Essential DMPK Strategies for IND Success

Ninety per cent of drug candidates that enter clinical trials never reach patients. Developing a single new medicine now costs more than a billion dollars. Too often, the culprit is a molecule whose absorption, clearance, drug–drug interactions or translation from animals to humans was never fully understood. The reality is that many of these liabilities can be identified, managed and potentially prevented when modality-based drug metabolism and pharmacokinetics (DMPK) are integrated into discovery from the outset.

The market reflects this growing need. Bioanalysis and DMPK studies continue to be one of the fastest growing segments of the preclinical contract research organisation (CRO) landscape, driven by the increasing complexity of modern therapeutic modalities and growth in biologics, antibody–drug conjugates (ADCs), oligonucleotides and gene therapies. Yet for many drug discovery researchers, particularly smaller biotechnology companies without dedicated DMPK expertise, the practical challenge is not simply accessing individual assays. It is determining which studies are needed, when they should be performed, how they should be integrated and how the resulting data should inform development decisions.

DMPK is no longer a downstream box to check, nor is it confined to supporting medicinal chemistry. The real challenge today is the increasing diversity of therapeutic modalities and the complexity of next generation agents such as ADCs, multi-specific biologics, T-cell engagers, protein degraders, molecular glues, gene therapies, peptides and oligonucleotides. Each presents distinct challenges in absorption, distribution, metabolism, clearance, bioanalysis and translation. Effective DMPK strategies therefore need to be integrated from the earliest stages of discovery, using mechanistic data to guide design, identify liabilities, establish exposure–response relationships and determine whether a promising molecule can become a viable therapeutic candidate.

Why DMPK Must Function as a Cohesive System

Many DMPK failures stem from relying on outdated assumptions or focusing too narrowly on potency. Compounds that perform well *in vitro* may fail *in vivo* due to poor permeability, rapid clearance or unexpected metabolites. Other common issues include missing early DDI risks, relying on animal models that do not reflect human metabolism or assuming plasma PK will predict tissue exposure, leading to weak translational confidence.

The challenge becomes even greater with complex modalities like ADCs, degraders, peptides and oligonucleotides. Standard small molecule DMPK cascades simply don't capture their unique distribution, clearance or biotransformation pathways. When PK, PD and biomarker strategies aren't tightly integrated, teams struggle to link exposure to biological effect or to select rational doses. Together, these failure modes highlight the need for modality appropriate assays, earlier mechanistic insight and a more holistic approach to DMPK across discovery and development.

Modern DMPK is centred on understanding exposure–response relationships and building translational PK/PD models that can reliably anticipate human outcomes. Effective DMPK today depends on generating the right mechanistic data early enough to support quantitative modelling, confident dose selection and a clear line of sight from preclinical studies to clinical performance.

Rather than functioning as a checklist of isolated experiments, DMPK should be organised as a sequence of connected decision gates that actively guide progression from early discovery through the Investigational New Drug (IND) filing. Each stage is designed to answer a specific, development-critical question, enabling teams to identify risks early, prioritise leads and make informed go/no-go decisions before costs escalate. Figure 1 below outlines these stages and highlights how targeted DMPK studies support decision-making at each point in the development lifecycle.

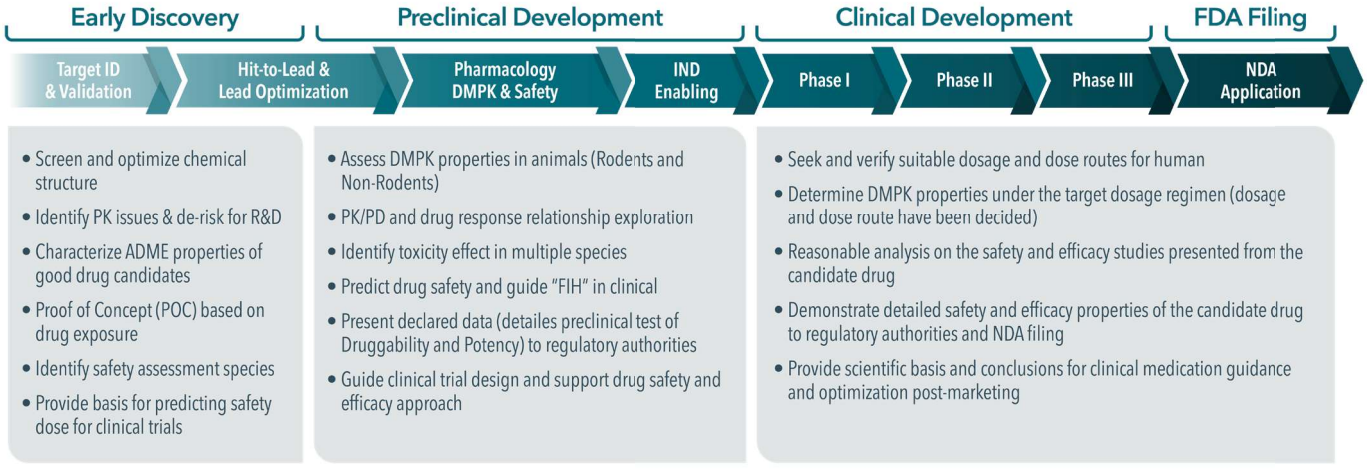


Figure 1. Stages of DMPK from Early Drug Discovery to FDA Filing

Figure courtesy of Crown Bioscience

Crucially, these stages are not independent, but interconnected. Data generated during lead optimisation, such as metabolite identification, should directly inform DDI strategies at candidate selection. Likewise, physiologically based pharmacokinetic (PBPK) models developed ahead of IND submission should provide a foundation for Phase I dose selection. Fragmented execution, where studies are conducted in isolation, by different vendors and without an integrated interpretive framework, can create gaps in understanding and delay critical decisions. Early DMPK integration transforms raw data into a clear roadmap, thereby reducing programme risk and accelerating the transition from bench to clinic.[1]

Modality Specific DMPK Considerations

A stage based DMPK framework remains important, but the most informative studies differ substantially by therapeutic modality. Taking a modality specific view helps clarify how DMPK priorities shift across classes of molecules.

