Quality by Design in Clinical Trials and Development Projects

Requirements and Recommendations for Designing, Conducting, and Monitoring Clinical Trials that Provide Quality Information

Table of Contents

Anastassios D. Retzios, Ph.D
Bay Clinical R&D Services, LLC
2417 Canyon Lakes Drive
San Ramon, California, 94582, USA

Tel: (925) 480-7497
Email: aretzios@adrclinresearch.com

Quality by Design in Clinical Trials and Development Projects

Requirements and Recommendations for Designing, Conducting, and Monitoring Clinical Trials that Provide Quality Information

I. Abstract

Key regulatory agencies such as the FDA and EMA began efforts in the first decade of this century to improve quality in pharmaceutical development. In the process, the International Conference on Harmonization (ICH) issued several guidances to industry based on the key notion of "Quality by Design" (QbD). The guidances applied QbD to both the manufacturing and quality control of pharmaceutical entities and the preclinical and clinical development of these compounds. This document focuses on QbD efforts in clinical development.

Obtaining high-quality results in clinical development programs is extremely challenging. This is because clinical trials involve numerous "actors" in the conception, planning, execution, and analysis of these studies. In addition, many programs involve substantial unknowns, including the pathophysiology of the disease under study, the event rate of the symptom being followed, and/or the clinical significance of the selected endpoint.

This brief examination of the issues related to the effort to build and emphasize quality in clinical studies focuses on: (a) the regulatory framework for Quality by Design (QbD) in clinical studies; (b) the significance of the target product profiles (TPPs) in the pursuit for quality; the methods and the effort for identification of risks to quality; (c) the sponsor's organizational issues that lead to quality failures; and (d) the conceptual issues that may affect clinical study design. Many quality failures occur because of our incomplete knowledge of the pathophysiology of the disease studied, the mechanism of action of the drug tested, and/or the event rate that underpins the primary endpoint. Therefore, this review examines methods for managing and mitigating scientific and conceptual risks, including selecting appropriate animal models, natural history studies, disease progression models, and real-world data collection. In that context, the review also discusses quality risks related to the choice and type of endpoints.

There is also a discussion of operational quality risks in clinical studies, risks that occur during study conduct and data collection. Such risks include the proper identification and management of study violations and deviations, and the selection of the type of monitoring for these clinical trials. The review also briefly discusses the quality issues and biases introduced by interim analyses and the early termination of clinical trials.

This review concludes with a discussion of the possible effects of artificial intelligence (AI) on clinical development and the quality of clinical trials.

II. Introduction

The key aim of the QbD initiative is to proactively build quality into clinical studies to enhance data quality and design integrity, avoiding post-trial efforts to correct problems and inaccuracies. The rationale for this effort is the high failure rate in the clinical stage of development, which prevents many potentially efficacious drugs from reaching the market. Medical schools in universities, clinical study organizations, regulatory agencies, and collaborations among them, such as the Clinical Trials Transformation Initiative (CTTI), are driving efforts to improve the quality of clinical studies.*

Quality by design (QbD) in clinical studies is a response by industry and regulators to the high rate of failure in drug development

The clinical development of new therapeutic agents has a remarkably high rate of failure. Smietana et al. (2016) [1] analyzed the R&D productivity of the pharmaceutical industry by examining the fate of more than 9,200 compounds from 1996 to 2014. These researchers determined that approximately 90% of potential drugs that enter clinical development are destined to fail. There were similar observations by Harrison (2016).[2] This was despite efforts by various leading pharmaceutical companies to increase the success rate by more carefully selecting their candidate drugs, such as AstraZeneca's 5R process.[3] There were certain improvements, of course. Dowden and Munro (2019) [4] were able to show, based on customer relationship Management (CRM) system data, that success rates in late-stage development — from phase III through to launch — have increased from just under a one-in-two chance of getting to market at the beginning of the decade to almost two in three in the most recent time period for which data are available, 2015–2017. On the other hand, the same authors found that success rates in Phase 2 had not improved.

Despite the efforts s by pharmaceutical companies to better position test drugs, compounds that appear safe and efficacious in preclinical testing fail to demonstrate efficacy or adequate safety in clinical studies. The failure rate, despite modest improvements, remains high. Therefore, such failures may occur not because the test compound lacks efficacy or has unacceptable toxicity but because of inadequacies in the clinical development program. The sheer complexity of planning and executing these studies, due to the multiplicity of actors and processes, makes maintaining quality difficult. Therefore, methods to improve the rate of clinical development success by improving data quality and strengthening the scientific validity of studies warrant careful examination.

* The Clinical Trials Transformation Initiative (CTTI) is a public-private partnership co-founded by Duke University and the U.S. Food and Drug Administration in 2007. It currently includes >70 organizations and has > 500 partners in the clinical studies business area (CTTI Homepage - CTTI).

What is meant by the term "quality" in clinical studies, and why would quality increase the likelihood of success? If a clinical study has been performed ethically, the subjects have been provided with detailed and accurate informed consent, all procedures have been conducted per protocol, and all data have been collected and accurately reflect the source material, is this a "quality" clinical trial? Not necessarily. The scientific validity of the study is also an important consideration. This validity is reflected in the enrollment criteria, the disease progression elements assessed (signs, symptoms, biomarkers, organ function, etc.), the selected endpoints, and the adequacy of the trial's sample size. However, most clinical studies are not individual units but exist within the nexus of an overall development effort. Each clinical trial is an element of an overall clinical development plan with defined goals. In most cases, development targets marketing authorization for the intervention/therapeutic compound from a regulatory authority. A clinical study that does not address the correct questions regarding the drug's safety and efficacy within the attributes and parameters set by the regulatory authority has not met the definition of quality, regardless of the robustness of the data it collected.

Considering the above, an appropriate definition of **"quality by design" in clinical studies is the combination of planning, operational, conceptual, and organizational processes and procedures that are science-driven and provide robust and accurate data, fully reflective of the actual record, which are adequate for the appraisal of the safety and efficacy of the tested intervention/therapeutic agent.** The Clinical Trials Transformation Initiative (CTTI) defines quality as "the absence of errors that matter". I do not think this is an all-encompassing definition, as a clinical study may have no errors that matter, yet still be uninformative or unhelpful in defining the efficacy or safety of a novel drug.

This document attempts to define the processes that imbue quality not only to individual studies but also to the clinical development program overall. It also discusses methods and approaches to improve quality and prevent failures.

III. The Regulatory Framework

The progressive revisions of the ICH E6 Guidance on Good Clinical Practices and the emphasis on quality by design

The requirements for clinical studies (both of conducting and reporting them) have typically been included in sections 312 and 312 of Title 21 of the US Code of Federal Regulations (21 CFR 312 & 314). These sections define the current Good Clinical Practice (GCP) regulations. In further defining these regulations, key regulatory agencies under the International Conference on Harmonization (ICH) umbrella have developed additional guidelines. In the ICH efficacy guidelines, the ICH E6 defines the key principles and requirements of GCP. The original guideline (E6(R1)) was published in 1996. However, progress in the methodology of clinical trials and pharmaceutical development necessitated a substantial amendment in 2016 [E6(R2)]. In this revision, an

annex (Annex 1) has been added to provide guidance on quality in the conduct and reporting of clinical trials.[5] Annex 1 of E6(R2) goes beyond the general principles of Good Clinical Practice (GCP). It provides detailed instructions on the roles and obligations of ethics committees, the investigator, and the sponsor, including the specific documents required for the full application of GCP processes. In addition, it standardizes the information that should be presented in clinical trial protocols, investigator brochures, and clinical study reports.

Because ICH E6 is so important to pharmaceutical development, ICH issued a 3rd revision of the guideline in January 2025 [E6(R3)]. E6(R3) contains a major revision of Annex 1 of the E6(R2) guideline and introduces a substantial number of new elements. It emphasizes, more than its predecessors, the need for quality by design. This guideline was expected to eventually include Annex 2, which will address non-traditional elements of clinical trials, such as decentralized activities, alternative designs, and the use of Real-World Data (RWD). Annex 2 of E6 is still in draft. The 2nd draft (Step 2) was released in November 2024. Please note that the FDA has already issued guidance to industry on the use of Real-World Data in pharmaceutical development. This guidance is discussed in Section V.C.1 of this document.

The ICH E8 addresses specific issues on quality by design

The ICH E8 guideline addresses general considerations for clinical studies. A recent revision of this guideline (ICH E8(R1)), finalized in October 2021, includes specific suggestions/instructions for building quality by design (QbD) in clinical trials. A key reason for updating the E8 guideline was to provide a table that lists the Critical-to-Quality Factors (CtQFs) and to cross-reference them to other ICH E (Efficacy) guidelines. Although the finalized E8(R1) includes an annex on the process for identifying CtQFs, it does not provide a comprehensive cross-reference to these CtQFs beyond the summaries of typical activities in sections 5 and 6 of this guideline. It is, however, a very valuable document, and it should be carefully examined by clinical development professionals.

The ICH E9 guideline on quality risk management is essential for designing a risk-based approach to development

Since properly powering a clinical study to provide reliable data is an important proactive quality consideration, the ICH E9 guidance, Statistical Principles for Clinical Trials, is essential. The FDA has expanded on certain issues of this guideline, including the recent guidance on multiple endpoints in clinical studies.

Although it does not address clinical studies per se, the ICH Q9 guideline on Quality Risk Management is also essential in clinical development. In fact, it is not possible to design quality in clinical studies without risk management. A risk-based approach, strongly indicated in clinical development, directs attention, monitoring resources, and testing in areas of risk. Without defining the risks, such an approach is not possible. Therefore, risk identification and proactive mitigation are essential activities for building quality into clinical study design.

In 2011, the FDA and EMA launched a pilot program to assess the effects and potential complications of implementing the QbD guidelines. The program concluded in 2017; as a result, a series of Q&A documents was produced to further clarify the QbD guidelines. Q&A (1)[6,78] was published in 2013 and has been amended several times, including this year; it covers extended information on the quality target product profile (QTPP), critical quality attributes (CQAs) and additional items related to manufacturing processes and analytical methods. Q&A (2) was also published in 2013 and covered issues regarding Design Space both for active substances and finished products. Q&A (3) was published in 2014 and addressed issues such as the details required for risk assessment and Design Space. The final report concluded that there was a "solid" alignment of these agencies regarding the implementation of concepts included in the ICH Q8, Q9, and Q10 guidelines

Although the above-mentioned guidances, guidelines, and Q&A documents comprise the umbrella regulatory framework, major regulatory agencies (such as the FDA and EMA) have issued additional guidance to industry on various issues related to improving clinical development. We will refer to these guidelines in individual sections of this document.

IV. Other Initiatives

To build quality prospectively in clinical studies, one would be remiss not to mention the Clinical Trial Transformation Initiative (CTTI), a Duke University–FDA partnership. CTTI was founded in 2007 and is partially funded by the FDA, although its membership has expanded to include a substantial number of pharmaceutical and biotech companies and contract research organizations. Certain CTTI publications can be very helpful in interpreting regulatory documentation.

V. Methodology for Quality-by-Design (QbD) in Clinical Studies

Several key activities are necessary to build quality into a clinical development effort. A crucial one is to fully conceptualize the final goal of development in a comprehensive target product profile (TPP); To examine the main areas of quality risk; and a third would be to define the "critical to quality factors" (CtQFs) in the proposed clinical development effort, specific to each clinical study protocol.

