Tachinardi, U. (2026) Navigating the bumpy road from data to health: a worthwhile journey? *Proceedings 13th World Congress on Genetics Applied to Livestock Production*: https://doi.org/10.31274/wcgalp.24482 © 2026 *Tachinardi* CC BY-NC-ND

Navigating the bumpy road from data to health: a worthwhile journey?

U. Tachinardi[1]

[1]*University of Cincinnati College of Medicine, Department of Biostatistics, Health Informatics and Data Sciences, Cincinnati, USA; corresponding author: umberto.tachinardi@uc.edu*

Introduction

Data availability has transformed human health sciences and livestock production. In medicine, electronic health records, clinical registries, genomics, imaging, and patient data revolutionize research and care. Similarly, livestock systems now leverage genomics, high-throughput phenotyping, environmental sensing, and detailed production data. The promise is to learn from data, personalize decisions, and improve outcomes at scale. However, translating insights into practice is challenging. Success depends not only on data and algorithms but also on data quality, integration, governance, skills, incentives, and operational systems. Human healthcare experiences offer lessons for livestock, respecting their unique biological and economic contexts.

Learning from data: from discovery to tailored interventions

a. Accessing data and data quality as foundational challenges

Across both human healthcare and livestock systems, the ability to learn from data is constrained from the outset by challenges in data access and data quality. In human healthcare, much of the relevant data is generated as a byproduct of care delivery and operations, rather than for research or analytics. Electronic Health Records (EHRs), Picture Archiving and Communication Systems (PACS), laboratory systems, and a growing ecosystem of digital tools generate vast quantities of data, but these data are often fragmented, incomplete, noisy, duplicated, or inconsistent. Similar issues arise in livestock systems, where data are collected and stored (many data siloes) across heterogeneous production environments, different players (e.g., nutrition and genetics companies), technologies and management practices.

Data quality issues—such as missingness, variable accuracy, temporal misalignment, and shifting definitions—are not merely technical inconveniences. They directly shape what can be learned and how reliable models can be. Models trained on poor-quality data do not simply underperform; they risk producing outputs that appear precise while masking underlying uncertainty. As data volumes grow, these risks can be amplified rather than mitigated.

b. Feature extraction from unstructured and multi-modal data

A growing share of high-value information in both human healthcare and livestock systems resides in unstructured or weakly structured data. In human healthcare, this includes clinical narratives such as admission notes, discharge summaries, operative reports, pathology reports, and radiology interpretations. Images, waveforms, genomic sequences, and free-text documentation all require sophisticated feature extraction before they can be incorporated into learning pipelines.

Advances in natural language processing, computer vision, and bioinformatics have enabled substantial progress in extracting signal from these sources, but the process remains complex, computationally intensive, and sensitive to context. Extracted features are shaped by documentation practices, instrumentation, and local conventions, making portability and

generalization challenging. Similar complexities arise in livestock systems when integrating genomic data, sensor streams, images, and observational records.

These feature-extraction challenges reinforce the importance of multimodal integration. Outcomes are rarely defined by a single data type; rather, they emerge from interactions among phenotypes, genotypes, environments, and behaviors. Learning systems that successfully combine these modalities are better positioned to support both discovery and intervention, but they require deliberate design and governance.

c. Building the pipeline from data to actionable interventions

Moving from data and features to effective interventions requires a disciplined pipeline that treats translation as a continuous process rather than a one-time deployment. Across human healthcare and livestock systems, this pipeline begins with robust data management and curation practices that ensure traceability, versioning, and fitness for purpose. Without these foundations, downstream analytics inherit ambiguity and bias that are difficult to correct later.

Equally critical are rigorous approaches to model development. Training, testing, and validation must be designed to reflect real-world use, including temporal drift, population heterogeneity, and operational constraints. Overfitting to historical data or narrowly defined benchmarks can yield models that perform well in development but fail in practice. As models move closer to decision-making, expectations for transparency, robustness, and reproducibility increase.

Once deployed, actionable algorithms require ongoing recalibration and localization. Models trained in one context rarely transfer unchanged to another. Differences in populations, workflows, data capture practices, incentives, and environmental conditions necessitate local adaptation, followed by re-validation. Continuous monitoring is essential to detect performance degradation, unintended consequences, and shifts in underlying data distributions. In both healthcare and livestock genetics, the absence of such monitoring has frequently led to erosion of trust and value over time.

