

Next-Gen Statistical Programming in Clinical Trials: Driving Faster, Reliable, and Smarter Outcomes

Authors: Mrityunjay Kumar, Shashikant Kumar
Presentation No. PD05

Stream: Professional Development (PD)
PHUSE APAC 2026, Hyderabad, India 19 - 21 February

Disclaimer: The opinions expressed in this talk are solely those of the authors and do not reflect the views of their employer

Agenda

- Introduction
- Current Trends in Statistical Programming
- What Does “Next-Gen Statistical Programming” Mean?
- Key Components of Next-Gen Statistical Programming
- Real Time Case Studies
- Challenges, Risks, and Mitigation Strategies
- Future Trends and Directions
- Key Takeaways

- ✓ Traditional programming methods are **manual**, **time-consuming**, and prone to **inconsistencies**.
- ✓ Next-Gen Statistical Programming integrates **automation**, **AI/ML**, and **standardization** across the analysis pipeline.
- ✓ Enhances **speed**, **accuracy**, and **reproducibility** in generating ADAM datasets and TLFs.
- ✓ Ensures **regulatory-ready outputs** with **traceability**, **version control**, and automated validation.
- ✓ Empowers programmers to focus on **interpretation** and **insights**, not **repetitive** coding tasks.



Current Trends in Statistical Programming

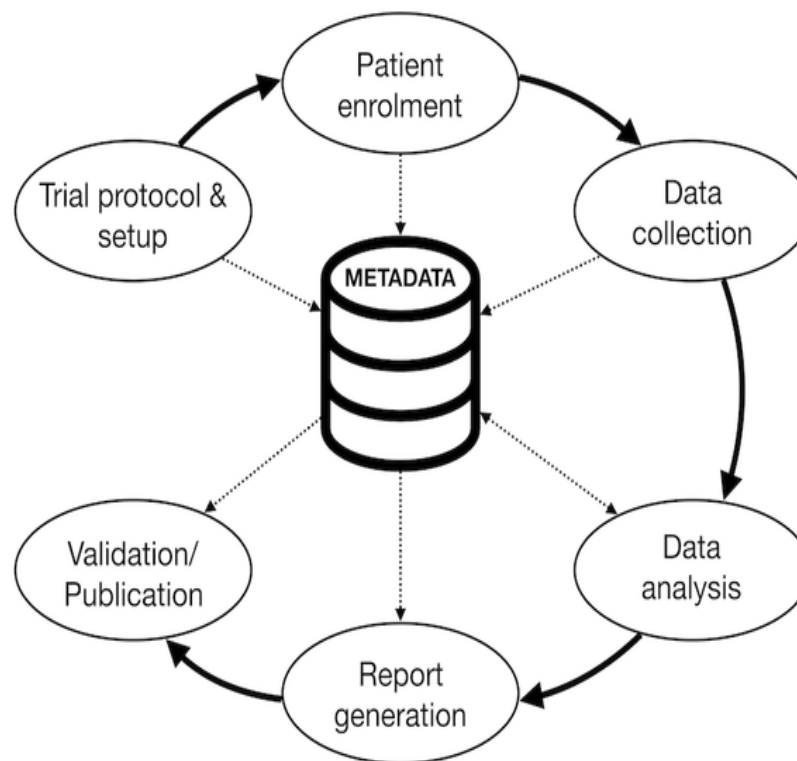
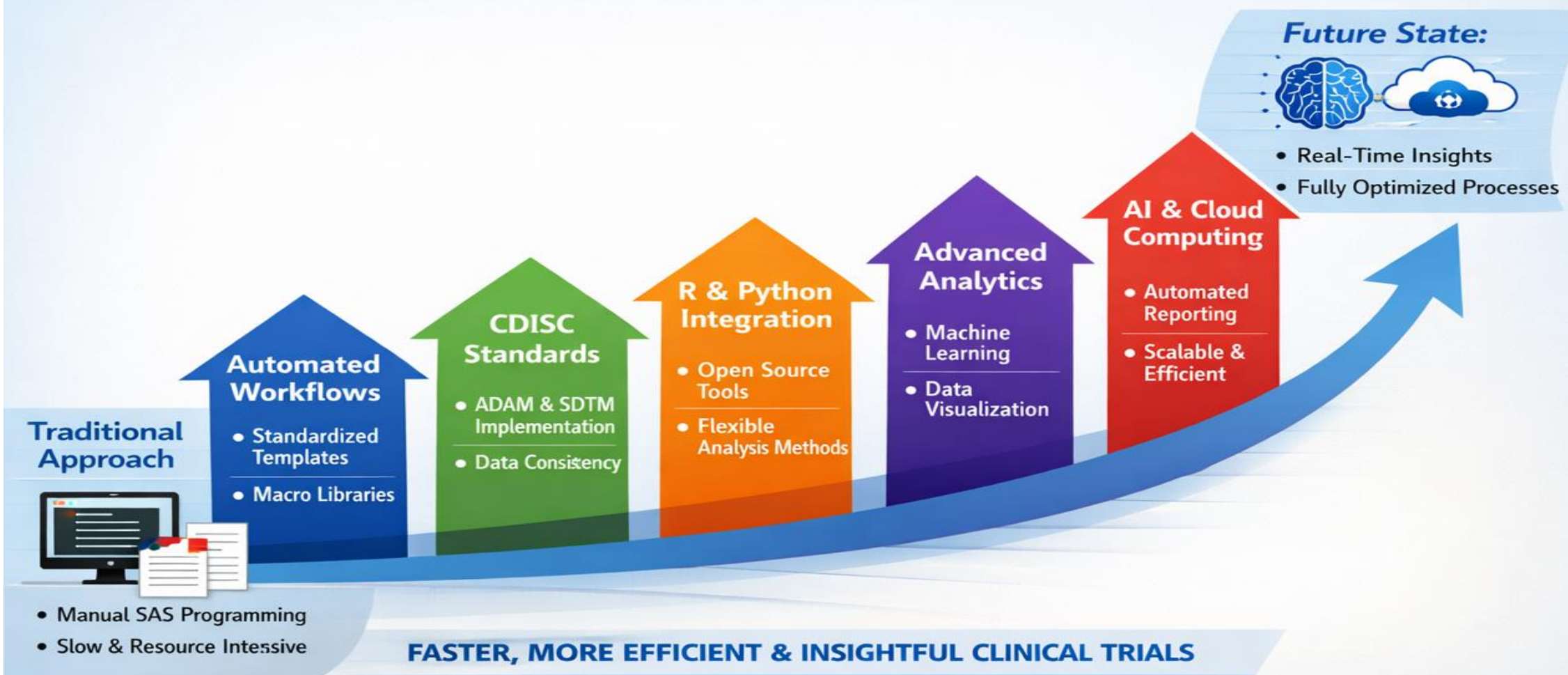
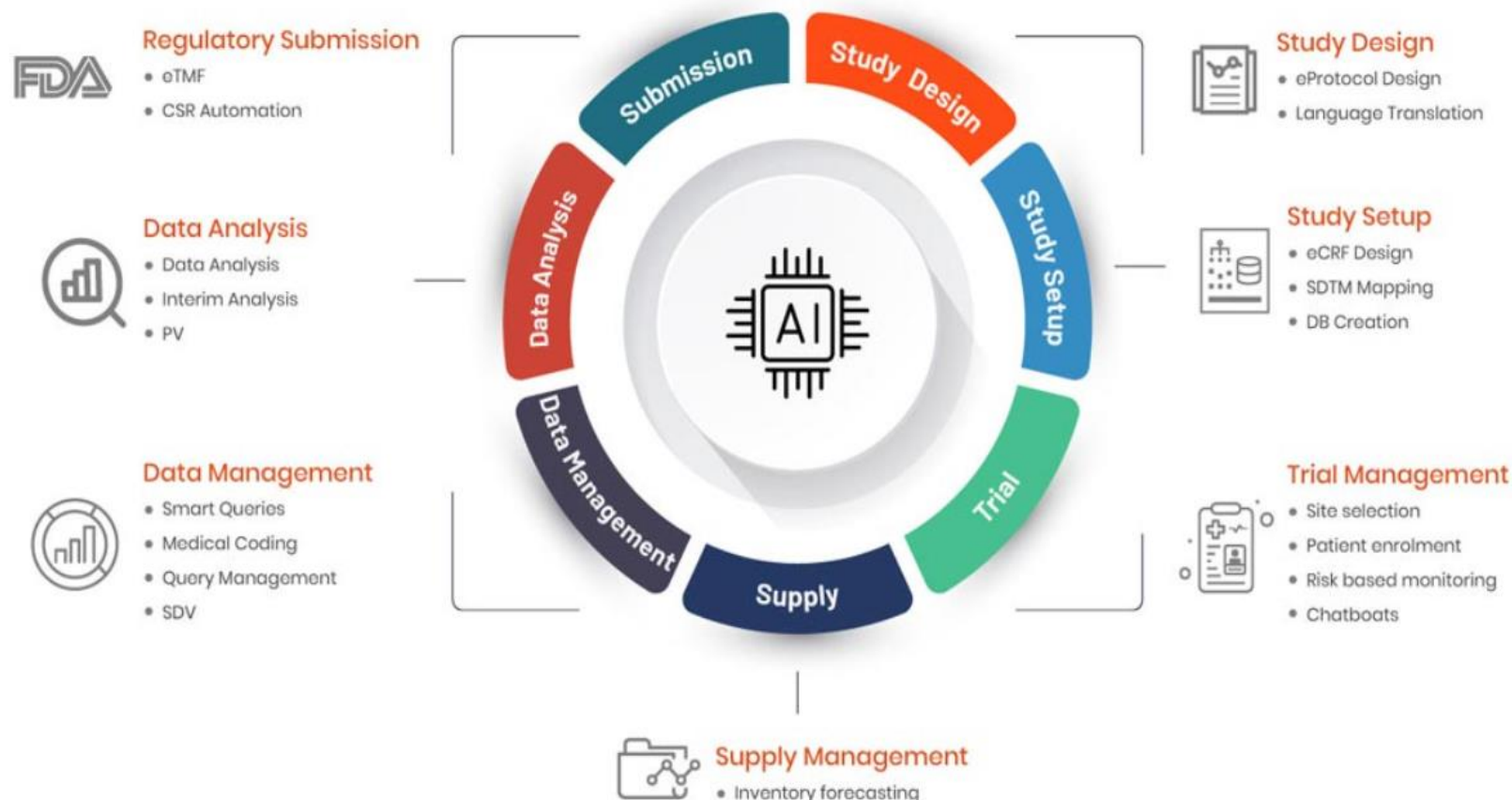