A. Target Product Profile

Detailed target product profiles (TPPs) are essential in facilitating quality by design in clinical development

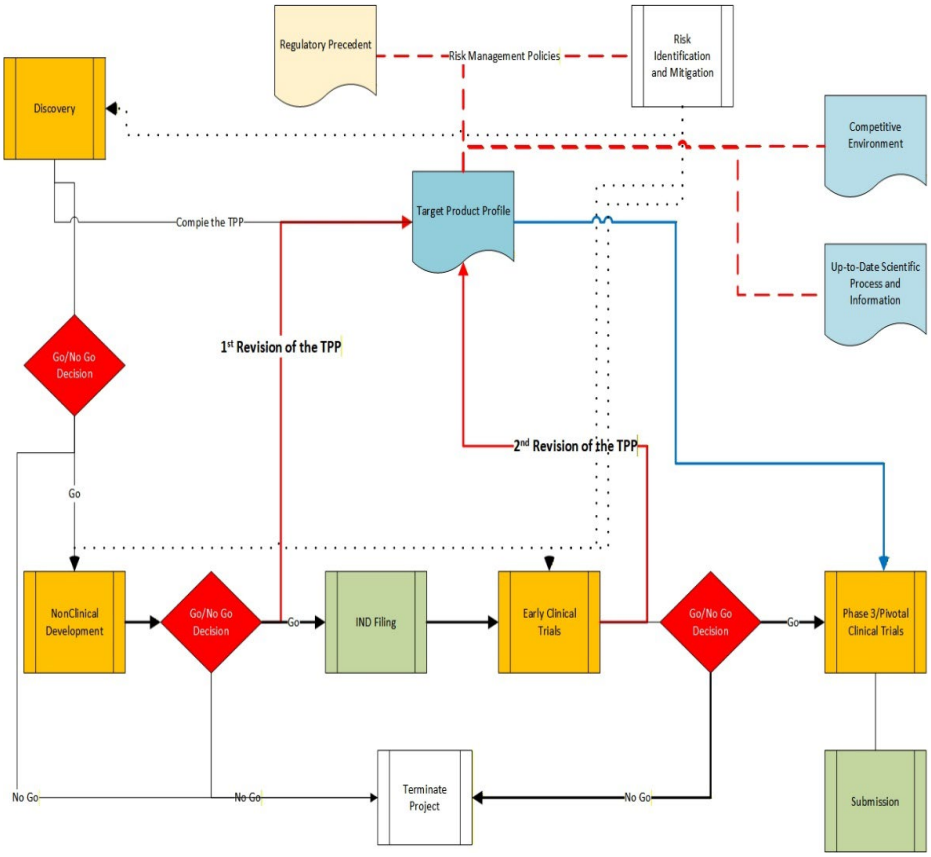
Achieving quality by design in clinical trials is challenging without a comprehensive target product profile (TPP) that clearly defines the intended outcome of the clinical development process, including the indication that is both desirable and feasible for the test compound. A well-designed TPP should contain, in precise regulatory terminology, the intended claims for physicochemical characteristics, safety, and efficacy of the test compound/device. The TPP should provide the acceptable range of essential attributes/targets of the investigational product (maximal, base, minimal). The TPP should also include extensive annotations explaining the rationale for the values assigned to various product attributes. The role, design elements, and methods to complete an effective TPP are discussed by Retzios (2020).[9,] In addition, more information on the function and use of TPP is available in various publications and monographs.[10]

A complete TPP is usually accompanied by the Quality TPP (QTPP) and a Risk Identification and Mitigation document. The QTPP identifies key quality elements of the formulation and manufacturing of the test drug/biologic/device (the CMC element of development). The Risk Identification and Mitigation module of the TPP details the possible risks to the development process and rates them for likelihood and severity. It also provides alternative routes through the development process if the identified risks materialize.

Because a TPP introduces clarity to the development effort (if done correctly), it can enable specific go/no-go decisions at specific stages of the project. In addition, such go/no-go decision nodes can be introduced at any time if the therapeutic/device in question fails to achieve the minimally specified value for a key product attribute or claim.

Although the TPP is a valuable tool for designing and achieving quality in clinical trials and for facilitating pertinent comments from regulatory agencies, it remains underused. This problem is caused by a variety of factors, such as organizational and operational constraints, a lack of focus among the development team, and a lack of training.[10]

Figure 1: The Key Role of Product Target Profile in a Quality-Driven Clinical Development Effort



B. Identification of Quality Risks

A risk-based approach to development concentrates attention and effort in areas of risk to quality

It is important to have a good understanding of the quality risks before commencing clinical development, especially risks that would affect the adequacy of the informed consent process, the protection of patients, the reliability and completeness of the data collected, and their interpretability in conjunction with the objectives of the clinical trial in question and the overall clinical development program. A risk-based approach to clinical trials concentrates management attention on areas of potential risk. In identifying these risks, one needs to define the following:

- The reason for the risk
- The severity of the risk (minor, moderate, major)
- The likelihood of the risk (Unlikely, Possible, Likely/Probable, Certain)
- The actions that need to be taken to eliminate or reduce the severity of the risk (mitigation actions)
- The consequences of these actions (several actions may lead to different regulatory considerations)

In general, risks can be organizational, conceptual, and operational. A brief analysis of these risks follows below:

1. Organizational Risks

Organizational problems in various companies inhibit communication and do not promote easy sharing of information, thus increasing the risk that problems may not be solved in a timely manner

To achieve quality by design in clinical trials, it is important to take a good, hard look at the organization and its ability to achieve and maintain quality. In the recent ICH E8(R1) harmonized Guidance on General Considerations on Clinical Trials, regulatory agencies support the notion of establishing “a culture that supports open dialog.” Only through open dialogue and the opportunity for all stakeholders to challenge each other can quality be both designed and achieved in clinical studies. However, in many organizations, this is a really difficult task, not only because of organizational barriers (of which there are many), but also because of personalities. Therefore, it is essential to create this "culture of open dialogue" in pharmaceutical development by removing organizational barriers and empowering team members. A good way to do this is to map the organization's SOPs in a way that allows all stakeholders to provide input and ensures that input is properly weighted by the development team and company management.

a) Clinical Study Management Team

Given the number of people involved in drug development, organizational and management issues are important because the process requires input from individuals with different areas of expertise. Therefore, organizing and managing the development teams to optimize their operations is always a remarkably difficult task.

The development team should be staffed and organized with a clear understanding of each member's responsibilities. It should empower all members to communicate freely and share opinions. How best to organize "parallel" teams with members of all appropriate skill sets from various departments remains a challenge when "vertical" (departmental) organization creates potential conflicts. Various types of "matrix" schemes exist as companies strive to create the optimum environment for drug development.

b) Training

A dedicated team cannot achieve much without appropriate training. Therefore, training is essential in Good Clinical Practice (GCP), study protocols and associated procedures, and the symptoms and pathophysiology of the disease addressed in the development project. This training should be comprehensive, documented, and repeated at regular intervals. All new team members must be fully trained in all aspects of the disease treated by the test drug and the clinical tests involved. It is not uncommon for clinical research associates who

monitor investigational sites to not understand the implications of certain test values (normal or abnormal) and, therefore, to fail to seek necessary clarification and explanation from site personnel.

Certain training should also be extended to investigational sites that participate in the development project. The investigator meetings and study initiation procedures at sites provide some training, but this often falls short of being comprehensive.

Unfortunately, training in biotech is one of the activities usually cut or drastically reduced when organizations try to achieve "economies". Considering the multiplicity of "moving parts" in a clinical study, personnel turnover and the occasional flurry of amendments can be "fatal" to the project.

c) SOPs

SOPs that address clinical development should carefully map the organization's actual responsibilities and activities. They should also be developed to minimize human error and increase quality.

2. Conceptual/ Scientific Risks

Conceptual/scientific risks are the most difficult to overcome in the process of enhancing the quality of the development program

In the author's experience, most development projects fail because of conceptual and scientific reasons. This is not to say that the organizational and operational risks are negligible; of course not. But the operational gestalt is far better understood, and its rules are well defined. On the other hand, conceptual issues present many varied and difficult challenges. Our knowledge (or even, occasionally, the definition of certain diseases) is fragmentary, absent, or contradictory. Fully understanding the pathophysiologic pathways affected by a disease remains challenging because biological systems are complex. Our "conceptual tools" leave much to be desired. For example, we are now on our third definition of sepsis;[11] we are not doing that much better in ARDS[12], two diseases with high mortality and challenging development history. Also, clinical design approaches and endpoints may be suboptimal in certain situations, failing to provide appropriate information or accurately measure a drug effect. In summary, a well-designed and adequately controlled clinical trial remains a difficult goal to reach.[13] Below, we will attempt to identify the areas where these risks may occur and discuss possible ways to mitigate them.

a) Poorly Understood Progression and Physiopathology of the Disease

Clinical studies leading to the marketing approval of any therapeutic agent need to be well-designed and controlled. This cannot happen when the scientific information in the therapeutic area of development is deficient. One of the most difficult aspects of clinical development is

structuring a clinical program for a disease/syndrome whose pathophysiology is poorly or incompletely understood. Some therapeutic areas, due to our limited understanding of the diseases involved, make clinical development extremely difficult. These fields often have a high failure rate for test drugs/biologics. These include (but are not limited to) sepsis, acute respiratory distress syndrome, stroke, Alzheimer's disease, Parkinson's disease, and various cancers. Beyond these well-known areas of severe developmental challenge, thousands of rare disorders remain poorly understood due to limited interest or very small patient populations. Some approaches discussed below may reduce the risk in these cases.

Inadequate understanding of the disease is one of the main causes of clinical study failures in high-failure clinical presentations such as sepsis and ARDS

(1) Understanding the Disease: Staying on Top of Scientific Advances

Mitigating risks from an imperfect understanding of the disease's pathophysiology can be challenging. However, for those "brave enough", there is a substantial reward if successful. Science is advancing, of course, and our understanding of a particular disease is improving with successive failures in development. Therefore, it is important to examine these advances and understand how they may "modify" development and regulatory precedent in the targeted disease. For example, there are now more accurate definitions of sepsis and ARDS.[14,15] They are not perfect, but they are an improvement over what was there before and, to a degree, can facilitate further development by more accurately defining the eligibility criteria for patients who may benefit from treatment.

(2) Understanding the Disease: Utilization of Appropriate Preclinical Models

In addition, substantial evidence should be placed on preclinical pharmacology data. Advances in this area have also produced new models that approximate human disease better than earlier ones. Using such models may provide important information for planning the clinical program.[16]

Beyond examining the interplay between new information and prior development, other potential approaches exist for diseases that are not well understood. These consist of natural history studies, disease progression models, and real-world evidence collection.

(3) Understanding the Disease: Utilization of Natural History Studies

These observational studies, usually designed by the sponsor, attempt to define the natural history of a disease. If carefully planned and executed, they can provide substantial information

about a poorly understood disease: its epidemiology, symptoms, penetrance, genetics, environmental factors, progression, and outcome measures.[17] Importantly, when designing such studies, one needs to be comprehensive in the information sought and potentially identify biomarkers relevant to the natural history of the disease. Beyond the substantial information these studies can provide to the development program, they can occasionally be used as external controls when ethical considerations require it, and the number of patients is too small to allow concurrent control groups. The FDA recently released a draft guidance on the design and conduct of natural history studies to assist drug development.[18]

(4) Understanding the Disease: Utilization of Disease Progression Models

Well-constructed disease-progression models (DPMs) can assist substantially in the design of successful clinical studies. They can be incorporated in Model-Informed Drug Efforts (MIDDs)

These models use mathematical algorithms to quantitatively determine, over a defined period, parameters related to disease pathophysiology (patient characteristics, symptoms, organ function, biomarkers) and outcomes, based on available information about disease progression.[19] Depending on what one wants to evaluate, these models can also include treatment-induced modifications to disease progression (and effects on symptoms, organ function, and biomarkers), based on the treatment agent's mechanism of action and general pharmacological principles.[20] These simulations can provide important information for clinical study design. For example, in such a model, one can assess the effects of various enrollment criteria and thus select a population more likely to respond to treatment, thereby limiting the sample size (this approach, of course, has regulatory consequences). Of course, if the mechanism of action of a test compound is known, one can potentially obtain useful information on dosing and dosing frequency. Such information may shorten the duration of the learning phase (Phase 2) of drug development. As information about the disease increases, these models can become increasingly elaborate and predictive. For example, several models for ARDS provide useful information on the predictive capacity of certain biomarkers.[21] Various progression models are also available for Alzheimer's Disease.[22]

Disease progression models (DPMs) can also be incorporated into Model-Informed Drug Development (MIDD) efforts. MIDDs[†] have

[†] Model-Informed Drug Development comprises a set of statistical and disease models, pharmacokinetic and pharmacodynamic information, information from preclinical and clinical studies to structure an effective development framework that may require fewer studies or resources to provide convincing answers to the posed development hypotheses.

received extensive attention from development groups in the pharmaceutical industry and the regulatory agencies. Although they are outside the bounds of this document, AI/ML utilization in these models is further explored in Section V (Artificial Intelligence and Quality by Design in Clinical Studies)

(5) Understanding the Disease: Collecting Real-World Data and Evidence (RWD/RWE)

Real-world data can provide useful information on the progression of the disease, key characteristics of patients and relevant biomarkers.