For small molecules, early work typically centres on core *in vitro* ADME assays (such as solubility, lipophilicity, metabolic stability, permeability and protein binding) followed by IV and PO PK studies to understand clearance, half life, exposure and bioavailability. As programmes advance, tumour-bearing mouse PK/PD, tissue distribution, metabolite identification, transporter assessments and IVIVE/PBPK modelling help strengthen translational confidence.

For PROTACs and other targeted protein degraders, early solubility, lipophilicity, stability and permeability assays remain useful, though classical permeability and clearance models are often less predictive. Early *in vivo* PK and tissue exposure studies are especially important to determine whether adequate systemic or target site exposure is achievable. Later translational work may include target degradation studies, PK/PD, efficacy models, tissue distribution and tailored modelling approaches.

ADCs require early bioanalytical method development, measurement of key analytes (total antibody, conjugated antibody and free payload) and stability assessments across matrices to characterise linker stability, drug-to-antibody ratio (DAR), payload release, tissue distribution and species differences in target binding and disposition, all of which require consideration. For next-generation ADCs, including bispecific and dual-payload formats, DMPK becomes even more complex because different targets, payloads and release mechanisms may contribute independently to exposure, efficacy and toxicity. Humanised mouse models can clarify PK when species differences matter, and tissue distribution, efficacy studies and monkey PK/toxicology support later stage translation.

For bispecific antibodies and T-cell engagers, DMPK must account for target-mediated drug disposition, receptor occupancy, immune-cell interactions and nonlinear PK. Plasma exposure alone may be insufficient to predict pharmacological activity. Integrated PK/PD studies incorporating target occupancy, T-cell activation, cytokine release, tumour exposure and immune biomarkers can provide a more meaningful translational framework. Human-relevant immune models and mechanistic modelling may be particularly important where species differences limit conventional animal studies.

For oligonucleotides and siRNA therapeutics, stability, fit for purpose bioanalysis and tissue distribution are central, with PK/PD relationships, persistence and target tissue exposure guiding development. Repeat dose behaviour and higher species studies often become necessary as programmes mature.

Peptides and cyclic peptides typically require early plasma and enzymatic stability assessments, route specific PK evaluation and consideration of formulation or half life extension strategies to achieve viable exposure profiles.

Gene therapy products demand a focus on biodistribution, vector analytics and transgene expression, with shedding and persistence informing safety and translational interpretation; immunogenicity and higher species toxicology become increasingly important later in development.

Taken together, a modality specific lens enhances the traditional stage based DMPK framework by helping teams

prioritise the studies that most effectively improve translational confidence and support sound development decisions.

Bridging Preclinical to Clinical: PBPK, IVIVE and Biomarkers

For decades, DMPK science was optimised primarily around small molecules. Lipinski's Rule of Five served as a practical framework for oral drug likeness; hepatic CYP450 metabolism dominated clearance pathways, as well as a well established toolkit of microsomal stability assays, Caco 2 permeability studies and rodent PK could reliably characterise most development candidates. That entire framework has now fundamentally evolved for the increasingly diverse therapeutic landscape. ADCs, PROTACs, oligonucleotides, siRNA and cyclic peptides occupy a rapidly growing share of modern pipelines and each challenges at least one foundational assumption of classical DMPK. As a result, modality specific DMPK strategies have become essential rather than optional.

PBPK provides a mechanistic framework for integrating physicochemical properties, *in vitro* ADME data, tissue physiology and species-specific parameters to predict human exposure. Regulatory agencies now actively encourage PBPK submissions and, in some cases, accept PBPK analyses in lieu of certain clinical pharmacology studies, underscoring their growing role in de risking early development. IVIVE provides a simpler but complementary path, scaling hepatocyte or microsomal intrinsic clearance to predict *in vivo* hepatic clearance. However, its well documented tendency to underestimate clearance for low turnover compounds (<10 mL/min/kg) highlights the need for enhanced methodologies such as hepatocyte relay assays and long term co culture systems.

Recent work on heterobifunctional degraders reinforces this point. PROTACs, which sit far outside traditional rule of five space, challenge standard permeability and clearance assays.[2] Caco 2 assays lack predictive value for absorption, and conventional IVIVE systematically underpredicts clearance unless experimentally determined fraction of unbound drug fu inc values are used.

Computational DMPK approaches are also increasingly valuable for prioritising compounds and identifying liabilities earlier in discovery. However, predictions of pKa, LogD, solubility, permeability, metabolic stability and protein binding remain highly dependent on chemical space and model quality. *In silico* approaches are therefore most powerful when used to prioritise experiments rather than replace high-quality experimental data.[3]

The need for robust translational frameworks extends beyond small molecules. Gene therapy programmes face their own DMPK challenges, including biodistribution, shedding, immunogenicity and dose projection. For example, recombinant adeno-associated virus (rAAV) modalities require specialised bioanalytical platforms and PK/PD thinking that accounts for vector biology rather than classical clearance pathways.[4] These considerations are now central to FIH dose selection and risk mitigation in an increasingly crowded gene therapy landscape.