Figure 1: A conceptual diagram illustrating metadata-driven clinical trial lifecycle

NextGen Statistical Programming Evolution in Clinical Trials



AI & ML in Clinical Trials



1.1 Metadata-Driven Frameworks

1.2 Automated QC and Validation

1.3 Modular Code Reuse

1.4 AI/ML Integration

1.5 Scalable Cloud Infrastructure

1.6 Governance and Traceability

1.1 Metadata-Driven Frameworks

- A metadata framework is a structured system for **organizing, categorizing, and managing** metadata across an organization.
- It defines standards, tools, and processes to ensure consistency and usability of data.

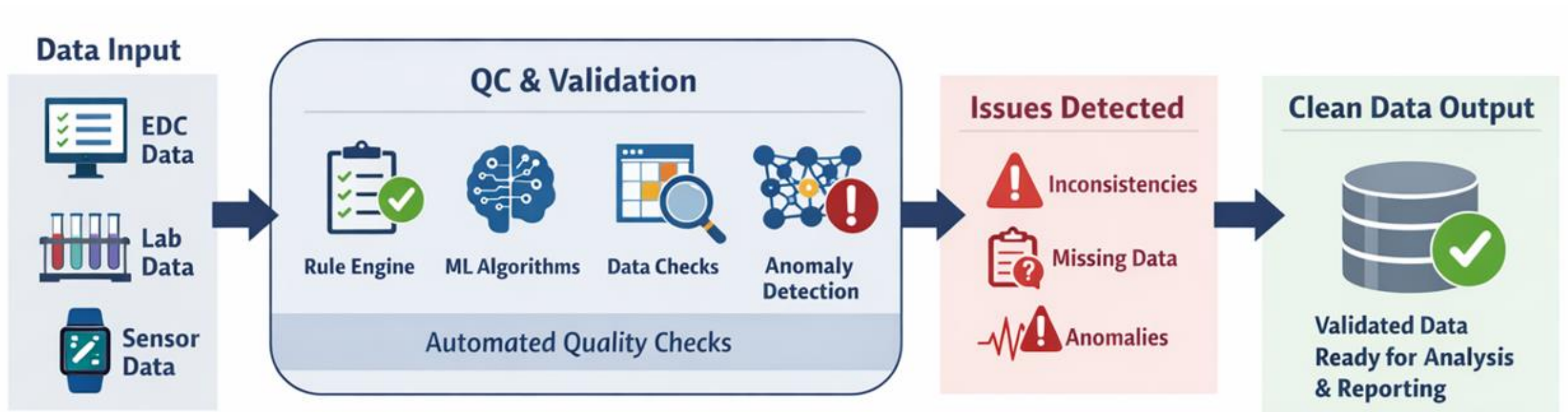
Figure 3: A simplified structure of data and metadata