RWD are data on patients with the disease under investigation and their treatment, collected from various sources, mainly registries, health records, medical claims data, and/or publications. Analysis of such data leads to Real-World Evidence (RWE) on the safety and efficacy of the treatment agent in question. The FDA has issued guidance on the utilization of RWD and the specific requirements for supporting RWE in regulatory submissions.[23,24] The FDA guidance on the use of RWD and RWE in regulatory submissions is specifically for new indications for drugs that have already been approved and for satisfying post-approval requirements.

Regarding the use of RWD for investigational drugs, these can be used as data from natural history studies to define more accurately the progression of the disease, the demographics and characteristics of patients (to facilitate the incorporation of appropriate inclusion criteria), select appropriate biomarkers, as well as clinically relevant endpoints, all key elements of a well-designed study. RWD can also be used to determine treatment parameters for various drugs used to treat the disease in question and to select an appropriate control for clinical studies. It should be considered, however, that registries, the main source of RWD, and publications may suffer from various biases stemming from their design (for example, by targeting specific demographics). Natural history studies, designed specifically to provide the information needed to design a clinical development program, may be a better approach. Of course, combining information from RWD and from natural history studies may well be the best approach.

The risks in a therapeutic area with a high failure rate should be fully covered in the Risk Definition and Mitigation plan. One would need to provide various "exits" from the program when there is adequate information that the project is failing to meet set minimal values for key attributes. Another approach is to structure a flexible study design and institute futility analyses.

Poor understanding of the drug's mechanism of action can result in substantial errors not only in study design but also in the overall direction of the development program

b) Poorly Defined Mechanism of Action

It is difficult to implement quality controls without a clear understanding of the drug and its mechanism of action, specifically, the cellular/molecular entities it affects (the drug target). Identifying these should be a very important step in drug discovery and development. A good number of beneficial drugs do not have a well-understood mechanism of action, but for most of these, the "target" of the drug, the molecular entity affected or modified by the drug's action, is known.[25] Efforts have been ongoing to define such targets for a variety of drugs for which these targets were not identified, even during the process of approval.[26] Since the beginning of this century, drug discovery has mainly been based on identifying specific targets that, when acted upon by the appropriate drug/biologic, (target-directed discovery).[27] However, there has also been renewed interest in phenotypic drug discovery that dispenses with defining the drug target or the molecular process by which the drug affects the disease.[28] However, the author believes that defining the drug's mechanism of action (irrespective of the method of initial discovery) is very helpful in developing a preclinical/clinical testing strategy. Without a clear understanding of the drug's dynamics, one may target the wrong indication or patient population. In that context, clinical development efforts are doomed to fail. A good example is the development of Dimebon (Latrepirdine) by Medivation (San Francisco, California, USA).[29] Dimebon was assumed to be an effective NMDA receptor inhibitor. In Alzheimer's Disease, early synaptic dysfunction is associated with NMDA receptor-dependent synaptic depression.[30] A phase 2 study provided hopeful results, but a Phase 3 study failed to detect any activity. Further investigations into Dimebon's activity revealed that its NMDA receptor inhibition was too weak to have an effect in the targeted indication.[29] Pfizer's torcetrapib is another good example of a development failure caused by a poor understanding of the drug's mechanism. Torcetrapib inhibited CETP (cholesterol ester transfer protein), which greatly increased HDL ("good") cholesterol and lowered LDL ("bad") cholesterol. Researchers did not fully understand the mechanism linking HDL levels to cardiovascular protection. HDL turned out to be much more complex than a simple biomarker. Increasing HDL concentration did not necessarily improve HDL function or promote cardiovascular benefit. In addition, torcetrapib had side effects, including raising blood pressure and aldosterone levels that were not "discovered" until relatively late in the development effort. The result was that in the Phase 3 clinical trial, patients receiving torcetrapib experienced more deaths and more cardiovascular events despite the apparently favorable cholesterol profile. The study and the project were discontinued.[31]

These examples clearly illustrate that poor understanding of the mechanism of action, the specific targets the drug affects, and the kinetics of interaction is likely responsible for many drug "efficacy" failures, the most common type of failure in development.[2]

In summary, it is imperative to define, as well as possible, the drug's mechanism of action and its likely targets (intended and unintended) during the discovery process. The earlier the development team has the relevant information, the easier it is to design appropriate clinical studies.

c) Selecting the Proper Endpoints

Non-validated or non-regulatorily sanctioned surrogate endpoints will need to be carefully considered and should be supported by substantial and convincing data of their validity

Endpoints are measures of treatment outcomes and, therefore, key elements of the development process. There is a substantial volume of information on endpoints; this section is not meant to be a full exploration. It incorporates recommendations and tools for selecting appropriate endpoints to improve the quality of clinical studies and the overall development project.

It is strongly recommended to start thinking about the endpoints of the learning and pivotal phases of clinical development at the planning stage, specifically during the assembly of the target product profile. If the validity and usefulness of the selected endpoints are discussed early in the development effort, the resulting considerations can inform planning across the preclinical and clinical program. In the planning stage, one must also identify potential risks and complications associated with specific endpoints and develop strategies to mitigate them within the clinical development program. If, for example, the development effort would depend on an unvalidated or previously non-utilized surrogate endpoint, it is incumbent on the managers of the development effort to define the data and studies required to convince the regulatory agencies of the validity of such an endpoint. The earlier one starts considering this issue, the higher the quality of the resulting studies will be.

Endpoints need to be clinically relevant and sensitive to intervention

It is imperative to use specific endpoints from regulatory precedent, if available and current; pivotal studies targeting superiority or non-inferiority to already approved products may depend on them. In therapeutic areas with no previous developments, or for drugs with a novel mode of action, selected endpoints must be carefully considered. They must be clinically relevant, sensitive to intervention, and accepted by regulatory bodies, policymakers, and the treatment community. Endpoints can be clinical or nonclinical, measuring clinical changes or reporting on proximal/distal pharmacodynamic measures and/or various biomarkers.

(1) Clinical Endpoints

Clinical endpoints directly assess the patient's symptoms, physical function, and/or survival and, as such, are the preferred option when feasible. They can be reported by the treating physician or by the patient. The emphasis on clinical endpoints is that they must be "meaningful". Because many clinical programs utilize endpoints of limited clinical value, certain tools have been developed to define what is meaningful in specific clinical endpoints.

(2) Nonclinical Endpoints

The FDA has provided a number of aids, tools and guidances regarding the use of surrogate endpoints, especially in oncology

If nonclinical endpoints have a proven association with the clinical condition being assessed, they are surrogate endpoints. Surrogate endpoints can be used when the specific pharmacodynamic measure is strongly associated with a clinically meaningful outcome and appropriate studies validate that relationship. The FDA provides a list of surrogate endpoints that serve as the basis for approval and licensure.[32] However, non-validated surrogate endpoints with no well-defined connection to clinical outcomes are used extensively in accelerated development programs. For surrogate endpoints that are new or not included in this list, the FDA proposes a Type C meeting with the sponsor to discuss key elements of the proposed surrogate endpoint. The agency has published considerations that the sponsor should address in such a meeting.[33] Surrogate endpoints are used extensively in oncology drug development. The "gold standard" in oncology endpoints is "overall survival (OS)", but the observation period required to define it in each study is lengthy. Therefore, a variety of surrogate endpoints are typically used, such as Objective Response Rate, Complete Response, Disease-Free Survival, Progression-Free Survival, and Time to Progression. However, their relationship to OS is often poor and varies across cancers and populations. To improve the relationship between these endpoints and OS, regulatory authorities have issued various guidances, such as the FDA's "Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics."[34] Overall, specific criteria must be considered when using surrogate endpoints; it is important to evaluate these criteria before deciding to proceed with such endpoints.

Various guidelines, tools, and other resources can assist those who are involved in the clinical development program in defining the optimal endpoint. For example, the FDA, in collaboration with the NIH, has produced a useful guide for the utilization of biomarker-based surrogate endpoints called BEST[‡] (for Biomarkers, Endpoints, and other Tools).[35] As discussed previously, the FDA has issued guidance to industry regarding the use of surrogate endpoints in oncology clinical trials.[34] This guidance discusses various considerations for using such endpoints, as well as statistical considerations for analyzing them. The EMA maintains a site that lists all guidelines issued for determining efficacy across various diseases and symptoms.[36] The European Network for Health Technology Assessment (EUnetHTA) has produced several documents analyzing endpoint issues that those designing clinical studies should carefully consider.[37]

Composite endpoints may be required in the study of certain diseases. Although they may be able to provide increased statistical power in certain cases, they may also make interpretation of results difficult or impossible, unless they are carefully assembled

(3) Composite Endpoints: Advantages and Disadvantages

In certain cases, because of disease complexity or the relative rarity of a cardinal symptom or major outcome, a single endpoint (clinical or surrogate) is inadequate to define the intervention's effectiveness and treatment outcomes. In these cases, more than one endpoint is needed to assess the drug's effect on various symptoms, organs, or functions. In such cases, clinical study designers could use either a composite endpoint that combines multiple aspects of the disease into a single entity or multiple endpoints to assess specific elements of the outcome individually.

Composite endpoints usually consist of two or more individual events combined into a single outcome that represents the clinical benefit of the treatment. Because using composite endpoints can significantly affect the quality of a study or development program, it is worth examining this issue in some detail.

Composite endpoints are employed in two key circumstances: (a) in diseases that present various disparate symptoms, and (b) when a single measurement for the outcome of treatment is inadequate in describing its overall effect. Diabetic gastroparesis is such a disease. It presents various symptoms such as nausea, vomiting, epigastric pain, early satiety, and others, not all of which may be manifested in any given patient. One would need to describe all of them to get an appropriate understanding of the efficacy of the test medication; (b) when endpoint component outcomes have a very low rate of occurrence, combining them into a single endpoint increases the

‡ It should not be confused with FDA CBER's BEST initiative

statistical power of the study. For example, composite endpoints are frequently used in cardiovascular and heart failure studies because they capture more treatment-related events in less time than a single endpoint, such as survival, would.[38]

Composite endpoints carry many potential risks, especially when used to inform treatment decisions, which are the key driver of most development projects.[39] Not all components of a composite endpoint may have the same clinical "value." Interpretation may be vexing. For example, in a clinical study for the treatment of chronic obstructive pulmonary disease (COPD), researchers who investigated the use of corticosteroids in this disease combined death from all causes with intensification of treatment. Based on this combination, the treatment group was statistically significantly better than placebo. However, if one examined just mortality (death from all causes), there was no difference between treatment and placebo.[40] Because composite endpoints can be difficult to interpret, especially when the event rate of certain cardinal symptoms differs substantially from others, they should not be used if a single primary endpoint can be used. Composite endpoint scores are difficult to interpret clinically. A good example of the dangers of composite endpoints is the clinical program investigating the efficacy and safety of domperidone (Janssen Pharmaceutica, Beerse, Belgium), a dopamine antagonist, in diabetic gastroparesis.[41,42] The study group created a composite endpoint by combining the known cardinal symptoms of the disease, such as vomiting, nausea, early satiety, distention/bloating, and abdominal pain. Symptom intensity was scored on a scale of 0 to 3. The double-blinded phase of the study found that the aggregate deterioration symptom score was 1 point higher in the placebo group than in the domperidone group (1.84 vs 0.85, respectively; P=0,025).[41] But what was the actual clinical benefit of this difference? The composite endpoint made it impossible to determine this, and the FDA did not approve the drug for marketing in the US.

