is especially important when planning translational and pivotal preclinical studies. To minimize errors in extrapolating data from animal models to humans, the statistical plans of translational investigations should be aligned with the design frameworks used in clinical trials (Table 4). This approach to systematizing statistical principles and design elements in preclinical and clinical studies will increase the likelihood of successfully translating pharmacological effects from animals to the human population.

By analogy with clinical studies, the core principles for the statistical planning of preclinical studies for data translation can be presented as follows [7, 16, 33]:

1. Definition of objectives: clear formulation of research questions and proposal of testable hypotheses.
2. Selection of appropriate models: selection of relevant *in vitro* and *in vivo* models that mimic the drug's action in humans. Statistical analysis should take the specifics of the chosen model into account.
3. Determination of sample size: calculation of the required number of animals or cell cultures to achieve statistical power sufficient to detect significant effects at a given alpha and beta error level.
4. Randomization: random allocation of animals or samples to treatment groups (drug, placebo, or control

group) to minimize systematic errors associated with individual differences.

5. Blinding: concealment of information about the assigned treatment from researchers conducting the experiment and analyzing the data to prevent subjectivity.

6. Selection of endpoints: identification of measurable parameters to assess drug effects (e.g., biomarkers, survival, or changes in physiological parameters).

7. Selection of statistical analysis methods: selection of appropriate statistical tests for comparing groups (e.g., Student's *t*-test, ANOVA, or non-parametric tests) based on the data type and study design.

8. Planning of repeated experiments: allowing for the possibility of repeating experiments to confirm the results and assess their reproducibility.

This approach will expand the possibilities for determining efficacy and safety parameters, while reducing the number of errors in extrapolating pharmacological effects observed in animals to the future patient population.

CONCLUSION

1. This analysis systematizes methodological approaches to the statistical and experimental planning of preclinical drug evaluations. The findings reveal that despite established international and national regulatory requirements (ICH, OECD, and GOST R), significant heterogeneity in statistical strategies persists in practice, reducing the reproducibility of results and complicating inter-study comparisons.

2. Based on this systematization, unified criteria for planning preclinical drug studies were proposed regarding the following aspects: sample size calculation depending on the type of endpoint and power; algorithm for selecting primary and secondary evaluation criteria; rules for selecting statistical models according to the experimental design; randomization and blinding procedures; and methods for the assessment and control of systematic errors.

3. The findings demonstrate that the unification of statistical planning at the pre-registration and registration stages of drug development is necessary for ensuring data continuity, increasing the reliability of pharmacological effect assessments, and correctly extrapolating outcomes from experimental models to the target population.

4. The proposed concept of systematization and unification creates a methodological basis for the standardization of preclinical study protocols for drugs, minimizing uninformative experimental designs, reducing time and resource costs at the development stage, and, ultimately, increasing the reliability of results significant for clinical practice.

Table 4. Statistical plans applicable in preclinical study designs

Study design	Hypotheses and sample size formulas for assessing the difference in means			Hypotheses and sample size formulas for assessing the difference in proportions		
	H_0	H_a	Formula	H_0	H_a	Formula
One-sample Equality of means (proportions)	$\bar{X} - \bar{X}_0 \neq 0$	$\bar{X} - \bar{X}_0 = 0$	$n = Z_{1-\alpha/2}^2 \times s^2$	$p - p_0 \neq 0$	$p - p_0 = 0$	$n = \frac{[Z_{1-\alpha} \times \sqrt{p_0(1-p_0)} + Z_{1-\beta} \times \sqrt{p(1-p)}]^2}{d^2}$
Two-sample Non-inferiority	$\bar{X}_1 - \bar{X}_2 \geq d$	$\bar{X}_1 - \bar{X}_2 < d$	$n = \frac{(Z_{1-\alpha/2} + Z_{1-\beta})^2}{(d/s)^2}$	$p_1 - p_2 \geq d$	$p_1 - p_2 < d$	$N = (Z_{1-\alpha/2} + Z_{1-\beta})^2 \times \frac{p_1(1-p_1) + p_2(1-p_2)}{d^2}$
Two-sample Superiority	$\bar{X}_1 - \bar{X}_2 \leq d$	$\bar{X}_1 - \bar{X}_2 > d$	$n = \frac{(Z_{1-\alpha/2} + Z_{1-\beta})^2}{(d/s)^2}$	$p_1 - p_2 \leq d$	$p_1 - p_2 > d$	$N = (Z_{1-\alpha/2} + Z_{1-\beta})^2 \times \frac{p_1(1-p_1) + p_2(1-p_2)}{d^2}$