Across modalities, translational biomarkers provide an essential link between exposure and pharmacology. Establishing a preclinical link between drug exposure and a measurable biomarker response can enable the same pharmacodynamic endpoint to be monitored during early clinical development, helping to confirm target engagement and pharmacological activity at predicted exposure levels. This reduces the risk of entering Phase II with an incorrect dose or an unvalidated mechanism, complementing PBPK, IVIVE and modality-specific DMPK strategies to create a more resilient path from discovery to clinic.

Regulatory Expectations of an IND Package

Regulatory expectations for DMPK and clinical pharmacology are defined by the therapeutic modality, development stage and specific characteristics of the candidate. For a standard IND, the core DMPK package requires metabolic stability and clearance data, metabolite identification and reactive metabolite assessment, P450 phenotyping, CYP inhibition and induction screening, transporter interaction studies (P-gp, BCRP, OATP1B1/1B3, OAT1/3, OCT2), plasma protein binding and blood-to-plasma ratio, and *in vivo* PK in two species.

DDI liability deserves particular attention. If a compound is a strong CYP3A4 inhibitor, a time-dependent inhibitor, or heavily dependent on a single elimination pathway, a clinical development plan will be materially constrained by DDI risk. The FDA guidance recommends a tiered approach: *in vitro* screens, then static mechanistic modelling, then PBPK-based clinical prediction. Characterising this risk at candidate selection, not during IND-enabling studies, offers time to modify the scaffold or design the clinical programme intelligently.

The appropriate DMPK package depends on the molecule, modality, indication, route of administration, intended duration of treatment and known pharmacological and toxicological risks. Early regulatory interaction can therefore be valuable when novel modalities or unconventional DMPK strategies are being developed.

Choosing a DMPK CRO

In practice, few principles are more useful than broad service descriptions. For novel modalities, depth of experience in the specific modality is more important than the breadth of a CRO's offerings. Evaluation should focus on study director qualifications, publication history and client references rather than service lists.

The quality of scientific engagement also differentiates providers. Some CROs primarily execute predefined studies, while others contribute to protocol design, highlight unexpected findings and participate in interpretation across assays. The latter approach tends to add more value during development. Continuity of partnership can further improve outcomes, as knowledge accumulated on a compound is not easily transferred between vendors.

Desirable DMPK support includes *in vivo* pharmacokinetic studies across species and dosing routes; bioanalytical measurement of parent compounds and metabolites using methods such as LC–MS/MS, HPLC and immunoassays and supporting activities like method development and validation, dosing formulation analysis, tissue distribution, excretion studies and maximum tolerated dose determination.

Conclusion

Despite major advances in target biology, medicinal chemistry and therapeutic innovation, clinical attrition rates remain high.

In many cases, failure is not driven solely by inadequate biology, but by insufficient understanding of exposure, clearance, distribution, metabolism or the relationship between drug exposure and pharmacological activity. DMPK therefore plays a far more consequential role than simply supporting regulatory submission; it is a core translational discipline that directly shapes the probability of clinical success.

As therapeutic modalities become increasingly diverse, spanning small molecules, ADCs, PROTACs, oligonucleotides, gene therapies and cell-based approaches, traditional DMPK paradigms cannot be applied uniformly. The most effective strategies are increasingly modality-specific, mechanistic and translational, incorporating tissue and intracellular exposure, bioanalysis, biomarkers, PK/PD modelling, IVIVE, PBPK and human-relevant models where appropriate. The capabilities to improve outcomes already exist and are accessible to both small and large organisations. The key shift is therefore from asking "What DMPK studies are required?" to asking "What development decision must the DMPK programme enable us to make?" Early in discovery, that may mean identifying a chemical series with acceptable developability. During lead optimisation, it may mean balancing exposure against potency and toxicity. At candidate selection, it means demonstrating a credible human PK and safety profile. Before IND submission, it means integrating the available evidence into a defensible translational and clinical strategy.

For biotechnology companies in particular, access to the right expertise and capabilities can be as important as access to individual assays. Selecting partners with demonstrated modality-specific experience, strong bioanalytical capabilities and the ability to integrate DMPK with pharmacology, biomarkers and modelling can reduce uncertainty and accelerate critical development decisions.

In an increasingly complex drug development landscape, DMPK should no longer be viewed as a downstream regulatory requirement. When integrated early and used strategically, it becomes a powerful translational framework for identifying risk, optimising candidates, guiding clinical dose selection and ultimately increasing the probability that promising therapies reach patients

REFERENCES

1. Crown Bioscience. Beyond exposure: Integrating PK/PD for translational predictive power [Webinar]. Available from: https://www.crownbio.com/webinar-beyond-exposure-integrating-pk-pd-for-translational-predictive-power
2. Maurer CK, Fang Z, Schindler C, Pohl G, Machrouhi-Porcher F, Lecomte M, Petersson C, Duevel HM. *In vitro* and *in vivo* ADME of heterobifunctional degraders: a tailored approach to optimize DMPK properties of PROTACs©. RSC Med Chem. 2025 Feb 21;16(4):1746-1757. doi: 10.1039/d4md00854e. PMID: 39990161; PMCID: PMC11843253.
3. Handa K, Hirano M, Kageyama M, Bender A. Computational approaches to DMPK: A realistic assessment of current methods and their practical impact. Part I: Physicochemical and *in vitro* properties. Drug Discov Today. 2025 Aug;30(8):104422. doi: 10.1016/j.drudis.2025.104422. Epub 2025 Jun 30. PMID: 40602659.
4. Chen N, Sun K, Chemuturi NV, Cho H, Xia CQ. The Perspective of DMPK on Recombinant Adeno-Associated Virus-Based Gene Therapy: Past Learning, Current Support, and Future Contribution. AAPS J. 2022 Jan 31;24(1):31. doi: 10.1208/s12248-021-00678-7. PMID: 35102450; PMCID: PMC8817103.
5. Martins V, Fazal L, Oganesian A, Shah A, Stow J, Walton H, Wilsher N. A commentary on the use of pharmacoenhancers in the pharmaceutical industry and the implication for DMPK drug discovery strategies. Xenobiotica. 2022 Aug;52(8):786-796. doi: 10.1080/00498254.2022.2130838. PMID: 36537234.