Although composite endpoints may be essential in certain development projects, it is imperative to structure a discussion between the sponsor and the relevant regulatory authority; concurrence may be reached if the sponsor presents a well-founded rationale for using the composite endpoint. Without this concurrence, the use of the composite endpoint needs careful evaluation.

d) Multiplicity

Many circumstances require clinical studies to investigate a variety of events that cannot be combined into a composite endpoint. Interventions in various diseases can change multiple parameters that need to be assessed. This is the case, for example, in Alzheimer's disease, where both cognition and functioning must be evaluated in different ways. Therefore, using multiple primary endpoints becomes necessary. Consequently, one must address the issue of multiplicity when analyzing study results. This is a serious issue because multiple comparisons increase the chance of erroneous associations unless the study design and statistical methods are adjusted to account for it. For this reason, regulatory agencies have issued guidance on addressing multiplicity. The FDA recently released such a guidance,[43] and the EMA introduced a draft update to its original guideline on this issue.[44]

Failing to account for multiplicity may reduce the study's power. The conservative approach (the Bonferroni correction) of simply dividing the alpha (the significance level) by the number of comparisons may substantially increase sample size and render the study impractical.[45] Other, less conservative approaches can yield acceptable sample sizes. These include the Holm,[46] Hochberg,[47] Hommel,[48] Benjamini, and Hochberg[49] adjustments and others. Careful consideration of the statistical analysis approach is essential here.

e) The Study Design Elements' Effects on Endpoint Values

Not understanding and quantitating carefully the effects of elements of study design, such as enrollment criteria, on the endpoints can easily undermine the statistical validity of a study

After selecting the appropriate endpoint(s), the key step in any clinical study protocol is deciding how to detect the effect of the clinical intervention. Some studies are purely descriptive, but most studies, especially those in the pivotal stage, involve hypothesis testing.[50] Usually, hypothesis testing compares the null hypothesis (no difference between tested groups) to the alternative hypothesis (a difference between tested groups). A clinical study should provide enough statistical power to discern which hypothesis, the null or alternative, is true. Usually, a significance level of 5% is adopted to minimize Type I error (false positive: difference detected when there is no difference), and a power of 80-90% to minimize Type II error (false negative: no difference is detected when there is actually a difference).

The statistical considerations involved in this exercise require certain estimates of the primary endpoint to calculate sample size. Statistical considerations and testing require an estimate of the value and variability of the primary endpoint in patients receiving the control treatment (to frame the null hypothesis). These estimates typically come from earlier clinical studies or epidemiological surveys. For framing the alternative hypothesis, an estimate is required for the likely

(or desired) value of the endpoint following intervention with the test treatment. Such estimates are typically derived from earlier studies in humans or animals.

Unfortunately, these estimates are often inaccurate because the effort to define them has not been diligent enough, or because specific study design elements may modify them (occasionally, to an unknown degree). This is why many studies fail, as is the case in sepsis. Typically, all-cause mortality at 28 days is the primary endpoint in most sepsis studies. However, sepsis mortality has been decreasing substantially over the last twenty years, and a failure to take this into account resulted in many underpowered studies that failed to pass regulatory master. For example, in the study examining the role of Eritoran (Eisai Co., Tokyo, Japan) in severe sepsis, the observed mortality in the placebo arm (27%) was lower than the anticipated mortality in the study design (40%). Thus, because the sample size was based on an estimated placebo mortality of 40%, the trial did not have enough power to detect a significant difference between treatment and placebo.[51] But the problem is not confined to a single occurrence. Kaukonen et al.(2014)[52] examined the mortality rate of severe sepsis in Australia and New Zealand between 2000 and 2012. These researchers found that mortality progressively decreased during this period and that studies using mortality estimates from previous years were routinely underpowered.

In other cases, enrollment criteria significantly modify the key outcome measure. This has been the case with the pivotal study of rBPI21 (Xoma Corporation, Emeryville, California, USA) in pediatric meningococcemia. Again, the key endpoint variable, mortality, was substantially reduced in the control arm by enrollment criteria, resulting in an underpowered study that could not support a filing for marketing approval.[53]

f) Dealing with Uncertainty: Adaptive Designs in Clinical Studies

Because of problems defining the event rate in the control population and the extent of the treatment effect in clinical studies, various methods have been introduced to modify elements of the study design while the study is in progress. These design elements, defined a priori in the clinical study protocol, are called adaptive or flexible. There are various such designs, fully explored in the literature.[54,55] Some introduce interim reviews by independent boards to assess futility or adjust the sample size. These are highly recommended when there is substantial uncertainty about key estimates for the statistical considerations of a pivotal clinical study, which may lead to setting the sample size too low and thus underpowering the study.

3. Operational Risks

The sound execution of a well-thought-out design is extremely important. In many cases, if adequate attention is not paid to the operational end of a clinical trial, implementation deficiencies can defeat the intentions of a solid design. In addition, operational risks can be minimized by carefully considering key elements of study design that may affect the rate of intercurrent events.

Various metrics can help study management staff monitor the proper conduct of a clinical study. Establishing these quality and performance metrics allows study management personnel to assess investigative sites' performance in implementing the study. Key metrics involve patient recruitment rate, patient retention, patient adherence to visits, and the accuracy and completeness of data collection and reporting. Study management must decide the acceptable performance level for each metric and take decisive action when performance drops below it. These metrics allow managers to identify possible risks and assign more resources to higher-risk areas.

In this paper, we will focus primarily on operational issues and metrics that affect data quality and, therefore, the validity of estimates of the study's primary endpoint(s). Operational issues of primary importance are (a) protocol violations; (b) missing data; (c) rate and number of patients withdrawn and/or lost to follow-up; (d) the type of study monitoring.

a) Deviations and Violations

Key metrics for assessing clinical study quality include the number and rate of clinical protocol deviations and violations, the amount of missing data, and the number of patients withdrawn or lost to follow-up. These can affect the assessment of specific measurements and the interpretation of the data, including the primary endpoint. The FDA, in various guidances, defines these events as "intercurrent events".[56]

There are various definitions for protocol deviations and violations. Although the existing FDA and ICH guidances do not include a specific definition of violations and deviations, one can be found in the ICH E3 Q&A (R1) companion document.[57] In that document, the regulatory authorities define violations as "important protocol deviations," which "are a subset of protocol deviations that might significantly affect the completeness, accuracy, and/or reliability of the study data or that might significantly affect a subject's rights, safety, or well-being".

In summary, a deviation occurs when certain planned events, such as study visits/assessments, do not occur in accordance with the protocol, but the deviation is not significant enough to render the subject ineligible for the per-protocol dataset or to affect the interpretation of the study results. Deviations that may potentially affect the clinical question of interest or introduce bias should be described as violations.

Violations may include abnormalities in informed consent, non-adherence to the study inclusion/exclusion criteria, unscheduled unblinding, use of prohibited medication, substantial non-adherence to visits, which renders this patient's dataset impossible to evaluate for the primary endpoint, and others.

Study management personnel should prospectively define what constitutes a deviation or violation in the conduct of each study. For example, in a pharmacokinetics/pharmacodynamics study, the loss of certain assessments and visits may prevent definitive determination of PK and PD parameters, while other intercurrent issues may not result in an adverse event. Study-dependent issues need to be specifically addressed a priori in the study management/monitoring plan and in the study statistical plan.

Other key problems that may affect the estimate of the primary endpoint and interpretation of the results of the study are (a) missing data and (b) the rate of patient withdrawal or that of patients lost to follow-up.

Missing data can result from several causes. Patients may miss scheduled visits or assessments; patients may not complete questionnaires fully or appropriately; data may also be lost because tissue samples have been lost, degraded, or contaminated, or because testing equipment has malfunctioned. Data is also lost because patients withdraw or are lost to follow-up. Patients may exit the study for a variety of reasons: the treatment may be ineffective; it may cause severe or serious adverse events; patients may be withdrawn by the investigator for non-adherence or may relocate to an area that makes it impossible to visit the investigative site; patients may die or develop another health issue for which treatment is incompatible under the provisions of the protocol.

There are actions that study designers and operational staff can take to limit the number of deviations/violations. Many of these intercurrent events can be prevented or their rate minimized by proper attention to elements of the study design. The length of the study, the extent of testing for each patient, the number of study visits, the proper “windows” for each test/study visit, and/or “tolerance limits” for specific variables need to be carefully assessed to minimize intercurrent events. Regarding missing data, there are, of course, imputation methods, but they are not always the best approach or a good approximation of reality.

b) Monitoring Clinical Studies

Although designing an effective clinical study is important for quality, so is the monitoring of its conduct. This is essential to providing high-quality data to answer the key questions posed by the study.

Clinical trial monitoring by the sponsor, the investigative site, and other relevant authorities (such as institutional review boards or ethics committees) is a key component of Good Clinical Practice (GCP), a set of principles within the regulatory framework that govern the design and conduct of clinical studies.[58] The FDA provides a list of guidance documents for clinical studies at Clinical Trials Guidance Documents | FDA.

Proper monitoring of the clinical study ensures that:

- The procedures of the study outlined in the clinical study protocol are consistently applied in the participating investigative sites, and the safety of the participating patients is not compromised
- The personnel of the investigative sites have the education, expertise, and capability to perform the procedures of the clinical study
- The resulting data are collected appropriately and in accordance with the applicable laws and regulations.

The type of monitoring performed for a clinical study depends on the quantification of risk and the availability of resources. Briefly, these types of monitoring can be summarized as follows:

- On-site monitoring that ensures 100% source data verification by appropriately trained study personnel
- Off-site, remote centralized monitoring (with or without statistically assisted monitoring)
- On-site or remote (or combination of on-site and remote) risk-based monitoring

A risk-based approach to monitoring clinical studies can provide savings in operational costs, but much depends on the training and acuity of the managing personnel for its successful implementation. In complex systems like clinical studies, the simplicity of the 100% source data verification in study monitoring has substantial advantages.

Following the release of the ICH E6 guideline and the initial FDA guidance on monitoring in 1988,[59] it became the norm for most biotech and pharmaceutical companies to perform on-site monitoring targeting 100% source data verification (SDV) of entries in case report forms (CRFs) at regular intervals in the investigative sites conducting the clinical study. This approach requires substantial resources in both personnel and funds. It is a conservative approach and should be the obvious choice for any corporation or organization with the funds and personnel to support it. However, in the last decade or so, significant attention has been paid to a risk-based approach to clinical study monitoring (Risk-based Monitoring; RBM). Various regulatory agencies

have provided guidance or comments on RBM. The FDA issued guidance to industry on RBM in 2013.[60] The agency also recently released a "Questions and Answers" follow-up document to this guidance.[61] Obviously, the purported advantage of RBM over an effort that targets 100% SDV is reduced costs for sponsors. The EMA has also issued a reflection paper on this monitoring method. Despite significant encouragement from regulatory agencies, risk-based monitoring is not as widely used as expected because of key issues that any company attempting to use it must address. In the first place, it is far more complex to implement than the on-site 100% SDV; it would require certain revisions to clinical operations SOPs; it is far more technologically demanding; its success (or failure) will strongly depend on specific assumptions regarding the potential risks; in addition, there is a lack of personnel familiar with it or well-trained for its implementation that a company can access.