Table prepared by the authors based on data from [16]

Note. n — sample size; s — standard deviation of the sample; d/s — effect size; d — magnitude of difference ($\bar{X} - \bar{X}_0$) or ($\bar{X}_1 - \bar{X}_2$), ($p_1 - p_0$) or ($p_2 - p_1$), representing the minimum expected significant difference; $\bar{X}$ — mean value of the response variable distribution (response to exposure); $\bar{X}_0$ — corresponding baseline (initial) mean; $\bar{X}_1$ — new treatment mean; $\bar{X}_2$ — standard treatment mean; $\bar{X}_1 - \bar{X}_2$ — magnitude of sample mean difference; p_0 — baseline proportion, p — new proportion, p_1 — new treatment proportion; p_2 — standard treatment proportion; $Z_{1-\alpha/2}$ and $Z_{1-\beta}$ — z-values for estimating validity and power, respectively; z is the value to be substituted into the sample size formula (it depends on the chosen desired significance level and power); p_0 is the baseline proportion, p is the proportion following exposure, and $d = p - p_0$ is the difference under the alternative hypothesis (where d is the change relative to baseline); p_1 is the new treatment proportion, p_2 is the standard treatment proportion, and $d = p_1 - p_2$ — difference under the alternative hypothesis (clinically significant magnitude). A one-sided test concentrates 5% in one tail of the normal distribution (z-score 1.645 or higher). The critical value at 5% is ± 1.96 since 5% consists of 2.5% in each of the two tails of the standard distribution. For a 90% confidence level, $Z_{\alpha/2}$ is determined as follows: given a 90% confidence level, α would be $1 - 0.9 = 0.1$, that is, $\alpha/2 = 0.1/2 = 0.05$; the corresponding critical z-value is determined from the z-table; for an absolute value of 0.05, $z = 1.64$ (for one-sided hypothesis testing, $Z_{0.1/2} = 1.645$). Given a 95% confidence level, α would be $1 - 0.95 = 0.05$, that is, $\alpha/2 = 0.05/2 = 0.025$, and the corresponding critical z-value for the absolute value of 0.025 is 1.96 (for one-sided hypothesis testing, $Z_{0.05/2} = 1.96$). Given a 99% confidence level, α would be $1 - 0.99 = 0.01$, that is, $\alpha/2 = 0.01/2 = 0.005$, and the corresponding critical z-value for the absolute value of 0.005 is 2.58 (for one-sided hypothesis testing, $Z_{0.01/2} = 2.58$). This assumes group sizes of at least 30 and a distribution sufficiently close to normal (e.g., an excess kurtosis coefficient of no more than a few units¹⁴).

¹⁴ Gerasimov AN. Medical statistics. Moscow: MIA; 2007.