Rajendra Kumari

As Executive Director of Integrated Solutions, Rajendra leads the focus of cross-disciplinary solutions for drug discovery and development. She has over 20 years of oncology research experience with a PhD in Molecular Pharmacology from the University of Leicester. Before joining Crown Bioscience, Rajendra held an Assistant Professor position in the Division of Preclinical Oncology within the Medical School, University of Nottingham and co-founded PRECOS Ltd., serving as COO and then, following the merger with Crown Bioscience, as CSO and General Manager for Crown Bioscience UK. Rajendra then became the Global Head of Scientific Communications for all platforms at Crown Bioscience Inc., working closely with R&D, multiple divisions and global departments to support the commercialisation of R&D activity and capabilities.

Dr. Xiaolong Tu

Dr. Xiaolong Tu earned his PhD from Nanjing University School of Life Sciences in 2016. He now leads the DMPK Tox Platform at Crown Bioscience, overseeing integrated DMPK and bioanalytical services. With over a decade of experience in preclinical oncology drug development, he has managed a wide range of programmes including ADCs, biologics and novel modalities. His expertise encompasses nonclinical PK/PD, toxicology and translational oncology. Dr Tu is dedicated to delivering high quality, regulatory ready data that bridge discovery and clinical development.

From Signal to Standard: *The Rising Structural Role of Biomarkers in Oncology Trials*

From Afterthought to Design Driver
Over the past decade, biomarkers have moved from the margins of oncology clinical trials into their structural core. Where they were once treated as exploratory endpoints or optional correlative work, they now shape who is eligible for a study, how early activity is interpreted and which dose ultimately reaches clinical practice. In the era of precision oncology, trials increasingly ask whether a drug works in a biologically defined subgroup, using biomarkers as the organising framework for design and analysis.

From the perspective of an early-phase oncologist, the most important shift is not simply that biomarkers are used more often, but that they are integrated earlier and more deliberately into development. When a biomarker and a drug are conceived and developed together, sometimes even biomarker first, compound second, the odds of selecting the right patients, demonstrating proof of mechanism and identifying a biologically appropriate dose are significantly better.

Biomarkers at the Gate – Patient Selection
In modern oncology, virtually every targeted agent is developed against a specific gene, protein, or pathway, making biomarker-based patient selection the most visible and consequential use of biomarkers in early-phase trials. Enrichment strategies improve signal detection in small cohorts, but they carry trade-offs; narrow eligibility criteria can hamper recruitment and may exclude patients who might benefit. In practice, broad molecular profiling panels assessing multiple alterations from a single biopsy or blood draw can open several potential trial pathways for each patient while addressing the underlying biological heterogeneity of cancer.

Beyond Inclusion – Biomarkers as Proof of Mechanism
Once a patient is enrolled, biomarkers take on a second critical role: confirming that the drug is doing what it is supposed to do. In early-phase trials, where traditional efficacy endpoints are immature, pharmacodynamic biomarkers are the primary evidence that a compound is engaging its target, modifying relevant pathways and producing measurable biological effects. This proof-of-mechanism work demands repeated blood sampling, extended pharmacokinetics visits and serial biopsies, a significant burden on patients and sites but one that is essential for sound go/no-go decisions. That burden still requires a clear justification for every sample. Each biopsy, blood draw or additional timepoint should answer a defined development question that cannot be addressed through a less invasive approach. Protocol teams must balance the value of richer biological data against procedure-related risk, travel, time in clinic and cumulative fatigue, particularly for patients who are already heavily treated. The value of this approach is well illustrated by the development of pembrolizumab; retrospective analysis of PD-L1 expression in a Phase 1 cohort drove adaptive redesign of the study and PD-L1 ultimately became a standard first-line treatment biomarker in non-small cell lung cancer, demonstrating how mechanistic evidence generated in early trials can reshape an entire therapeutic class.

Tissue, Liquid Biopsy and the Expanding Biomarker Toolbox
The tools available for biomarker assessment in oncology trials have expanded considerably over the past decade and early-phase studies are at the forefront of integrating them.

Histological assessment of tissue biopsies remains the gold standard for proof of mechanism; regulators expect tissue-based evidence of target engagement in the early-phase setting and this is unlikely to change. Blood-based approaches including circulating tumour DNA, cell-free DNA and whole exome sequencing complement tissue assessment by enabling dynamic monitoring across treatment cycles, capturing resistance mechanisms and tumour heterogeneity that a single biopsy may miss. Complementing tissue-level data with liquid biopsy is now standard practice in many Phase 1 studies. Looking ahead, multi-omics integration, spanning genomics, transcriptomics, proteomics and immune profiling, will create more complete biological portraits of individual tumours. However, realising that potential requires standardised data collection and regulatory frameworks that can keep pace.