An alternative to on-site monitoring is remote centralized monitoring in which a central location, either at a sponsor or contract research organization, collects and evaluates data and performance metrics for the clinical study from the participating investigative sites. This method requires the use of an electronic data capture (EDC) system. Centralized monitoring was widely used during the Covid-19 pandemic. Statistical algorithms can substantially support centralized remote monitoring by identifying misconduct, anomalies, outliers, and errors. In several circumstances, centralized monitoring can be combined with risk-based monitoring, further reducing the need for on-site SDV.

The key question, of course, is how good RBM is compared to 100% SDV. Few definitive studies address this. Andersen et al.[62] evaluated the error rate of on-site 100% source data verification and risk-based monitoring involving targeted partial source data verification across three phase III clinical trials, totaling 3 million data fields. The researchers found that on-site 100% SDV yielded an error rate of 0.25%; risk-based targeted and partial SDV resulted in an error rate of 0.53%. The difference was statistically significant ($P<0.0001$). The authors discussed that no method yielded a 0% error rate and that the difference between the industry standard 100% on-site SDV and the risk-based monitoring approach was "marginal.

Klatte et al.[63] compared several monitoring methods in interventional clinical trials, including the typical 100% on-site SDV, risk-based monitoring, central monitoring with triggered on-site visits, central statistical monitoring with annual on-site visits, and targeted SDV either on-site or remote. There was low certainty regarding the effectiveness of RBM versus extensive on-site SDV, although resource utilization was significantly higher for the latter. The same low certainty of

effectiveness applied to the comparison of central monitoring versus regular on-site visits with 100% SDV. However, despite the low certainty of evidence that these methods result in effective monitoring, the authors concluded that both RBM and central monitoring are promising strategies that require more extensive examination.

The key problem is, of course, that RBM attempts to reduce the resources required for effective study monitoring, which should be attractive to companies with limited resources. However, these companies may lack the technology and personnel needed to use RBM. If these essentials are not present, it is advisable to fall back to 100% on-site SDV.

c) Data/Safety Monitoring Boards

DSMBs play a key role in monitoring for safety and in introducing changes to key study parameters in flexible designs

A key element of effective monitoring is the role of Data/Safety Monitoring Committees/Boards (DMCs/DSMCs/DSMBs). This type of board is usually composed of independent experts who review safety data, including adverse events, changes in vital signs, and clinical laboratory parameters. The FDA has recently issued draft guidance on the organization and function of these boards, expanding on the 2006 guidance on DMCs.

DMCs can have access to unblinded data and interim analyses. They can provide several recommendations to the study sponsor; these may include early stopping of a clinical study for safety grounds (the number and severity of adverse events), futility, or a higher-than-expected treatment effect. A detailed charter for this board is essential; it should clearly outline the responsibilities of the DMC members, its composition, the mode of communication with the sponsor, documentation of these communications, and precise rules for early study stopping or other proposals.

In studies that use flexible designs, the DMC can potentially be assigned adaptation responsibilities if there is no adaptation committee, provided that the DMC has the appropriate expertise. In this case, the board can examine comparative efficacy data during interim analyses and recommend actions such as increasing sample size or changing primary and secondary endpoints.

The proper functioning of this board is essential for the quality of the study and the protection of patients. In studies using a flexible design, it is imperative that there is a “Chinese wall” between this board and the sponsor, and that the required changes are clearly spelled out in the standard operating procedures. All DMC interactions with the company should be carefully documented.

d) Planned Interim Analyses and Early Termination of Clinical Studies

Although early termination of studies is typically not addressed as a "quality by design" issue in clinical trials, this relatively frequent event can mask several quality issues that should be addressed by the responsible steering committee or a DMC. Several discussions address the effect of early termination on study data; the treatment below is a brief introduction to these issues, not a comprehensive review.

A substantial number of studies are terminated before completion, and this certainly influences the quality of the data collected. Many factors can lead to the early discontinuation of clinical studies.[64] The most common is slower-than-expected patient accrual.[65,66] Studies are also terminated early for safety reasons when a higher-than-expected rate of serious adverse events occurs. They may also be terminated if the monitored event occurs at a much lower rate than anticipated.[67] In such cases, there is little to do beyond ensuring that all appropriate data are collected correctly and assessing the causes of the low accession rate or the higher-than-anticipated safety concerns.

However, in this discussion, we will focus primarily on early discontinuations following protocol-mandated interim analyses.[§] In this case, the number and anticipated timing of interim analyses, sample size, statistical methodology, and precise instructions for the DMC need to be carefully detailed in the study protocol. All these should be considered carefully to minimize potential bias.[**]

Assuming that the interim analysis methodology is well spelled out, such early terminations can occur because of:

- Insufficient promise for the treatment effect
- Treatment is superior to control, and continuation of the study presents ethical issues

In early stopping for efficacy, the primary goal is to minimize the risk of a Type I error. Two key approaches exist: Group sequential methods with predefined bounds[68] and alpha spending functions. The former are well established and carry very low regulatory risk. The latter allows more flexibility in the setting of the timing of the interim analyses.

[§] Interim analyses can also be used for various requirements of adaptive study approaches, such as sample size recalculation, changes in primary endpoint, study enrichment and others. Such issues fall outside the purpose of this monograph

[**] Bias is a significant issue in clinical trials. Although not explored in detail in this paper, practitioners should be aware of it. Interim analyses, even if carefully constructed statistically, may introduce potential bias to the conclusions of the study

When the interim analysis indicates insufficient promise of treatment benefit, clinical studies are terminated for futility. Futility analyses utilize group sequential methods (as discussed above) or conditional power. In a review of studies terminated early for futility, Walter et al. (2020) determined that the early terminations were reasonable in the cases examined, but the documentation of the process was inadequate.

Interim analyses pose a significant challenge to quality. They may introduce a bias depending on when they occur within the study. For example, if a study is terminated early for efficacy, the smaller sample size may not be representative of the patient population as a whole. If that population includes mostly patients from only a fraction of the participating sites, it may not represent the typical treatment approach across a wide variety of institutions that treat the disease/syndrome in question. Therefore, those who manage and analyze the clinical study must carefully address potential bias. A great example of the bias introduced by early termination is the story of the development of recombinant activated Protein C (raPC). The pivotal clinical study of this biologic agent in severe sepsis was terminated after the 2nd interim analysis (1690 patients randomized and treated out of the 2280 planned). This interim analysis found that the raPC-treatment arm showed a decrease in mortality compared to placebo.[69] Based on these results, the study was terminated at this stage, and this therapeutic agent was eventually approved for marketing. However, a follow-up study that randomized 2640 patients and was completed normally did not find any mortality benefit for raPC in severe sepsis;[70] This biologic was eventually withdrawn from the market in 2011.

VI. Artificial Intelligence and Quality by Design in Clinical Studies

A. Introduction and Definitions

There is currently a lot of excitement in many fields of scientific endeavor about the use of Artificial Intelligence (AI) and Machine Learning (ML). AI consists of computer processes designed to imitate human intellectual reasoning, discover patterns, and learn from past experiences (data). ML is a key AI process that builds models based on existing data and uses them to make predictions. ML is a subset of AI that can analyze large amounts of data, extract patterns, and make predictions without being explicitly programmed for these tasks. These technologies have generated significant interest in their potential to transform and accelerate development efforts. There is a lot of hype regarding AI and ML in clinical development (and many ardent advocates), but what is the reality? This is a difficult question to answer. We are at the beginning of using these modalities, and key regulatory agencies appear to have “thoughts and ideas” but not much definitive guidance, at least not yet.

B. Regulatory Response to AI in Pharmaceutical Development

The various regulatory agencies are trying to understand the challenges and opportunities of AI/ML in clinical development

There is little doubt that the advent of AI and ML in drug discovery and development has challenged regulatory agencies such as the FDA and EMA. The FDA has recently released a couple of concept documents intended to facilitate discussion on the challenges of using AI in medicinal products, as well as on the use of AI and ML in drug discovery and development. One of these documents discusses how the agency intends to establish a framework to regulate the use of AI in medical products.[71] The FDA has also released a discussion paper regarding the potential use of AI and ML in the design and conduct of clinical trials.[72] The agency has identified several areas in drug discovery and development where AI/ML may be assistive.

The EMA has also produced several documents on the use of AI/ML in medicinal product development. EMA's key reflection paper on the use of AI in the medicinal product lifecycle examines the use of AI/ML in product development and marketing authorization and in the post-authorization period.[73] The EMA proposes a risk-based approach for AI/ML in drug development, and it promises that advice on risk management will be included in future regulatory guidances.[74]

Although the use of AI/ML in drug discovery is probably better delineated, its role in development, given existing regulations, remains ambivalent. For the purpose of this discussion, "development" will be defined as preclinical and clinical testing of a pharmaceutical compound

C. AI/ML in Preclinical Testing

Preclinical testing is an essential part of development for any drug or biologic. It determines the efficacy, toxicity, and pharmacokinetic/pharmacodynamic (PK/PD) profile of these medicinal entities across various animal species prior to their entry into first-in-human testing. Technologies that can accelerate this stage of development and/or provide more dependable results can have a critical impact on overall development. Thus, the interest in AI/ML, again within the milieu of regulations, which certainly create specific bounds for these technologies

The New Approach Methodologies (NAMs) for preclinical testing may be helped substantially by AI/ML, but such a possibility is purely theoretical at this stage

The FDA's Modernization Act 2.0 of 2023 has removed the mandatory requirement for animal pharmacology and toxicology studies and has allowed the use of validated non-animal-based methods, providing impetus for the use of AI/ML in preclinical testing.[75] To be sure, the FDA has outlined a "roadmap" to achieve this, and its key provisions currently concern monoclonal antibodies.[76] The "new approach methodologies" (NAMs) introduced with this modernization include AI-based/assisted toxicology, organs-on-chips, and human-relevant in vitro systems to assist in exploring the pharmacology and toxicology of drugs and biologics before first-in-human testing. To be acceptable to the agency, NAMS need to be validated (or be capable of validation) and/or possess a strong scientific rationale. The truth is that, at the time of writing this document, we are nowhere near that, and the FDA has not even amended the code of regulations governing the filing of INDs. This may happen in Modernization Act 3.o, whenever the FDA plans to release it. For the time being, NAMs are an interesting notion, and the use

of AI in any of them is alluring, but nothing is close to reality. Those involved in development should be aware of this; and they should be aware that this is a fast-changing field that needs to be monitored closely[††]

Beyond the potential use of AI/ML in NAMs, it can also be useful in providing information in other areas of preclinical testing, such as identifying the best animal models to use and how data from testing in these models is interpreted and utilized.[77] However, one should be very careful regarding possible conclusions or predictions from AI/ML. It is certainly strongly recommended that results are examined carefully by a toxicologist or other well-trained professionals. The problem is that the data sets available to be used by these methodologies are small and many are proprietary. In addition, any predictions of toxicity based on AI/ML information must dovetail with a mechanistic explanation of how the drug works, to make them plausible.