References

- Thomas R, Paules R, Simeonov A, Fitzpatrick S, Crofton KM, Casey WM, et al. The US federal tox21 program: a strategic and operational plan for continued leadership. *ALTEX*. 2018;35(2):163–8. <https://doi.org/10.14573/altex.1803011>
- Aladysheva ZI, Belyaev VV, Beregovykh VV, Brkich GE, Greybo SV, Demina NB, et al. *Industrial pharmacy. The product creation path*. Moscow; 2019 (In Russ.). EDN: [KWJJUK](#)
- Ko MJ, Lim CY. General considerations for sample size estimation in animal study. *Korean Journal of Anesthesiology*. 2021;74(1):23–9. <https://doi.org/10.4097/kja.20662>
- Kargopol'tseva DR, Kryshen KL, Makarova MN, Makarov VG. Regulatory and methodological aspects of preclinical and clinical studies of animal feed additives. *Laboratory Animals for Scientific Research*. 2018;3:100–12 (In Russ.). <https://doi.org/10.29296/2618723X-2018-03-11>
- Suvorov AY, Latushkina IV, Gulyaeva KA, Bulanov NM, Nadinskaya MYu, Zaikin AA. Basic aspects of meta-analysis. Part 1. *Sechenov Medical Journal*. 2023;14(1):4–14 (In Russ.). <https://doi.org/10.47093/2218-7332.2023.14.1.4-14>
- Waine K, White C, Dean R, Hudson C, Huxley J, Brennan ML. Assessing the feasibility of retrospective and prospective clinical audit in farm animal veterinary practice. *Veterinary Sciences*. 2021;8(4):62. <https://doi.org/10.3390/vetsci8040062>
- Vasiliev AN, Niyazov RR, Gavrishina EV, Dranitsyna MA, Kulichev DA. Problems of planning and conduct of preclinical trials in the Russian Federation. *REMEDIUM*. 2017;9:6–18 (In Russ.). EDN: [ZOLGIP](#)
- Moher D, Dulberg CS, Wells GA. Statistical power, sample size, and their reporting in randomized controlled trials. *JAMA*. 1994;272:122–4. <https://doi.org/10.1001/jama.1994.03520020048013>
- Thabane L. *Sample size determination in clinical trials*. HRM-733 Class Notes. Hamilton: St. Joseph's Healthcare; 2004.
- Karkishchenko NN. *Biomedicine Alternatives. Vol. 1. Fundamentals of biomedicine and pharmacomodeling*. Moscow: Interacademic Publishing House; 2007 (In Russ.). EDN: [SOPUYN](#)
- Shreder OV, Bunyatyan ND, Goryachev DV, Subayev RD, Engalycheva GN, Kuznetsova AD, Kosenko VV. Mathematical prediction of drug efficacy in nonclinical studies. *Bulletin of the Scientific Centre for Expert Evaluation of Medicinal Products. Regulatory Research and Medicine Evaluation*. 2022;12(3):315–30 (In Russ.). <https://doi.org/10.30895/1991-2919-2022-12-3-315-330>
- Vasiliev PM, Golubeva AB, Koroleva AR, Perfiliev MA, Kochetkov AN. *In silico* forecast of the toxicological and pharmacokinetic characteristics of medicinal compounds. *Safety and Risk of Pharmacotherapy*. 2023;11(4):390–408 (In Russ.). <https://doi.org/10.30895/2312-7821-2023-11-4-390-408>
- Aban IB, George B. Statistical considerations for preclinical studies. *Experimental Neurology*. 2015;270:82–7. <https://doi.org/10.1016/j.expneurol.2015.02.024>
- Krewski D, Acosta D, Andersen M, Anderson H, Bailar JC, Boekelheide K, et al. Toxicity testing in the 21st century: a vision and a strategy. *Journal of Toxicology and Environmental Health, Part B*. 2010;13(0):51–138. <https://doi.org/10.1080/10937404.2010.483176>
- Levitt T. Exploit the product life cycle. *Harvard Business Review*. 1965;43:81–94.
- Reynolds PS. *A guide to sample size for animal-based studies*. Hoboken: John Wiley & Sons; 2023.
- Ampilogova IN, Karlina MV, Makarov VG, Makarova MN. The interconnection between pharmaceutical development and preclinical research. *Drug Development & Registration*. 2023;12(2):155–63 (In Russ.). <https://doi.org/10.33380/2305-2066-2023-12-2-155-163>
- Sieburg H, McCutchan J, Clay O, Caballero L, Ostlund J. Simulation of HIV infection in artificial immune systems. *Physica D: Nonlinear Phenomena*. 1990;45:208–27. [https://doi.org/10.1016/0167-2789\(90\)90184-Q](https://doi.org/10.1016/0167-2789(90)90184-Q)
- Vasilyev AN. Good preclinical study, as an obligatory stage in design and clinical use of new medicinal preparations. *Antibiotics and Chemotherapy*. 2012;57(1–2):41–2 (In Russ.). EDN: [QCRZIE](#)
- Henderson VC, Kimmelman J, Fergusson D, Grimshaw JM, Hackam DG. Threats to validity in the design and conduct of preclinical efficacy studies: a systematic review of guidelines for *in vivo* animal experiments. *PLoS Medicine*. 2013;10(7):e1001489. <https://doi.org/10.1371/journal.pmed.1001489>
- Tannenbaum J, Bennett B. Russell and Burch's 3Rs then and now: the need for clarity in definition and purpose. *Journal of the American Association for Laboratory Animal Science*. 2015;54(2):120–32.
- Garber J, Barbee R, Bielitzki J, Clayton L, Donovan JC, Hendriksen CFM, et al. *Guide for the care and use of laboratory animals*. Washington, D.C.: National Academies Press; 2011. <https://doi.org/10.17226/12910>
- McInnes E, ed. *Pathology for toxicologists: principles and practices of laboratory animal pathology for study personnel*. Hoboken: John Wiley & Sons; 2017. <https://doi.org/10.1002/9781118755174>
- Lambert L, Newman D. Construct development and validation in three practical steps: recommendations for reviewers, editors, and authors. *Organizational Research Methods*. 2023;26(4):574–607. <https://doi.org/10.1177/10944281221115374>
- Huang W, Percie du Sert N, Vollert J, Rice ASC. General principles of preclinical study design. In: Bepalov A, Michel M, Steckler T, eds. *Good research practice in non-clinical pharmacology and biomedicine*. Cham: Springer; 2019. https://doi.org/10.1007/164_2019_277
- Cohen J. *Statistical power analysis for the behavioral sciences*. Mahwah: Lawrence Erlbaum Associates; 2013.
- Lenth RV. Some practical guidelines for effective sample size determination. *The American Statistician*. 2001;55(3):187–93. <https://doi.org/10.1198/000313001317098149>
- Petri A, Sabin K. *Visual statistics in medicine*. Moscow: GEOTAR-MED; 2021 (In Russ.).
- Karp N, Reavey N. Sex bias in preclinical research and an exploration of how to change the status quo. *British Journal of Pharmacology*. 2019;176:4107–18. <https://doi.org/10.1111/bph.14539>