Co-Development of Drugs and Diagnostics
Perhaps the most important conceptual shift in modern oncology drug development is the recognition that a drug and biomarker cannot be developed independently. When a biomarker is identified retrospectively, after a compound has already moved through early- and mid-stage development, the result is often a validation gap that delays or undermines the ability to use it at scale. Companion diagnostic platforms, assay cutoffs and validation plans should be defined during Phase 1 design, not after. Yet early assays and cutoffs often need refinement as dose-escalation data reveal which biological signals matter. Waiting for complete validation can make early use impractical, while postponing validation until expansion may leave a promising drug without a dependable way to select patients. Parallel development allows analytical validation, cutoff refinement and regulatory discussion to progress alongside dose escalation. Equally important is practicality: a biomarker that works in a specialist centre but is unaffordable or unavailable in routine practice is of limited value to the field. Regulators, particularly in Europe, need frameworks that support parallel validation during dose escalation so that by the time expansion cohorts open, the tools to identify patients are already fit for purpose.

Project Optimus and Dose Optimisation
One of the most consequential regulatory shifts in recent years is the move away from the assumption that the highest tolerated dose is the best dose. Project Optimus, the US Food and Drug Administration's initiative to reform dose selection in

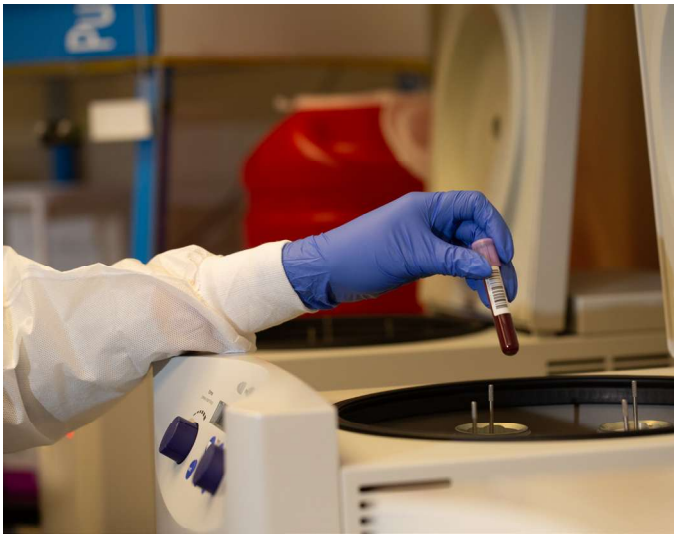
oncology, requires sponsors of targeted therapies to conduct structured dose-optimisation work and justify dose based on a broader evidence base than tolerability alone. Biomarkers are central to this. When two doses appear similarly safe clinically, pharmacodynamic and molecular readouts can reveal whether higher exposure adds meaningful biological activity or whether a lower dose already achieves full target engagement. Nearly every early-phase protocol now includes a formal dose-optimisation cohort following escalation, with biomarker endpoints built in to support the comparison. This approach has significant patient benefit, reducing toxicity levels and improving quality of life without sacrificing efficacy.

Operational Realities of Biomarker Inclusion

Making biomarkers structural to trial design increases operational complexity across screening, testing, logistics and site coordination. Experienced multidisciplinary teams bringing together investigators, pathologists, molecular laboratory staff and sponsors are essential to interpreting biomarker data within clinical context and translating findings into real-time decisions. The challenge intensifies when a target attracts broad simultaneous interest. When multiple sponsors pursue the same biomarker, each may impose different assay platforms, cutoffs or central testing requirements, creating a fragmented landscape for sites trying to screen patients across trials. Trials targeting MTAP are a current example; choosing between expensive central next-generation sequencing and more affordable local immunohistochemistry reflects both scientific and practical pressures. When each sponsor accepts a different method, sites may repeat testing on the same limited tissue, extend turnaround times and increase screening costs. Shared performance standards and broader acceptance of validated local results could reduce duplication while preserving assay quality. MTAP therefore illustrates a wider need to harmonise testing for biomarkers pursued across several competing programmes. The answer lies in collaboration. In an integrated early-phase network, a toxicity signal or biomarker finding at one centre informs and can alter trial management across the network to the benefit of both patients and sponsors. For biomarkers that are clinically important and widely pursued, treating assay development and usage optimisation as a collective priority would benefit all stakeholders considerably.

Toward More Precise, Equitable, Biomarker-Driven Clinical Trials

The number of actionable molecular targets in oncology is growing rapidly and so is the pressure to be precise about which biomarkers and drugs are truly worth pursuing. Artificial intelligence is beginning to contribute to target identification, compound prioritisation and complex data analysis; its near-term impact is strongest in basic research and trial design, though its use across all aspects of trial execution is not too far in the future. Multi-omics integration, combining genomic, transcriptomic, proteomic and immune profiling data, is creating richer pictures of tumour biology and drug response. However, its systematic adoption in early-phase trials requires further standardisation and regulatory clarity. Antibody-drug conjugates add another important test case. As these agents expand across tumour types and into earlier treatment settings, development will need to define not only the target but also the relevant expression threshold, tumour context and changes under treatment. Target expression alone does not consistently predict benefit,



so biomarker strategies must also account for heterogeneity and resistance rather than assume a simple drug-to-target match. The ultimate goal is to use these advances to benefit patients, an outcome that will require sustained commitment to validation, standardisation and equitable access to testing. The structural integration of biomarkers into oncology trials will be the foundation on which more precise, more effective and more equitable cancer care will be built.