NAME	AGE	GENDER	HEIGHT (CM)	
A	20	MALE	172	
B	21	MALE	168	
C	19	FEMALE	160	
D	20	MALE	163	

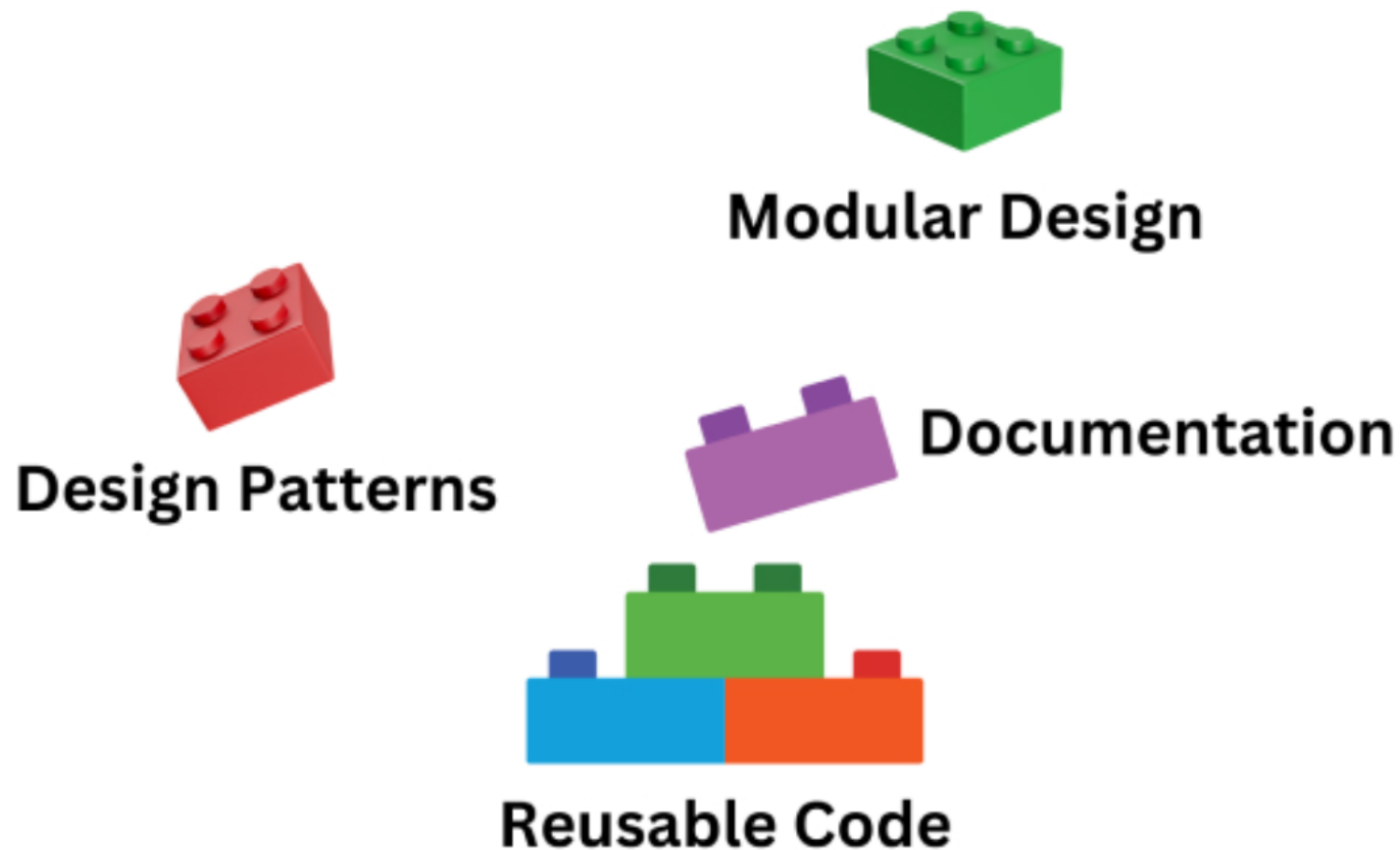
→ METADATA

→ DATA

1.2 Automated QC and Validation



1.3 Modular Code Reuse



1.4 AI/ML Integration

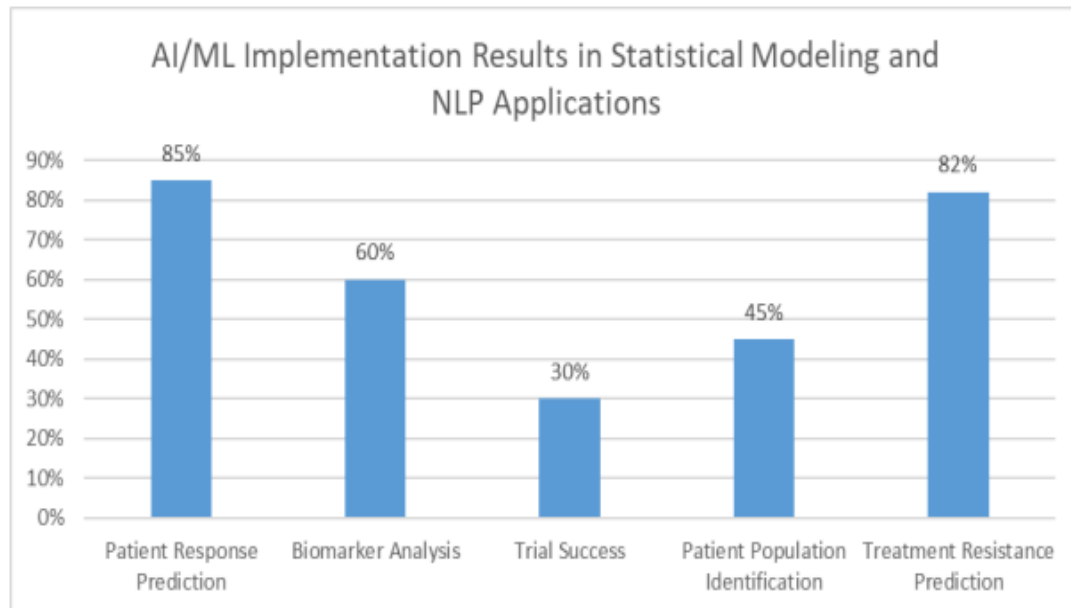


Figure 4: Quantitative Analysis of AI Performance in Clinical Trial Analytics and NLP Applications