PK/PD modelling may be assisted by AI routines, but there are potential problems with ML models that lack biological considerations

Another potential application of AI/ML is in PK/PD modelling. Although it is relatively routine to predict the pharmacokinetics of compounds destined for clinical trials based upon preclinical studies, the real value of such predictions only comes when these are linked to drug effects, i.e. pharmacodynamics. In summary, PK/PD modelling is an essential activity during preclinical testing. Of course, PK/PD modelling is not new. PK/PD information can certainly be collected *in vivo*, although such an effort is certainly challenging and expensive. Models can complete information partially collected from animal testing. The modelling effort, based on the accumulation of information from actual experiments can provide an interesting capability in defining the drug "movement" within the body through various tissues. These models are mathematical algorithms that can predict the concentration of a given drug/biological in various compartments. A more intelligent dosing can be devised for both animal and human testing based on that information. In addition, knowing or estimating the amount of drug/biological compound in the tissue of interest and correlating it to the recorded reaction of the said tissue provides valuable information on the utility of the tested drug. Several animal species can be modelled, if of interest. A variety of

†† It is interesting to note here that in 2016, the EMA released a document on its ideas of the reduction of animal testing. This 3Rs document (Replacement, Reduction, Refinement) is in "spirit" similar to the FDA's road map for reduced testing. But there was very little that was concrete in that document. There were certain suggestions for regulatory acceptance of NAMs but matters were not concrete. So, recently the EMA has released a concept paper that intends to provide some ideas to the industry as to where the agency is heading. In the concept paper, there is discussion about the utilization of microphysiological system (MPS) models, but the agency is frank in stating that the incorporation of these technologies is challenging with a critical hurdle being the establishment of robust qualification of these systems for context of use (CoU). The agency proposed to draft a document for tightening definitions and terminology and produce a road map for the general qualification of NAMs. For the purposes of all these, a drafting group may be organized to amend the existing guideline. An industry consortium (the International Consortium for Innovation and Quality in Pharmaceutical Development) has been organized and the summary of a 2022 Workshop with the FDA is generally available to those interested (Baran et al; ALTEX 2022; 39: 2)

pharmacokinetic/ pharmacodynamic models and associated routines have been developed over the last few decades. There are two key approaches with the utilization of ML in modelling. One of these tends to replace the current approach with data-driven approaches, while the other simply assists traditional modeling processes. In general, data-driven ML models lack biological/physiological considerations and have generated doubts that they are reliable enough or accurate enough.[78] Future developments will determine how effective AI/ML may be in PK/PD modeling.

D. AI in Clinical Development

The focus of the possible use of AI/ML, is, of course, on clinical development, considering the expenses[79] and the rates of failures in this area of pharmaceutical research.[1,80] We described in Section V.2 the many conceptual risks in the clinical development of drugs. These arise from various information gaps, such as an incomplete understanding of the pathophysiology of the disease, disease progression and penetration and by poor understanding of the mechanism of action of the test drug. AI/ML can help by accessing substantial amounts of information that many investigators have difficulty collecting considering the number of publications and databases in the field of biomedicine. In addition to that information, AI can provide extensive scientific information pertinent to the drug of interest that would have been challenging to obtain otherwise. It can examine and summarize various related clinical studies and provide pertinent real-world data (RWD).[81] The latter includes disease registries, health records, medical claims data, and publications of cases studies.

AI can predict the possible outcomes of clinical trials and assess risks, enhancing the possibilities of a successful development project

Beyond providing and summarizing existing information pertinent to the field of study, AI/ML can potentially have other uses. For example, it can assist in the effort to design better protocols for clinical studies. There are AI tools and routines that can predict the likelihood of a successful outcome of a given clinical study. They achieve this based on information available for the drug target, the results of the Phase 2 studies (for Phase 3 study predictions), or from previous clinical trials in the same or similar indications.[82,83,84] ASCO has recently issued a challenge for AI tools to predict the outcome of specific endpoints in oncology clinical trials.[85] Also, AI/ML can suggest specific biomarkers that influence the outcome. These can assist in patient stratification in certain clinical trials. Li et al. (2015) [86] developed and validated a model predicting response to the cancer drugs Erlotinib and Sorafenib using a tumor cell line panel to identify gene expression that was sensitive or resistant to Erlotinib treatment (using the median IC_{50} value as a cutoff). This is just one example of the use of omics[‡‡] to discover and validate biomarkers that can be used to predict outcomes both in the clinic and in clinical studies. A review of such studies by Glaab at al. (2021)[87] highlights their utility in such efforts, in utilizing the validated biomarkers for patient stratification in a variety of clinical studies. Using such tools can produce a protocol with a higher possibility of providing a successful outcome for the test compound, although the regulatory

[‡‡] Omics refers to large scale studies of biological molecules in an organism and it includes genomics, transcriptomics, proteomics, metabolomics, and epigenomics.

acceptance may not be particularly assured. Human judgment and experience remain crucial!

AI processes have also been used to predict possible study outcomes based on the selected design and existing data on the disease and pathophysiological processes.[88] AI/ML can also help identify risks in the designed studies and the overall clinical development program. The risks that AI/ML can identify include toxicity, efficacy and regulatory approval and it can do this by carefully evaluating the chemistry of the tested drug/biologic, preclinical data, the disease that is targeted, the genomics that are known regarding the disease, gene expression (transcriptomics) and proteomics known, as well as the molecular pathways and interactions that constitute the drug target. In addition to all these, RWD and previous studies are examined for clinical outcomes for predictions of toxicity, efficacy, and regulatory approval. [89]

AI/ML can assist in defining more accurately patient inclusion/exclusion criteria and in finding the patients that may benefit from the drug in development

Another potential area for AI/ML is selecting the best dose (or dose range), mainly in first-in-human clinical studies. AI, using predictive algorithms, prior information, and preclinical testing data (PK/PD), can estimate potential risks, drug concentrations at target, and pharmacodynamic effects across doses of the test drug. It can therefore help define the number of doses to administer in dose-escalating Phase 1 studies, speeding development, reducing costs, and minimizing the number of patients that need to be enrolled.[90] The FDA has recently decided to proceed with a pilot program to assess the use of AI-based technologies in improving the efficiency and speed of early-phase clinical trials (mostly first-in-human clinical studies). It stated that it will attempt to do so without compromising the scientific rigor and compliance with existing regulations. To support this pilot program, the FDA has posted a Request for Information (RFI) at Regulations.gov (https://www.regulations.gov/document/FDA-2026-N-4390-0001). Those interested may post comments, although the closing date for these was May 29, 2026.

Areas in which AI/ML have shown some promise is in assisting with patient selection by refining the enrollment criteria for the contemplated clinical studies. Patient recruitment and retention are critical issues in clinical development and the cause of early termination of at least one third % of Phase III studies.[91,92] AI/ML may help by expanding eligibility criteria; Liu et al. (2022)[93] developed Trial Pathfinder, a computational framework that integrates real-world data and systematically analyzes the hazard ratio for overall survival across cohorts defined by different eligibility criteria. Analysis using this framework showed that many exclusion criteria in oncology studies do not contribute substantially to the effectiveness and robustness of the study, and that removing them from protocols would double the number of eligible patients. Finding the right patients is crucial to demonstrating a drug effect. Verstovsek et al. (2024)[94] used and validated an ML approach to identify polycythemia vera patients at higher risk for thromboembolic complications after starting hydroxyurea.

Other efforts have been tentative and mostly theoretical so far. One of those has been study enrichment, an effort to select patients for a study whose test results and analyses indicate they are more likely to respond to the test compound than others.

Obviously, there are many regulatory issues with an approach in which an AI algorithm monitors and selects these patients. The FDA has identified certain issues here and has outsourced a project to a select group of researchers.[95] Various groups have provided AI platforms for enrichment.

Still, anticipated AI/ML breakthroughs should be weighed against meaningful limitations and risks. AI/ML is constrained by the datasets available to it—datasets that often contain gaps and may embed systematic bias. Moreover, such data may not be representative: (a) much of the evidence comes from clinical studies with heterogeneous eligibility criteria, and (b) real-world data (RWD) are frequently drawn from narrow, fragmented populations. In practice, AI/ML summaries may also implicitly treat RWD variables as independent variables and similarly distributed across settings, an assumption that is commonly incorrect. Finally, when the underlying pathophysiology is novel, the target has not been previously addressed, or the mechanism of action is new, existing data are unlikely to provide reliable guidance or outcome predictions. For these reasons, the challenges remain substantial, and a cautious approach is warranted.

E. AI routines in clinical studies and regulatory compliance

Therefore, AI/ML routines that may be included in any given clinical study protocol, such as those that may determine dose or study "enrichment", must be fully documented. Although it is not "fleshed out" in the FDA discussion paper, the EMA document specifies that the use of AI/ML in clinical study procedures (patient enrollment, treatment allocation, etc.) must conform to the ICH E6 Good Clinical Practice guideline.[5]

The utilization of AI in the review of submissions by regulatory agencies poses specific challenges to the sponsors to utilize similar tools to examine their documentation prior to submission

Such efforts can also be rolled over in Model-Informed Drug Development (MIDD) programs. As discussed previously, MIDD approaches integrate nonclinical experience, dose optimization, and disease progression models (DPMs) to better inform drug development, reduce the number of clinical studies required, and design studies to better provide key answers to the proposed hypothesis. Very recently, ICH produced a draft guideline (ICH M15) on the use of MIDDs in development, which includes a proposed outline for sponsor interactions with regulatory agencies in clinical studies informed by MIDDs and associated activities.[96] Considering the inputs for MIDD, AI/ML appears as an optimal synergistic partner. However, much depends on the data used and the team's experience.

F. AI in Regulatory Submissions Review

Although not part of this discussion on QbD, designers, planners, and analysts of clinical studies should become increasingly familiar with regulatory agencies' use of AI. For example, the FDA uses its own AI tool, Elsa, to assist employees with a variety of regulatory activities, such as reviewing protocols and submissions and examining adverse events. In addition, the FDA has recently consolidated disparate applications, submission data sources, and ports into a new platform called HALO. HALO and Elsa 4.0 can be used to query and analyze submitted data. The FDA's AI tools can scan documents to identify and highlight potential inconsistencies, so those submitting data

to the FDA must carefully screen their submission packages using similar tools and correct any issues identified this way. However, these tools carry inherent risks, and AI-to-AI interactions can introduce errors.[97]

VII. Discussion

How effective was the effort to introduce elements of QdB in clinical development? Likely marginal, with large pharma companies benefiting more

Regulatory agencies and the pharmaceutical industry have placed significant emphasis on quality by design (QbD) over the last decade and a half. How effective has this strategy been? Higher quality would mean a higher rate of successful clinical studies and, therefore, more new drugs and biologics approved. Certain data seem to support this. Smietana *et al.* (2016) examined success rates by phase of development and found that the rate of cumulative success (Phase 1 to Phase 3) declined substantially between 2000 and 2011 but registered a modest increase from 2012 to 2014. There has certainly been an increase in new drugs and biologics approved starting from 2012 onward (with the glaring exception of 2016).[98] The average number of drugs and biologics approved per year between 2000 and 2012 was 20.8 and 5.5, respectively. But between 2013 and 2022, the annual approval rate for NMEs and biologics increased to 31.4 and 11.7, respectively.[98] But was the effort to instill QbD in clinical studies responsible for this? Around the same time approvals of new pharmaceutical entities began to improve, the FDA started its "modernization" drive (FDASIA). A variety of regulatory paths were devised that sped up reviews and lowered approval requirements.[99] In certain years, the plurality of approvals was obtained through these new accelerated pathways.[100,101] Therefore, the relatively small increase in Phase 3 success rates reported by Dowden and Munro from 2015 to 2017 may reflect the new regulatory pathways rather than an increase in study quality.[102] Looking at FDA approvals in 2024 gives a good idea of how changes in regulatory procedures may have increased the number of approved medicinal entities. Last year (2024), the FDA approved 50 new pharmaceutical entities: 32 small molecules (drugs) and 18 biologics. Of these, 27 had received priority review designation, and 7 were approved in an accelerated process based on success in surrogate endpoints.[101] Therefore, a case can be made that the increase in approvals in the last decade was due mainly to procedural changes and novel regulatory pathways, rather than to the introduction of QbD principles. In fact, considering the substantial increase in the number of clinical studies in the last decade,[103] the failure rate does not appear to have diminished.

The causes of clinical trial failures remain a field of intense investigation. For example, Hwang et al. (2016) investigated 640 pivotal (Phase 3) studies of novel therapeutics and found that 54% of them failed. Of those that failed, 58% did so because they did not demonstrate adequate efficacy.[104] Efficacy failures are caused by inadequate design, poor endpoint selection, and underpowered studies. As we discussed previously, it is not uncommon to underestimate the control's efficacy, leading to underpowered pivotal studies. Unfortunately, pressure to obtain approval as quickly as possible due to

funding and investment pressures undermines key elements of QbD, with no time allotted for an in-depth investigation of all key parameters in the planned studies.