REFERENCES

1. Tu Y, *et al.* Latest developments in adaptive enrichment clinical trial designs in oncology. Semin Oncol. 2024. PMID: 39271644.
2. Duan XP, *et al.* New clinical trial design in precision medicine. Signal Transduct Target Ther. 2024;9(1) :57.
3. Incorvaia L, *et al.* PD-L1 as a predictive biomarker for pembrolizumab. Front Oncol. 2019;9:1044.
4. US Food and Drug Administration. Project Optimus. FDA Oncology Center of Excellence. 2024. https://www.fda.gov/about-fda/oncology-center-excellence/project-optimus
5. Kang SL, *et al.* Institutional comparison of FDA, EMA, and MFDS review of companion diagnostics. Front Oncol. 2023. PMC10474566.
6. ctDNA as a biomarker for oncology drug development in phase I clinical trials. Expert Rev Mol Diagn. 2025. PMID: 40622305.
7. The Future of Multi-Omics in Cancer Clinical Trials. ACRP. 2024. https://acrpnet.org/2024/08/15/the-future-of-multi-omics-in-cancer-clinical-trials
8. Biomarker-Driven Oncology Trial Design and Subgroup Analysis. JCO Precision Oncology. 2024. https://ascopubs.org/doi/10.1200/PO.24.00116
9. Nicolò E, *et al.* Biomarkers for antibody-drug conjugates in solid tumors. Cancer Discov. 2024;14(3):442–456. PMID: 38363729.

Dr. Tatiana Hernández Guerrero

Tatiana Hernández Guerrero, MD, PhD, is Director of Clinical Research at START Barcelona, part of The START Center for Cancer Research's global network of early-phase oncology clinical trial sites. Dr Hernández's clinical focus is on DNA-repair targeted therapies and immunotherapy, and she has served as Investigator on more than 60 clinical trials. She received her MD from Carabobo State University in Venezuela and a PhD in Medicine and a master's degree in healthcare management from ESADE in Spain.

Turning Oncology Data into Better Development Decisions

Oncology remains one of the highest-attrition areas in drug development, with only a small minority of candidates entering clinical testing going on to reach approval. The reasons are familiar, including lack of efficacy, unexpected toxicity or insufficient differentiation, but the deeper cause is often a persistent disconnect between what is measured in the laboratory and what ultimately matters in patients.

The recent discontinuation of Roche's tiragolumab programme, an anti-TIGIT antibody, is instructive. An encouraging randomised Phase II signal supported development across several tumour types and thousands of patients. The results proved inconsistent. The pivotal NSCLC study missed its overall survival endpoint, while a parallel study in oesophageal squamous cell carcinoma met both of its primary endpoints. TIGIT is a well-characterised inhibitory receptor and the rationale for blocking it was biologically coherent. What the programme lacked was a way to anticipate where that rationale would hold. Trials enriched for PD-L1 expression, but no marker specific to TIGIT biology distinguished the settings in which the combination helped from those in which it did not.

Failures of this kind are rarely failures of data volume. The mechanistic evidence existed and the trials were large. What was absent was a chain connecting laboratory measurement to the decision it was meant to inform, so that the biology could be tested in the population where it applied. As data generation accelerates and computational methods mature, the central challenge is therefore no longer access to data. It is whether that data is captured, connected and interpreted well enough to guide confident decisions at every stage of a programme.

The contrast with genotype-directed development is stark. In the recent pivotal trial of daraxonrasib in previously treated metastatic pancreatic cancer, the primary endpoints were specified in patients carrying RAS G12 mutations and the trial reported a median overall survival of 13.2 months against 6.7 months with chemotherapy. The selecting biomarker was the target itself: measurable at baseline, causally implicated in the disease and defined before the trial began rather than sought afterwards.

In this article, we will explore four connected shifts reshaping how the field uses data: linking preclinical and clinical evidence, the rise of real-world and multimodal datasets, faster biomarker discovery and patient stratification and the opportunities and challenges of using data to de-risk oncology pipelines.

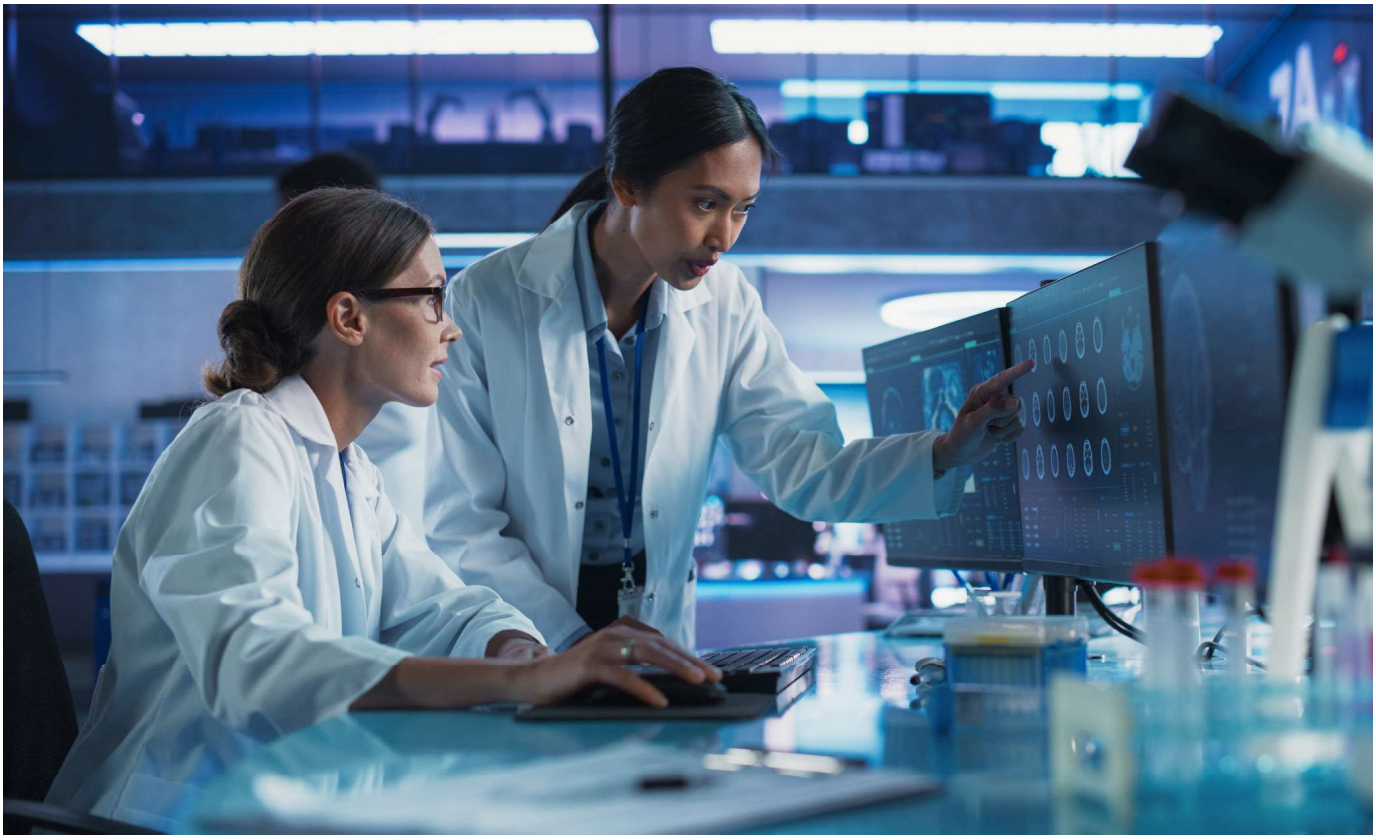
Connecting Preclinical and Clinical Data to Improve Decision Making