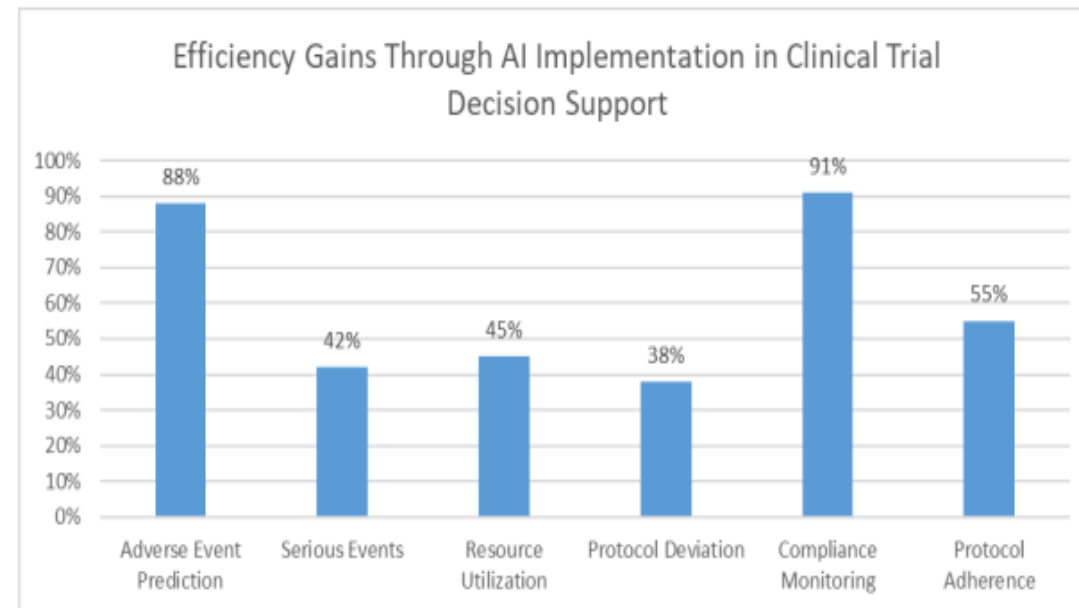
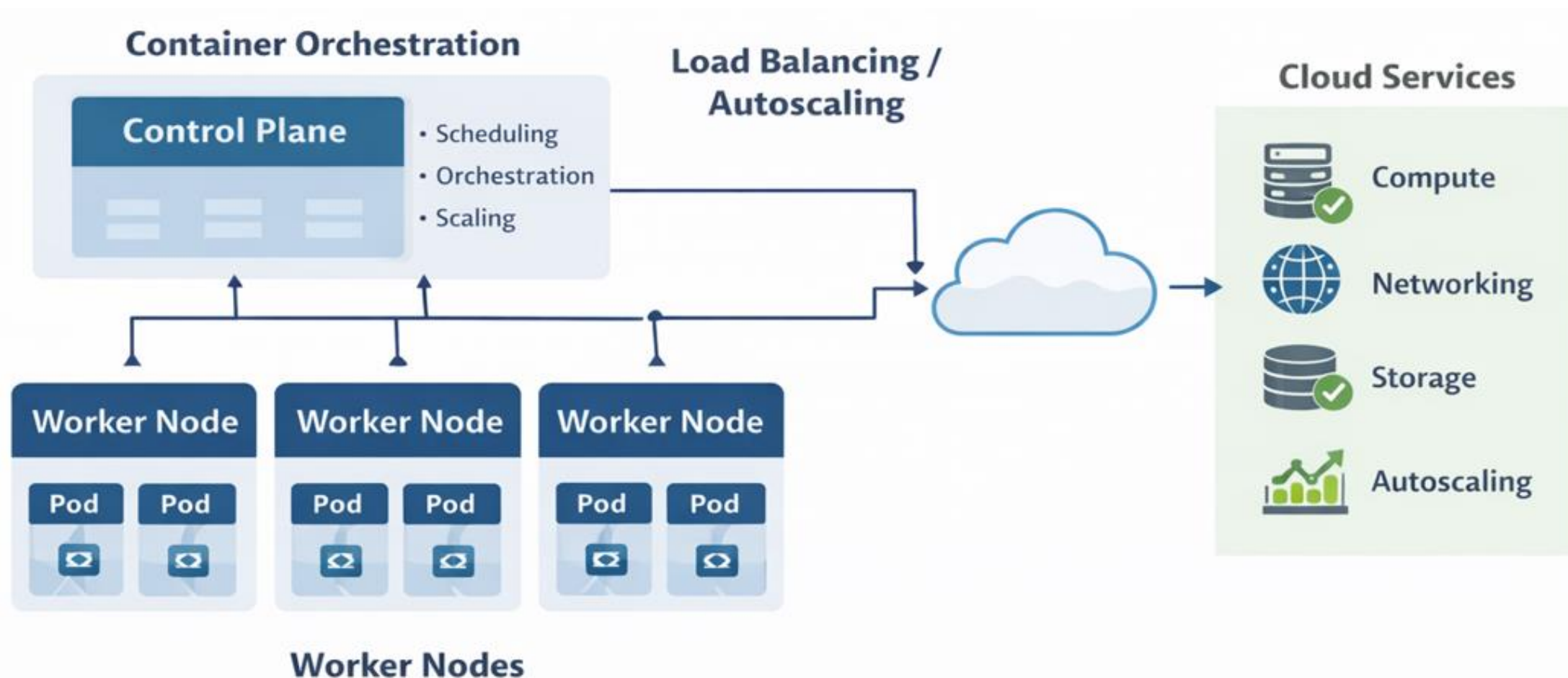
