EXECUTIVE SUMMARY

From Bench to Bedside and Back Again: Strategies for effectively leveraging preclinical data to accelerate R&D

PRESENTED BY: RockStep

PUBLISHED BY: FIERCE Biotech

NOVEMBER 2023

PRESENTED BY: RockStep

PUBLISHED BY: FIERCE Biotech

From Bench to Bedside and Back Again:
Strategies for effectively leveraging preclinical data to accelerate R&D

CONTENTS:

Facilitator: Julia Douthart, Fierce Biotech

Panelists:

- **Chuck Donnelly**, CEO and Co-Founder, RockStep Solutions
- **Amy Huff**, Senior Director of Global Operations, Charles River Laboratories
- **Szczepan Baran, VMD/MS**, Chief Scientific Officer, VeriSIM Life
- **Michael Hayward, PhD**, Senior Director, *In Vivo* Pharmacology, JAX Mice, Clinical & Research Services, The Jackson Laboratory

Overview

The drug discovery process is complex and time-consuming, with an average development time of over 10 years and hundreds of millions to billions of dollars spent to receive regulatory approval. As recently as 2015, researchers concluded that anywhere from 50-80% of pre-clinical animal *(in vivo*) research data was irreproducible, representing a monetary loss of more than **$28 billion**. For an industry renowned for advancing human knowledge, it is shocking to realize that 90% of *in vivo* study data is captured and recorded either on paper, disparate legacy software like spreadsheets, or other in-house developed systems.

Why is there a reluctancy to move away from antiquated technology and 40-year-old legacy systems?

Animal research is highly regulated. *In vivo* models are often developed with hands-on approaches in academic laboratories. Scaling these models to test multiple compounds reliably and consistently must align with OECD guidelines[1] and remain cost-effective for contract research organizations (CROs). "The CROs' pace cannot stop to transition to novel systems," remarked Amy Huff, Senior Director of Global Operations at Charles River Laboratories.

Modernization involves change management that encompasses data harmonization, animal welfare, laboratory process improvements, and budget. The panelists unanimously agreed that staying ahead involves parallel work with legacy systems and modern technology. They also pointed out that the fear of not achieving a return on investment underlies a hesitancy to move forward.

Modern data management systems facilitate access to aggregated data in real-time. Connected data can be transformed into actionable insights, but most current technology cannot engage with data across multiple legacy systems. Organizations may take years to digitize archived preclinical information and re-establish benchmark validation criteria, noted Dr. Szczepan Baran, CSO at VeriSIM Life.

Chuck Donnelly, CEO and Co-Founder of RockStep Solutions, highlighted the five main challenges of introducing new systems into *in vivo* research:

- **Change management.** Employees hesitate to embrace change in fast-paced workplaces.
- **Regulatory compliance.** Digital data management systems must ensure consistent and objective collection of data.
- **Adaptability.** Even robust workflows require flexibility to overcome problems that arise.
- **User-friendly interfaces.** Systems should be easy to navigate and configure.
- **Cloud-based software.** Researchers need quick and secure access to their data.

> *In vivo* research is considered the final frontier of informatics because it is so complex. Working with live organisms involves adapting to daily changes and the conflicting requirements of flexibility and precision.
>
> Chuck Donnelly, CEO and Co-Founder of RockStep Solutions

What is clean data?

"Clean data is collected under identical conditions of a standardized protocol," explained Dr. Michael Hayward, Senior Director of *In Vivo* Pharmacology at The Jackson Laboratory. "Quantitative measures, which can be analyzed by parametric statistics, are preferred over subjective clinical observations."

Semantically consistent and annotated data is imperative for preclinical research.

Semantic drift occurs when the meaning or spelling of a word evolves over time. In biomedical research, semantic drift is reflected by inconsistent terminology, taxonomy, and data granularity in the literature between organizations or within the same research group. Results from the same center may vary due to slight changes in environmental conditions or how individuals define which data is findable, accessible, interpretable, and reusable (FAIR Data Principles[2]). These inconsistencies, reflected in many public datasets, hinder upstream and downstream analyses. Similarly, unavailable metadata (e.g., the transgenic background of a mouse) may reduce the accuracy of machine learning (ML) predictions.

Collaboration between technology providers is key to reducing semantic drift and leveraging the data's full potential. Conversations between end-users in the pharmaceutical and biotech industry and regulators will help standardize terminology in disconnected clinical systems. State of the art tools, such as Climb™, organize vocabulary terms, used for annotating data, via a relational model creating a centralized ontology server. Relational databases facilitate the training of ML models[3] by arranging data points with defined relationships. CROs can train the end-users on how to use nomenclature and select the standardized terms from a drop-down list, noted Donnelly.

Why is clean data important?

Between 50% to 80% of preclinical animal research data is irreproducible,[4] which is both alarming and unsustainable. Reproducibility is often affected by deviations or lack of experimental metadata (e.g., animal genotypes, genetic drift, gender, individual microbiome, time points, environmental conditions). Reliable interpretation of study findings depends on each variable being well-documented.

However, fixating on reproducing certain *in vivo* model characteristics might not be the best use of resources; *in vivo* observations might not necessarily translate to what occurs in clinical trials. Genetic diversity exists among inbred rodent models and clinical trial participants. Not all physical traits matter for downstream scientific interpretation but recording them remains useful. One example from Huff: animal eye color may not be relevant for drug interactions in humans.