For decades, preclinical and clinical research have operated as largely separate worlds, each with its own datasets, endpoints and assumptions. Early-stage oncology is rich in hypothesis-driven experimentation but comparatively weak at building holistic, predictive views of tumour biology that anticipate downstream clinical questions. Individual data layers, including genomics, transcriptomics, proteomics and phenotypic readouts, are frequently analysed in isolation, producing conclusions that hold within a narrow experimental context, but fail to generalise to patients. Closing this gap requires treating preclinical work not as a standalone proof-of-concept exercise but as the first step in a continuous translational chain.

A recurring example is the assumption that high transcript abundance implies durable protein expression or functional relevance. The assumption persists in part because transcript is the data layer preclinical programmes most readily generate, while the question it is used to answer, whether a patient carries a target a therapy can engage, belongs to the clinic. Transcript and protein abundance correlate only moderately, and for target-directed modalities the discrepancy matters most at the cell surface, where the relevant question is not whether a gene is expressed but whether the protein is present, correctly localised and accessible to a binding agent. Surface and membrane proteomics remain among the least routinely generated layers in preclinical characterisation, despite being the layer that determines whether an antibody, a conjugate or a cell-based therapy has anything to engage.

The practical value of connecting these worlds is that early experiments become more decision-relevant. When preclinical study design is aligned with prospective patient stratification strategies, teams can interrogate the questions that will define clinical success, namely who responds, why resistance emerges and how a candidate compares to the standard of care, before committing significant investment. That last comparison is most informative when the reference arm is generated concurrently and in the same models, rather than inferred from historical or vehicle-controlled data, since a candidate's value is defined by its differentiation from the current standard rather than by activity in isolation.

Three design choices do most of the work and each addresses a specific failure mode. Human-relevant systems reduce the risk that a result reflects the model rather than the disease, longitudinal sampling captures what changes under treatment rather than what is true at a single endpoint and functional validation tests whether a molecular association actually drives response or merely accompanies it. "Human-relevant" covers a range of systems and the useful question is which one answers the question being asked. Patient-derived organoids retain much of the genomic and phenotypic character of the source tumour and can be screened at a scale that supports systematic hypothesis generation, which suits them to early target and combination triage. Patient-derived xenografts preserve tumour architecture and heterogeneity within a living host, which allows pharmacology, dosing and response to be studied *in vivo* and sampled repeatedly across the course of treatment. Where the



observed process is immune-mediated, systems carrying a human immune compartment are required. Used together, organoids and xenografts address different halves of the same problem, the first generating candidates at scale and the second testing whether those candidates survive the conditions of actual treatment. The choice of system also determines what can be asked later. Much of the value in translational data lies in the ability to go back to the same models and ask the next question, testing a hypothesis by perturbation rather than reanalysis. A static archive can only be interrogated statistically. A system that can be experimentally re-entered allows an association to be tested as a mechanism, which is the step that converts a correlation into something a development team can act on. Planned together, these choices produce a translational narrative that can be followed forward rather than reconstructed retrospectively from disconnected studies.

The Growing Role of Real-World and Multimodal Datasets

The digitalisation of healthcare has produced an unprecedented explosion of data types for each patient, spanning clinical records, multi-omics at bulk, single-cell and spatial resolution, medical imaging and, increasingly, wearable and sensor data. Real-world data, drawn from electronic health records, claims databases and patient registries, adds a further dimension by capturing how disease and treatment behave outside the controlled confines of a clinical trial. Used well, real-world evidence can confirm existing knowledge, generate new hypotheses about tumour biology and de-risk future studies by informing trial design and eligibility. The strategic shift underway is away from single-modality analysis and toward integrated approaches that combine these diverse layers into a coherent view.