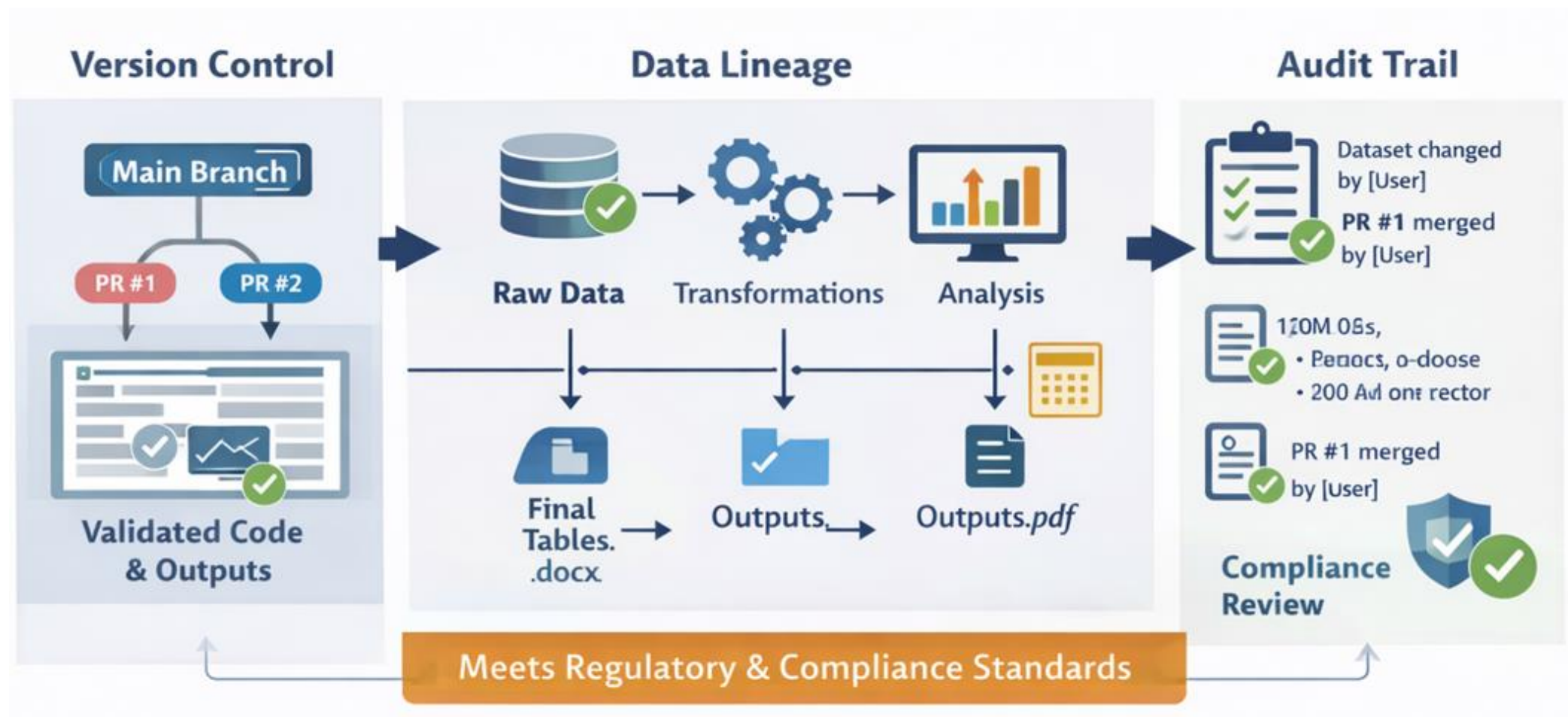