Some reasons for failures in the pivotal stage stem from flawed clinical development in earlier phases. To be fair, it may not always be the fault of those designing the studies. In certain therapeutic areas, such as oncology, typical Phase 2 protocols and endpoints do not provide enough information to design pivotal studies, as Retzios (2010) previously discussed.[105]

Large pharmaceutical companies are more likely to get approvals for experimental agents than small or medium-sized ones

However, QbD can overcome certain problems inherent in development in challenging therapeutic areas such as oncology. Kim et al. (2023) analyzed the success factors of clinical studies by examining data from approximately 24,000 clinical trials conducted between 2010 and 2020. The analysis showed that companies with substantial clinical research experience and several previous approvals were more likely to succeed in their development programs than companies that lacked this experience. This is further reinforced by the analysis of pivotal trial successes by Hwang et al.(2016)[104], who, after examining the success rate of late-stage clinical studies of investigational drugs from 2008 to 2015, found that experimental agents sponsored by small or mid-sized pharmaceutical companies have a lower likelihood of approval than those sponsored by large pharmaceutical companies. This study strongly emphasizes the need to build teams with development experience and expertise and to remove organizational and communication barriers to facilitate their work, as discussed in Section V.B.1 of this document. Pharmaceutical and biotechnology companies should build on the expertise they have developed, analyze failures, and identify organizational risk factors based on their accumulated experience. The analysis also showed that the presence of scientific collaborators improved success rates, strongly reinforcing the point that increasing scientific knowledge of the disease being addressed in development and the drug/biologic's mechanism of action is an essential element of quality. Yamaguchi et al. (2021)[106] further amplified these considerations by examining success rates by target. An interesting finding in that study was that success rates were lower in areas where the drug mechanism may not have been fully understood. We discussed earlier (Section V. 2) the negative effect of not fully understanding a drug's mechanism of action on success; this publication appears to confirm it.

Some pharmaceutical companies are better at this than others, although the record is not particularly clear. Schumacher et al. (2025)[107] analyzed the likelihood of approval (LoA) of new medicinal entities (NMEs) for 18 major pharmaceutical companies between 2022 and 2026. The average LoA was 14.3%, but it varied substantially among these companies, with the lowest being for AbbVie (8.1%) and the highest for Amgen (22.8%). There may be various reasons for this discrepancy among major pharmaceutical companies, but the results show that organizational factors are important to quality in pharmaceutical development. Unfortunately, no guidance can

address organizational matters, internal communication, interpersonal relationships, hiring competent people, etc.

Why is QbD difficult to introduce in clinical development? The main reason is that clinical studies have too many "moving parts." There are many actors, such as the sponsor's management teams, CROs, investigative sites, medical monitors, advisory boards, DSMs, and assorted experts, who participate in planning and conducting clinical trials. How well all these disparate characters mesh into planning and conducting a clinical study depends greatly on the sponsor's management team; in many cases, that team does not rise to the occasion. As mentioned in Section V.B.1, many companies simply lack the corporate culture and organization that allow the free exchange of information and ideas; training is uneven, decision-making is uncertain, and communication among study personnel is spotty. Imposing quality in that process requires determined individuals with excellent training and experience, and this is not always the case.

VIII. References

1 Smietana K, Statkowsky M, Moeller M: Trends in Clinical Success Rates. Nature Reviews: Drug Discovery, 2016; 15:379-380

2 Harrison, R. Phase II and phase III failures: 2013–2 015. Nat Rev Drug Discov 15, 817–818 (2016)

3 Morgan, P., Brown, D., Lennard, S. et al. Impact of a five-dimensional framework on R&D productivity at AstraZeneca. Nat Rev Drug Discov 17, 167–181 (2018)

4 Dowden, H., & Munro, J.: Trends in clinical success rates and therapeutic focus. Nat Rev Drug Discov. 2019 Jul;18(7):495-496.

5 ICH E6(R2) and Integrated Addendum – 9 November 2016: Accessed at https://database.ich.org/sites/default/files/E6_R2_Addendum.pdf.

6 FDA-EMA: Pilot program for parallel assessment of Quality-by-Design applications: lessons learnt and Q&A resulting from the first parallel assessment. August 2013; Accessed at extension://efaidnbmnnnibpcajpcglclefindmkaj/https://www.ema.europa.eu/fr/media/13532

7 FDA Guidance to Industry: Q8, Q9, and Q10 Questions and Answers- Appendix Q&As from Training Sessions; July 2012., Accessed at https://www.fda.gov/media/83904/download

8 FDA Guidance to Industry: Q8, Q9, and Q10 Questions and Answers (R5). May 2026 Accessed at https://www.fda.gov/media/78668/download

9 Retzios AD: Assembling and Utilizing Target Product Profiles: In "Global Pharmaceutical and Biologics Regulatory Strategy", Sietsema WK, Meacham MM, eds, Regulatory Affairs Professional Society, 2020

10 Tyndal A, Du W, Breder CD: The target product profile as a tool for regulatory communication: advantageous but underused. Nature 2019; 16: 156

11 Singer M, Deutschman CS, Warren Seymour C, et al. The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). JAMA 2016; 315: 801-810

12 Matthay MA, Yaseen A, Arroliga AC, et al. A New Global Definition of Acute Respiratory Distress Syndrome. Am J Respir Crit Care Med 2024; 209: 37-47

[13] Retzios AD: Why pivotal trials fail:
[14] Singer M, Deutschman CS, Seymour CW, et al.: The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). JAMA. 2016; 315: 801–810
[15] Ranieri M, Rubenfeld GD, Thompson BT, et al.: Acute Respiratory Distress Syndrome -The Berlin Definition. JAMA. 2012;307:2526-2533
[16] Honkala A, Malhotra SV, Kummar S, et al.: Harnessing the predictive power of preclinical models for oncology drug development. Nature Reviews Drug Discovery 2022; 21: 99–114
[17] Liu J, Barrett JS, Leonardi ET, et al.: Natural History and Real-World Data in Rare Diseases: Applications, Limitations, and Future Perspectives. Clin Pharmacol 2022, 62(S2): S38–S55
[18] FDA Guidance to Industry: Rare Diseases: Natural History Studies for Drug Development. March 2019. Accessed at: https://www.fda.gov/media/122425/download
[19] Barret JS, Nicholas T, Azer K, et al.: Role of Disease Progression Models in Drug Development. Pharm Res 2022; 39: 1803-1815
[20] Cook SF, Bies RR. Disease Progression Modeling: Key Concepts and Recent Developments. Curr Pharmacol Rep 2016; 2:221–230
[21] Bime C, Casnova N, Oita RC, et al.: Development of a biomarker mortality risk model in acute respiratory distress syndrome. Critical Care 2019; 23:410
[22] Lau Raket L: Statistical Disease Progression Modeling in Alzheimer Disease. Front. Big Data, 2020; 3:24
[23] FDA Guidance to Industry: Real-World Data: Assessing Registries to Support Regulatory Decision-Making for Drug and Biological Products Guidance for Industry. November 2021. Accessed at https://www.fda.gov/media/154449/download
[24] FDA Guidance to Industry: Considerations for the Use of Real-World Data and Real-World Evidence to Support Regulatory Decision-Making for Drug and Biological Products. December 2021. Accessed at https://www.fda.gov/media/154714/download
[25] Overington JP, Al-Lazikani B, Hopkins AL. How many drug targets are there? Nature Reviews Drug Discovery 2006; 5:993-996
[26] Gregori-Puigiane E, Setola V, Hert J, et al.: Identifying mechanism-of-action targets for drugs and probes. PNAS 2012; 109, no 28
[27] Lindsay MA: Target Discovery. Nature Reviews Drug Discovery 2003; 2:831-838
[28] Moffat JG, Vincent F, Lee JA, et al. Opportunities and challenges in phenotypic drug discovery: an industry perspective. Nature Reviews Drug Discovery 2017; 16:531-543
[29] Bezprozvanny I: The rise and fall of Dimebon. Drug News Perspect 2010; 23: 518-523
[30] Liu J, Chang L, Song Y, et al.: The role of NMDA receptors in Alzheimer's disease. Frontiers in Neuroscience 2019; 13: article 43
[31] Nature Reviews Drug Development Editorial: Learning lessons from Pfizer's $800 million failure. Nat Rev Drug Discov, 2011; 10:163–164
[32] FDA Table of Surrogate Endpoints That Were the Basis of Drug Approval or Licensure. Accessed at: https://www.fda.gov/drugs/development-resources/table-surrogate-endpoints-were-basis-drug-approval-or-licensure
[33] FDA: Considerations for Discussion of a New Surrogate Endpoint(s) at a Type C PDUFA Meeting Request. Accessed at extension://efaidnbmnnnibpcajpcglclefindmkaj/https://www.fda.gov/media/115120/download
[34] FDA Guidance to Industry: Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics. December 2018. Accessed at https://www.fda.gov/regulatory-information/search-fda-guidance-documents/clinical-trial-endpoints-approval-cancer-drugs-and-biologics

[35] FDA-NIH: BEST (Biomarkers, EndpointS, and other Tools) Resource. November 2021. Accessed at https://www.ncbi.nlm.nih.gov/books/NBK326791/

[36] EMA: Clinical efficacy and safety guidelines. Accessed at https://www.ema.europa.eu/en/human-regulatory-overview/research-and-development/scientific-guidelines/clinical-efficacy-and-safety-guidelines

[37] EUnetTHA: EUnetHTA 21 – Individual Practical Guideline Document - D4.4 – OUTCOMES (ENDPOINTS). Version 1.0, 25/01/2023. Accessed at extension://efaidnbmnnnibpcajpcglclefindmkaj/https://www.eunethta.eu/wp-content/uploads/2023/01/EUnetHTA-21-D4.4-practical-guideline-on-Endpoints-v1.0.pdf

[38] Anker SD, Schroeder S, Atar D, et al.: Traditional and new composite endpoints in heart failure clinical trials: facilitating comprehensive efficacy assessments and improving trial efficiency. European Journal of Heart Failure 2016; 18:482–489

[39] McCoy E: Understanding the Use of Composite Endpoints in Clinical Trials. Western Journal of Emergency Medicine 2018; 19:631-634

[40] Niewoehner DE, Erbland ML, Deupree RH, et al. Effect of systemic glucocorticoids on exacerbations of chronic obstructive pulmonary disease. Department of Veterans Affairs Cooperative Study Group. NEJM. 1999; 340:1941-1947.

[41] Silvers D, Kipnes M, Broadstore V, et al. Domperidone in the Management of Symptoms of Diabetic Gastroparesis: Efficacy, Tolerability, and Quality-of-Life Outcomes in a Multicenter Controlled Trial. Clinical Therapeutics 1998; 20:438-453

[42] Patterson D, Abel T, Rothstein R, et al. A Double-Blind Multicenter Comparison of Domperidone and Metoclopramide in the Treatment of Diabetic Patients With Symptoms of Gastroparesis. Am J Gastroenterology 1999; 94:1230-1235

[43] FDA Guidance to Industry: Multiple Endpoints in Clinical Trials. October 2022. Accessed at https://www.fda.gov/regulatory-information/search-fda-guidance-documents/multiple-endpoints-clinical-trials-guidance-industry

[44] EMA: Guideline on multiplicity issues in clinical trials. June 2017. Accessed at: extension://efaidnbmnnnibpcajpcglclefindmkaj/https://www.ema.europa.eu/en/documents/scientific-guideline/draft-guideline-multiplicity-issues-clinical-trials_en.pdf

[45] Bland JM, Altman DG. Multiple significance tests: the Bonferroni method. BMJ 1995; 310:170

[46] Holm M. A simple sequentially rejective multiple test procedure. Scand J Statist 1979;6:65-70

[47] Hochberg Y. A sharper Bonferroni procedure for multiple tests of significance. Biometrika 1988;75:800-802.