This end-to-end pipeline—spanning data curation, model development, deployment, and monitoring—defines whether analytical insight ultimately produces the “right” interventions and the “best” achievable outcomes. In livestock genetics and production systems, the same principles apply: genomic prediction models, health risk scores, and management optimization tools must be continuously recalibrated as populations, environments, and production practices evolve.

d. Recent breakthroughs in human healthcare applications

In recent years, human healthcare has also seen notable breakthroughs that illustrate both the promise and the complexity of data-driven transformation. Ambient clinical documentation systems reduce clinician burden by capturing and structuring human language during encounters. AI-driven patient monitoring leverages continuous data streams to detect early signs of deterioration. Advances in automated coding and documentation support more accurate billing and revenue cycle operations. These applications demonstrate how learning from data can improve efficiency, quality, and experience when tightly integrated with operational systems.

Translation as the new bottleneck

As scientific discovery accelerates, translation has emerged as the dominant bottleneck across data-intensive domains. In human healthcare, the gap between evidence generation and routine practice has been widely documented. Implementation science provides a formal framework for understanding and addressing this gap, emphasizing context, adoption, fidelity, sustainability, and continuous learning.

Research shows that effective AI interventions require more than technical performance. Models must integrate smoothly into workflows, align incentives, follow regulations, build trust, be interpretable, provide clear value, and be regularly evaluated. Similarly, livestock systems need predictive tools that fit decision-making processes, considering operational, economic, and human factors. Treating implementation as part of the science ensures data innovations lead to lasting improvements, not just short-term pilots.

Managing complexity in data-rich systems

Data-driven systems add complexity and capability. In healthcare, predictive models help anticipate sepsis, reduce readmissions, estimate length of stay, and optimize surgeries and consultations. Automation and analytics support supply chain, logistics, and revenue management. These show how data and AI can boost efficiency and quality when integrated into well-designed systems.

But they also reveal pitfalls. Models based on incomplete or biased data can give false signals. AI outputs may gain undue authority without explainability and controls. Costs can rise quickly with multiple overlapping tools. Managing these risks requires investment in governance, monitoring, and workforce development.These operational lessons are relevant to livestock production. Predictive health models, optimization of breeding and feeding strategies, and automation of logistics all promise value, but only if complexity is actively managed. Success depends on treating data quality, integration, and governance as ongoing responsibilities rather than one-time projects.

Multi-modal data and classification for better outcomes

Across domains, improved outcomes depend on accurate classification. In human health, defining phenotypes, genotypes, and exposomes has been critical to understanding disease mechanisms and treatment response. Multi-modal data enable more precise stratification, supporting both discovery and intervention. These capabilities help identify who is more susceptible to certain conditions, who benefits more or less from particular drugs or procedures, and who may be at risk of adverse effects.

Livestock systems similarly benefit from integrating genomic, phenotypic, and environmental data to classify animals and production contexts. These classifications inform breeding decisions, health interventions, and management strategies. As in human medicine, the challenge lies not only in building models, but in ensuring that classifications remain valid as systems evolve.

Discussion: shared lessons across domains

Several shared lessons emerge from comparing human healthcare and livestock systems. First, abundant data enable learning only when data access, data quality, and integration are actively managed. Second, translating insight into outcome requires organizational alignment, governance, and trust, supported by implementation science methods. Third, complexity and cost grow alongside capability, making prioritization and discipline essential. Finally, multi-modal classification is central to tailoring interventions and improving outcomes, but it demands continuous evaluation.

Human healthcare has invested heavily in translational infrastructures such as CTSA hubs and learning health system capabilities to shorten the path from discovery to routine practice. While contexts differ, livestock systems can benefit from anticipating translation challenges and designing systems that support iterative learning rather than one-time deployment.

Conclusions

The road from data to health—whether in human medicine, lifestyle sciences, or livestock production—is inherently bumpy. Scientific advances and analytical sophistication have outpaced our ability to translate insight into sustained outcomes. Experience from human healthcare suggests that this gap can be narrowed through deliberate investment in translation, governance, implementation science, and continuous learning systems.

Navigating this road is a worthwhile journey. When data are used not only to discover, but to guide action across individuals and populations, they can support healthier lives, improved disease control, and more efficient systems. The challenge ahead is not simply to generate more data, but to build the structures and practices that allow data-driven insight to reliably improve outcomes across domains.