The reason integration matters is fundamental: while single-layer datasets reveal association, combining molecular profiling with functional readouts of a deliberately perturbed system enables mechanistic conclusions. Bringing molecular, functional, imaging and clinical information together allows researchers to interrogate biology in context. Fragmentation is the practical obstacle to this, as high-quality data is often siloed across institutions, platforms and formats, generated under different assumptions and rarely harmonised. Realising the value of multimodal data therefore depends on several enabling conditions:

- **Matched design:** generating functional response and molecular profiling on the same models rather than associating them across separate cohorts, so that observed relationships can be interpreted together rather than inferred statistically from other models.
- **Interoperability:** harmonising structured, semi-structured and unstructured data so modalities can be analysed together rather than in isolation.
- **Provenance and quality:** ensuring datasets are clinically annotated, well documented and fit for the question being asked.
- **Clinical relevance:** prioritising data that reflects contemporary, pre-treated patient populations rather than idealised or historical cohorts.
- **Analytical capability:** deploying methods capable of reasoning across modalities without losing biological interpretability.
- **Governance:** managing privacy, consent and security, increasingly through federated and privacy-enhancing

approaches that allow learning without centralising sensitive data.

Accelerating Biomarker Discovery and Patient Stratification

Biomarkers sit at the heart of modern oncology, yet a stubborn gap persists between preclinical promise and clinical utility. Many candidate biomarkers fail to cross the divide because they were identified in systems that poorly reflect human biology, advanced without demonstrating reproducibility across cohorts or tested against uniform preclinical populations that mask real-world heterogeneity. Multi-omics profiling, integrating genomics, transcriptomics, proteomics and metabolomics, has expanded the search space considerably, enabling the identification of context-specific, clinically actionable markers that a single data layer would miss.

The most durable strategies pair this molecular depth with functional and longitudinal validation, capturing how biomarkers behave dynamically under therapeutic pressure rather than at a single static timepoint. Resistance biomarkers place a further demand on study design, since the informative signal appears between baseline and progression rather than at either end. Longitudinal sampling through treatment and into relapse can shift an observation of response to an explanation of why response was lost.

Patient stratification is where this work delivers tangible value. By linking molecular features to response and resistance, stratification strategies allow developers to define responder and non-responder populations, design rational combinations and select indications and expansion cohorts with greater precision. Computational methods, particularly machine learning, are increasingly used to support this process, as fragmented genomic, clinical and real-world evidence is unified into a single analytical layer, though the scale of their effect on discovery speed is still being established. The key discipline is preserving interpretability. A stratification hypothesis is only useful to a development team, a regulator or a clinician if the biology behind it can be understood and defended. As the field matures, the emphasis is shifting from generating ever more data toward generating stratification insight that changes trial design.

Opportunities and Challenges in Leveraging Data to De-risk Oncology Programmes

The overarching promise of these trends is de-risking, meaning better and earlier go/no-go decisions before costs escalate. When biology, function and clinical reality are connected, teams can pressure-test a hypothesis, anticipate resistance, refine dose rationale and strengthen competitive positioning long before a candidate reaches a pivotal trial. Data-rich approaches also support a more proactive development philosophy, one focused on sustaining response and pre-empting failure rather than reacting to it. For sponsors and investors alike, translationally grounded evidence increasingly serves as a differentiator in a competitive funding environment, offering a more objective basis for portfolio decisions. The most valuable outputs are those that carry a next step, showing not simply that resistance emerged, but through which mechanism, and therefore whether the rational response is a combination, a different modality directed at a newly exposed target or a decision to stop. The opportunity, in short, is to convert data from a descriptive record into a predictive asset.

The challenges, however, are real and should not be understated. Fragmentation, inconsistent data quality and limited interoperability continue to constrain even well-resourced organisations and the value of any model is ultimately bounded by the relevance of the data used to build it. The binding constraint is often not data volume but trustworthy, translatable ground truth, meaning measurements made in systems that behave enough like patients that a conclusion drawn from them survives contact with the clinic. That is a far scarcer commodity than data, and it does not become less scarce as data accumulates. Several tensions warrant particular attention as the field advances:

- **Data relevance over data volume:** large datasets built on treatment-naive or non-representative samples can mislead as easily as they inform. Models and cohorts carrying documented prior treatment and response history are considerably more informative for contemporary questions, where many patients have already progressed on at least one line of therapy.
- **Reproducibility and standards:** without shared validation frameworks, promising signals may not survive replication across cohorts.
- **Interpretability versus performance:** black-box models can undermine regulatory confidence and clinical trust if their outputs cannot be explained.
- **Privacy and governance:** integrating real-world and patient-level data demands rigorous consent, security and increasingly federated architectures.
- **Organisational readiness:** extracting value requires cross-disciplinary expertise that bridges biology, data science and clinical strategy.

Navigating these tensions is now a core competency rather than a peripheral concern. The organisations best positioned to benefit are those that invest not only in generating data, but in connecting, curating and interpreting it with clinical relevance in mind. Ultimately, the future of oncology development will be defined less by how much data a programme can accumulate and more by how effectively that data can be turned into confident, defensible decisions, marking the difference between measuring biology and genuinely understanding it.

Tammer Farid

Tammer Farid, PhD. is General Manager, Data at Champions Oncology. With 15 years at the intersection of data, technology, health care and life sciences, he has led strategy, advisory and operational roles. These have focused on portfolio development, commercial growth and translational services for precision medicine, including large-scale data, analytics and companion diagnostics. He was previously a Partner at Boston Consulting Group's Healthcare practice. Tammer holds a PhD from the University of Pennsylvania and resides in San Diego.