Figure 5: Quantitative Impact Analysis of AI-Driven Real-Time Processing and Adaptive Trial Management

1.5 Scalable Cloud Infrastructure



1.6 Governance and Traceability



Example 1:

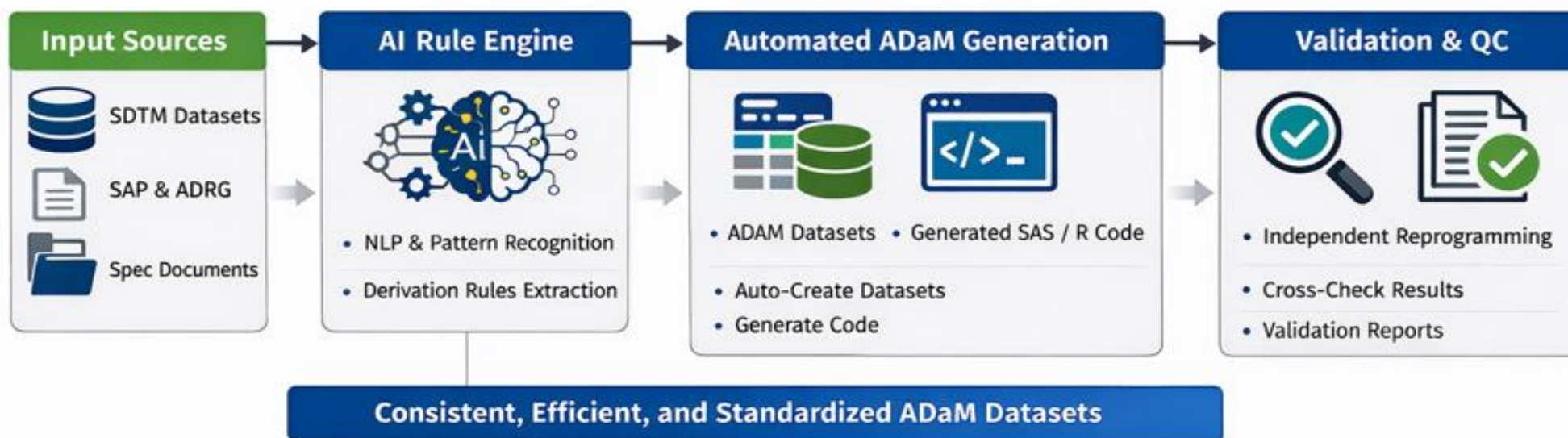
- Automated ADaM Derivation and QC

Example 2:

- AI assisted MAIC (Matching-Adjusted Indirect Comparison) Analysis

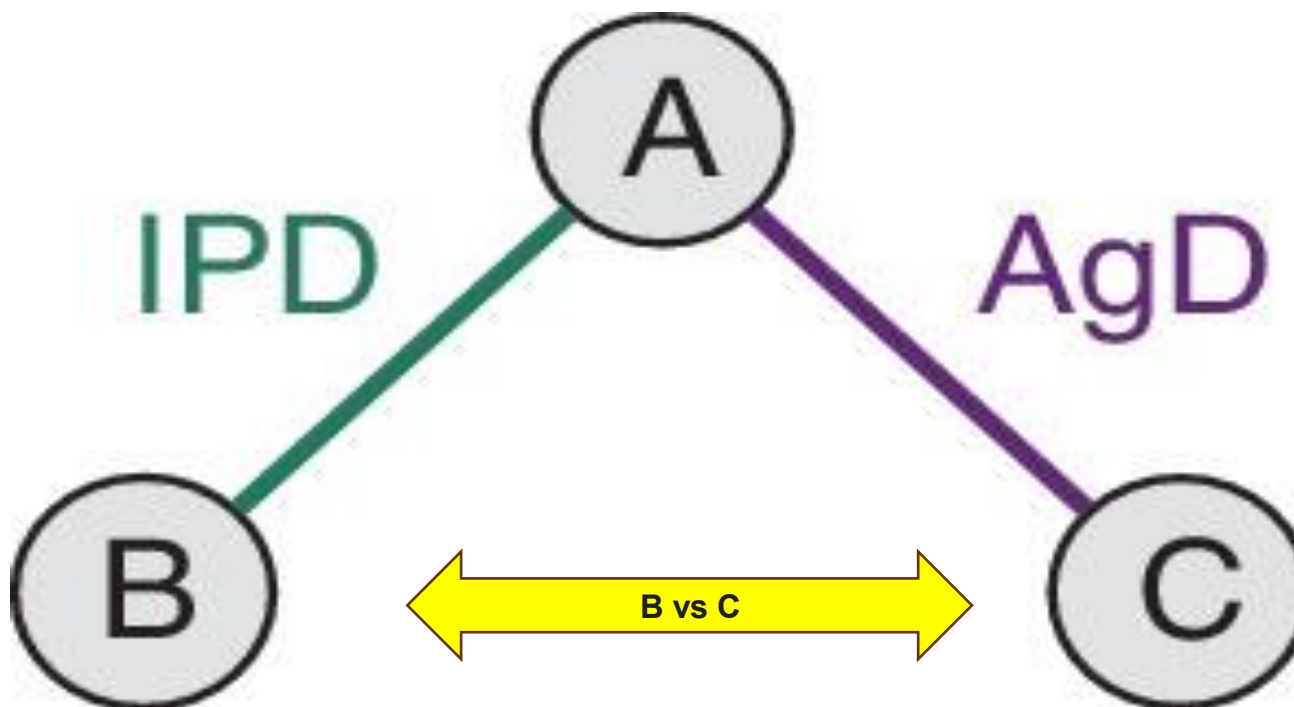
Example 1

Automated ADaM Dataset Generation Using AI Rule Engines



Example 2

AI assisted MAIC (Matching Adjusted Indirect Comparison) Analysis



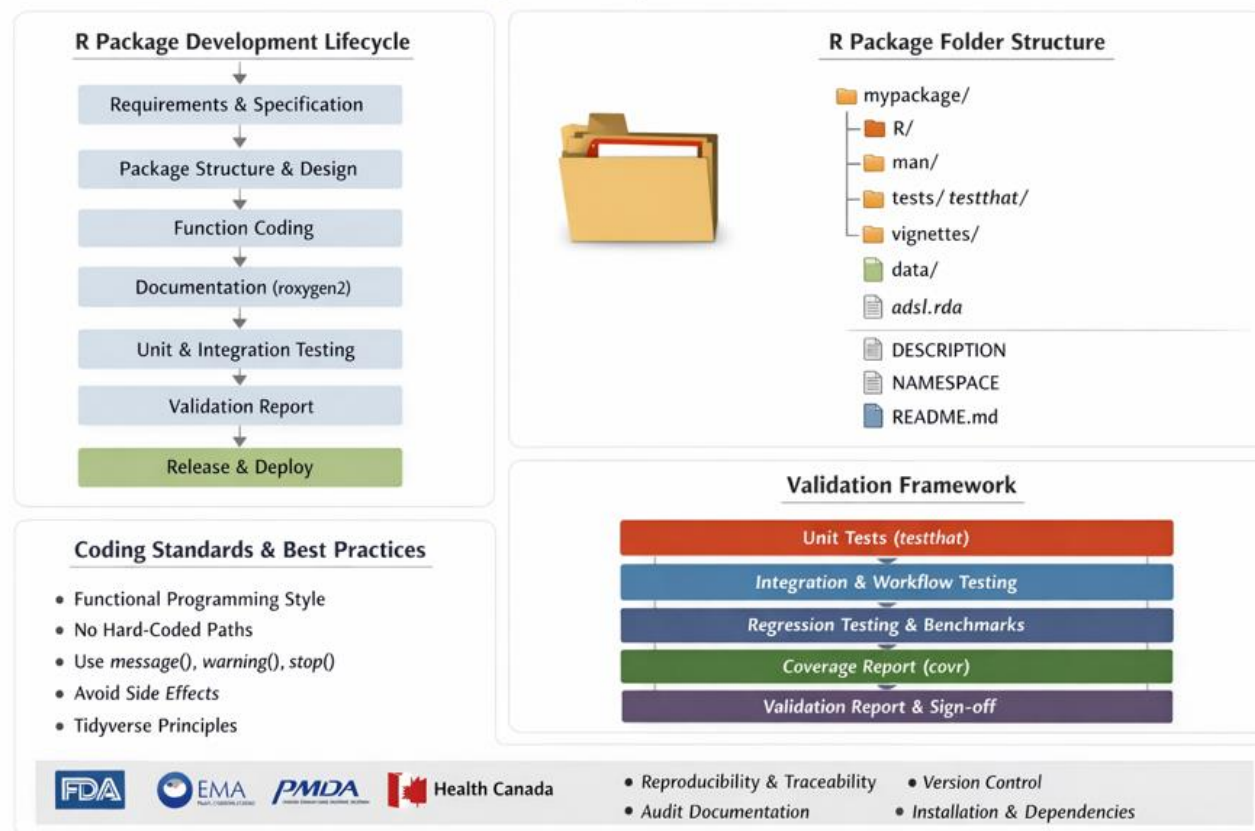
Source: Statistics in Medicine, Volume: 39, Issue: 30, Pages: 4885-4911, First published: 04 October 2020, DOI: (10.1002/sim.8759)

Challenges and Considerations with AI Workflow Automation



- Generative AI for auto-coding and documentation
- Federated learning for privacy-preserving collaboration
- Real-time analytics for adaptive designs
- R-Based Package Development and Validation

R-Based Package Development and Validation



Key Takeaways

The future of clinical trials lies in smarter, reproducible, and integrated statistical programming. Organizations must embrace innovation to remain agile and regulatory-ready. The journey begins with upgrading skills, systems, and mindset.

Next Generation Statistical Programming combines:

1. SAS, R, other open source softwares
2. Python, AI/ML Automation
3. Useful for RWE studies
4. Audit-ready clinical trial outputs
5. Timely Regulatory Submissions and Publications



THANK YOU