[48] Hommel G. A stagewise rejective multiple test procedure based on a modified Bonferroni test. Biometrika 1988;75:383-386.

[49] Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Series B Stat Methodol 1995;57:289-300

[50] Pugh SL, Torres-Saavedra PA. Fundamental Statistical Concepts in Clinical Trials and Diagnostic Testing. J Nucl Med 2021; 62: 757–764

[51] Opal SM, Laterne P-F, Francois B, et al.: Effect of Eritoran, an Antagonist of MD2-TLR4, on Mortality in Patients With Severe Sepsis.-The ACCESS Randomized Trial. JAMA 2013; 309: 1154-1162

[52] Kaukonen K-M, Bailey M, Suzuki S, et al.: Mortality Related to Severe Sepsis and Septic Shock among Critically Ill Patients in Australia and New Zealand, 2000-2012. JAMA 2014;311:1308-1316.

[53] Levin M, Quint PA, Goldstein B, et al.: Recombinant bactericidal/permeability-increasing protein (rBPI21) as adjunctive treatment for children with severe meningococcal sepsis: a randomized trial. The Lancet, 2000; 356:961-967

[54] Retzios AD: Innovation in Drug Development: Adaptive Designs for Clinical Trials. Accessed at https://adrclinresearch.com/adaptive%20designs.htm

[55] Pallmann P, Bedding AW, Choodari-Oskooei B, et al.: Adaptive designs in clinical trials: why usethem, and how to run and report them. BMC Medicine 2018; 16:29

[56] FDA: E9(R1) Statistical Principles for Clinical Trials: Addendum: Estimands and Sensitivity Analysis in Clinical Trials. June 2017. Accessed at: https://www.fda.gov/media/108698/download

[57] ICH E3 Structure and Content of Clinical Study Reports – Questions and Answers (R1). January 2013. Accessed at: https://www.ema.europa.eu/en/ich-e3-structure-content-clinical-study-reports-scientific-guideline

[58] ICH Integrated addendum to ICH E6 (R1)-Guideline for Good Clinical Practice E6 (R2). Accessed at: extension://efaidnbmnnnibpcajpcglclefindmkaj/https://database.ich.org/sites/default/files/E6_R2_Addendum.pdf

[59] FDA Guideline for the Monitoring of Clinical Investigations; January 1988

[60] FDA Guidance to Industry: Oversight of Clinical Investigations — A Risk-Based Approach to Monitoring. August 2013. Accessed at: https://www.fda.gov/media/116754/download

[61] FDA Guidance to Industry: A Risk-Based Approach to Monitoring of Clinical Investigations Questions and Answers. April 2023. Accessed at: https://www.fda.gov/media/121479/download

[62] Andersen JR, Byrljalsen I, Bihlet A, et al. Impact of source data verification on data quality in clinical trials: an empirical post hoc analysis of three phase 3 randomized clinical trials. Br J Clin Pharmacol 2014; 79:660-668

[63] Klatte K, Pauli-Magnus C, Love SB, et al. Monitoring strategies for clinical intervention studies (Review). Cochrane Database of Systematic Reviews 2021, Issue 12. Art. No.: MR000051

[64] Ellenberg SS, Shaw PA: Early Termination of Clinical Trials for Futility — Considerations for a Data and Safety Monitoring Board. NEJM Evid 2022; 1 (7)

[65] Zhang E, DuBois SG: Early Termination of Oncology Clinical Trials in the United States. Cancer Medicine. 2023; 12:5517–5525.

[66] Goldberg A, Bakhireva LN, Page K, et al. A Qualitative Scoping Review of Early-Terminated Clinical Trials Sponsored by the Department of Veterans Affairs Cooperative Studies Program From 2010 to 2020. Epidemiol Rev. 2022; 44:110–120

[67] Connors JM, Brooks MM, Sciurba FC, et al. Effect of antithrombotic therapy on clinical outcomes in outpatients with clinically stable symptomatic COVID-19: the ACTIV-4B randomized clinical trial. JAMA 2021; 326:1703-171

[68] Jennison C, Turnbull BW. Group sequential methods with applications to clinical trials. Chapman & Hall/CRC, Boca Raton, 1999

[69] Bernard GR, Vincent J-R, Laterre P-F, et al.: Efficacy and Safety of Recombinant Human Activated Protein C. N Engl J Med 2001;344:699-709

[70] Abraham E, Laterre P-F, Garg R, et al.: Drotrecogin Alfa (Activated) for Adults with Severe Sepsis and a Low Risk of Death. N Engl J Med 2005; 353:1332-1341

[71] FDA: Artificial Intelligence & Medical Products: How CDER, CBER, CDRH and OCP are working together. March 2024 (Revised February 2025)

[72] FDA: Using Artificial Intelligence and Machine Learning in the Development of Drug and Biological Products – Discussion Paper and Request for Feedback. May 2023 (Revised February 2025)

[73] EMA: Reflection paper on the use of Artificial Intelligence (AI) in the medicinal product lifecycle. September 2024

[74] EMA: Review of AI/ML applications in medicines lifecycle (2024); Horizon Scanning Short Report. July 2025

[75] Zuchin PJH, Mukherjee S, Wu JC: FDA Modernization Act 2.0: transitioning beyond animal models with human cells, organoids, and AI/ML-based approaches. J Clin Invest 2023; 133: e175824.

[76] FDA: Roadmap to reducing animal testing in preclinical safety studies: Accessed at https://www.fda.gov/files/newsroom/published/roadmap_to_reducing_animal_testing_in_preclinical_safety_studies.pdf

[77] Ghosh A, Choudhary G, Mehti D: The pivotal role of artificial intelligence in enhancing experimental animal model research: A machine learning perspective. Indian Journal of Pharmacology, 2024, Volume 56, Issue 1

[78] Huang Z, Denti P, Mistry H, et al. Machine learning and Artificial Intelligence in PK-PD modeling: Fad, Friend or Foe.

[79] DiMasi JA, Grabowski HG, Hansen RW. Innovation in the pharmaceutical industry: new estimates of R&D costs. J Health Econ. 2016;47:20–33.

[80] Wong CH, Siah KW, Lo AW. Estimation of clinical trial success rates and related parameters. Biostatistics. 2019;20:273–286.

[81] Chen Z, Liu X, Hogan W, et al.: Applications of artificial intelligence in drug development using real-world data. Drug Discov Today 2021; 26: 1256–1264

[82] Artemov AV., Putin E., Vanhaelen Q., et al.:. Integrated deep learned transcriptomic and structure-based predictor of clinical trials outcomes. bioRxiv DOI:10.1101/095653

[83] Feijoo F., Palopoli M., Bernstein J, et al.:. Key indicators of phase transition for clinical trials through machine learning. Drug Discov 2020; 25:414–421

[84] Qi Y, Tank Q: Predicting Phase 3 Clinical Trial Results by Modeling Phase 2 Clinical Trial Subject Level Data Using Deep Learning. Proceedings of Machine Learning Research 2019;106:1-14

[85] Vinogradov V and Vinogradova A. ASCO Clinical Trial Outcome prediction. https://kaggle.com/competitions/asco-clinical-trial-outcome-prediction, 2025. Kaggle

[86] Li B, Shin H, Gulbekyan G, et al.: Development of a drug-response modeling framework to identify cell line derived translational biomarkers that can predict treatment outcome to erlotinib or sorafenib. PLoS ONE. 2015;10: e0130700.

[87] Glaab E, Rauschenberger A, Banzi R, et al.: Biomarker discovery studies for patient stratification using machine learning analysis of omics data: a scoping review. BMJ Open 2021, 11: e053674.

[88] Qian L, Lu X, Haris P, et al.: Enhancing clinical trial outcome prediction with artificial intelligence: a systematic review. Drug Discovery Today 2025: 30; 4

[89] Badwan BA, Liaropoulos G, Kyrodimos E, et al.: Machine learning approaches to predict drug efficacy and toxicity in oncology, Cell Reports Methods 2023;3:100413

[90] Malheiro V, Santos B, Figueiras A, et al.:The Potential of Artificial Intelligence in Pharmaceutical Innovation: From Drug Discovery to Clinical Trials. Pharmaceuticals (Basel). 2025;18:788.

[91] Desai M: Recruitment and retention of participants in clinical studies: Critical issues and challenges. Perspect Clin Res. 2020; 11:51–53

[92] Kitterman DR, Cheng SK, Dilts DM et al.: The Prevalence and Economic Impact of Low-Enrolling Clinical Studies at an Academic Medical Center. Acad Med. 2011; 86:1360–1366

[93] Liu R, Rizzo S, Whipple S, et al. Evaluating eligibility criteria of oncology trials using real-world data and AI. Nature. 2021 ; 592: 629–633

[94] Verstovsek S, Krecak I, Heidel FH, et al.: Identifying Patients with Polycythemia Vera at Risk of Thrombosis after Hydroxyurea Initiation: The Polycythemia Vera—Advanced Integrated Models (PV-AIM) Project. Biomedicines 2023; 11, 1925

[95] FDA: Evaluate Application of Artificial Intelligence to Adaptive Enrichment Clinical Trials. October 2021: Accessed at https://www.fda.gov/science-research/advancing-regulatory-science/evaluate-application-artificial-intelligence-adaptive-enrichment-clinical-trials

[96] ICH: General Principles for Model-Informed Drug Development – M15: January 2026. Accessed at: https://database.ich.org/sites/default/files/ICH_M15_Step4_Final_Guideline_2026_0129.pdf

[97] Athni TS: US Food and Drug Administration Shifts to AI-Enhanced Regulatory Review With Elsa 4.0 and HALO. J Med Internet Res 2026; 28: e101884

[98] Seoane-Vasquez E, Rodriguez-Monguio R, Powers JH: Analysis of US Food and Drug Administration new drug and biologic approvals, regulatory pathways, and review times, 1980–2022. Scientific Reports 2024; 14:3325

[99] FDA Guidance to Industry: Expedited Programs for Serious Conditions – Drugs and Biologics – May 2014. Accessed at: extension://efaidnbmnnnibpcajpcglclefindmkaj/https://www.fda.gov/media/86377/download

[100] Ferries EA, Fleming WK, Shrank WH: FDA expedited approval and implications for rational formulary and health plan design. J Manag Care Spec Pharm 2021; 27:682-684

[101] Mullard A: 2024 FDA Approvals. Nature Reviews Drug Discovery 2025; 24: 75-82

[102] Dowden H, Munro J: Trends in Clinical Success Rates and Therapeutic Focus. Nature Reviews Drug Discovery, 2019; 18:495-496

[103] Gresham G, Meinert JL, Gresham AG et al.: Update on the clinical trial landscape: analysis of ClinicalTrials.gov registration data, 2000–2020. Trials 2022; 23:858

[104] Hwang TJ, Carpenter D, Lauffenburger JC, et al. Failure of investigational drugs in late-stage clinical development and publication of trial results, JAMA Intern. Med 2016; 176 1826–1833.

[105] Retzios AD: Why Do So Many Phase 3 Clinical Trials Fail? Part 1: The Effect of Deficient Phase 2 Trials in Therapeutic Areas with High Failure Rates in Phase 3 Studies. 2010; Accessed at https://adrclinresearch.com/index_htm_files/Why%20Pivotal%20Clinical%20Trials%20Fail%20-%20Part%201_v12L_a.pdf

[106] Yamaguchi S, Kameko M, Narukawa M: Approval Success Rates of Drug Candidates Based on Target, Action, Modality, Application, and their combination. Clin Trans Sci 2021; 14:1113-1122

[107] Schuhmacher A, Hinder M, Brief E, et al.: Benchmarking R&D success rates of leading pharmaceutical companies: an empirical analysis of FDA approvals (2006–2022). Drug Discovery Today 2025; 30:Number 2