30. McCarthy MM. Incorporating sex as a variable in pre-clinical neuropsychiatric research. *Schizophrenia Bulletin*. 2015;41:1016–20.
<https://doi.org/10.1093/schbul/sbv077>
31. Makarova MN, Matichin AA, Maticina AA, Makarov VG. Animal choice strategy for research. Report 1: animal choice based on phylogenetic relationships. *Laboratory Animals for Science*. 2022;5(2):58–70 (In Russ.).
EDN: [SBSPNR](https://elibrary.ru/sbbspnr)
32. Charan J, Kantharia N. How to calculate sample size in animal studies? *Journal of Pharmacology and Pharmacotherapeutics*. 2013;4:303–6.
33. Shreder OV, Goryachev DV, Merkulov VA. Basic principles for calculating the required number of participants in clinical trials. Part 1. Common approaches (review). *Regulatory Research and Medicine Evaluation*. 2024;14(3):338–50 (In Russ.).
<https://doi.org/10.30895/1991-2919-2024-14-3-338-350>
34. Kryshen KL, Gushchin YaA, Faustova NM, Gushchina SV, Ustenko ZhU, Katelnikova AE, et al. Interpretation of toxicological study data for determination of no observed adverse effect level (NOAEL). *Regulatory Research and Expertise of Medicinal Products*. 2025;15(3):262–77 (In Russ.).
<https://doi.org/10.30895/1991-2919-2025-15-3-262-277>
35. Makarova MN. Preclinical studies in the drug development cycle. *Regulatory Research and Expertise of Medicinal Products*. 2024;14(3):248–50 (In Russ.).
<https://doi.org/10.30895/1991-2919-2024-14-3-248-250>

Authors' contributions. All authors confirm that their contributions meet the ICMJE criteria for authorship. The greatest contributions are distributed as follows: Olga V. Shreder — conceptualization, manuscript drafting, editing, manuscript approval; Dinara M. Ivashova — preparation of tables, figures, text formatting.

Authors:

Olga V. Shreder, Cand. Sci. (Biol.), ORCID: <https://orcid.org/0000-0002-7926-6033>

Dinara M. Ivashova, ORCID: <https://orcid.org/0000-0001-6111-6961>