> "Experimental reproducibility is the keystone of life science–it affects the risk to humans."
>
> Chuck Donnelly, CEO and Co-Founder of RockStep Solutions

Applying the Pareto Principle (80:20 rule) to *in vivo* research, 20% of scientists' efforts can solve 80% of problems. Upholding rigorous experimental standards and using tools like Climb™ to consistently capture metadata will help reduce the compounding costs of human errors, currently estimated to be over $28 billion.

Can we utilize AI/ML when there is a tangible lack of raw data?

"Biology is too complex for artificial intelligence (AI) to solve everything," said Dr. Baran. The limitations and applications of AI must be recognized. First, AI-based models cannot include unknown biological processes, molecular functions, and pathways. Second, ML algorithms of AI assistants (e.g., Chat GPT or generative AI) must be continuously fed with information. In contrast, independent AI models are trained to make clinical predictions based on *in vivo* data. The independent AI models designed to answer specific scientific questions have the most significant impacts. Notably, hybrid mechanistic ML and deep-learning simulation models can be trained with fewer records if data is high-quality and well-curated.

Data management systems allow widespread access to organized digital records and efficient extraction of clinically relevant metadata for analysis. Historical data can be employed for virtual control groups in safety studies, while synthetic data fills knowledge gaps and reveals new nuances. In this regard, AI-generated data facilitates study design, reduces time to market, aids internal decision making, and supports R&D packages.

The panelists acknowledged that there is currently insufficient data to create fully digital human or animal models for drug testing. However, in May of 2023, the Food and Drug Administration (FDA) requested feedback on the use of AI/ML in the development of drugs and biologics[5], indicating regulators are open to discussing how this type of data is being interpreted and perceived by peer reviewers.

Live animal alternatives for *in vivo* research: a utopian ideal or R&D reality?

Initiatives for animal use alternatives (3Rs)[6] and the FDA Modernization Act 2.0[7] strive to create a reality where live experimental animal models can be replaced by computerized models or organ-on-a-chip technology. "We have a responsibility as the caretakers of these animals," said Huff. "Leveraging opportunities for virtual control groups helps us reduce the number of animals used experimentally."

"Animal experiments are not going away any time soon, but I would not underestimate the potential of the technology," added Dr. Hayward. The Jackson Laboratory, among other institutions, is working on creating smarter models aligning with the 3Rs. These efforts include humanizing the immune system by engrafting human immune systems into mice, genetically modifying animals that express human mutations, and producing heterogeneous mice to test aging, among others.

Data management systems allow widespread access to organized digital records and efficient extraction of clinically relevant metadata for analysis. Historical data can be employed for virtual control groups in safety studies, while synthetic data fills knowledge gaps and reveals new nuances.

Listen to the full webinar on-demand.

The motivation for the development of Climb™

Prior to RockStep, Chuck Donnelly was Director of Computational Sciences at the Jackson Laboratory (JAX), where he led work on intractable computational problems for numerous disease areas, including cancer and diabetes. Chuck's work at JAX was funded by the NIH, HHMI, and JAX. Prior to JAX, Donnelly was in the Space Astrophysics Group at UC Berkeley, where he directed the development of software systems for NASA orbiting platforms, worked on the Space Shuttle ground control, and led the development of software for imbedded systems and algorithms for radio astronomy.

The creation of RockStep Solutions and Climb was driven by the passion to save lives by helping get drugs and vaccines to market faster. In 2014, RockStep was selected for an SBIR Award (GM112206) from the NIH, based on innovative ideas around building lab information systems to transform the way laboratory workflows were managed. RockStep collaborated with Novartis, the lab of a Nobel Laureate (Dr. Prusiner at UCSF), and a sizeable public biotech, to design and create Climb™–now widely adopted by world-leading biomedical research and drug discovery organizations.

RockStep continues to work with frontier scientists and engineers on some of the world's most advanced technologies to ensure Climb™ leads the industry with an exceptional *in vivo* research and data management software for both the lab of now and the lab of the future.

Request a demo of Climb™ or explore additional resources from RockStep Solutions to learn how to unlock the value of preclinical data and expedite breakthroughs that improve patients' lives.

Endnotes

1. Guidelines for Multinational Enterprises on Responsible Business Conduct (2023) Organisation for Economic Co-operation and Development (OECD).
2. The FAIR Guiding Principles for scientific data management and stewardship. Wilkinson M, Dumontier M, Aalbersberg I, et al. (2016). Sci Data 3, 160018.
3. Learning models over relational data: A brief tutorial. Schleich, M. et al. (2019). In Scalable Uncertainty Management: 13th International Conference, Compiègne, France, December 16–18 (2019). Proceedings 13 (pp. 423-432). Springer International Publishing.
4. The Economics of Reproducibility in Preclinical Research. Freedman LP, Cockburn IM, Simcoe TS et al. (2015). PLOS Biology 13(6): e1002165.
5. Artificial Intelligence and Machine Learning (AI/ML) for Drug Development (2023) FDA.
6. Animal Use Alternatives (3Rs) (2023) USDA National Agricultural Library.
7. The FDA Modernization Act 2.0: Drug Testing in Animals is Rendered Optional Adashi EY, O'Mahony DP, Glenn Cohen I et al. (2023). The American Journal of Medicine 136(9):853-854.

RockStep™

Climb was developed by scientists for scientists to modernize *in vivo* research and harness the value of preclinical study data.

Originally funded by the NIH, RockStep works in collaboration with industry leaders to design and develop Climb – an innovative, cloud software solution that streamlines *in vivo* workflows and ensures robust, high-integrity preclinical study data to accelerate time to clinical trial.