

The Journey Beyond Anonymization: Exploring the Secondary Use of Clinical Trial Data

Yuliia Cherednichenko & Tetiana Skorokhod

Agenda

1

Scenarios for Data Reuse

2

Benefits & Challenges

3

Key Stakeholders

4

Data Repositories

5

Standards & Guidelines

6

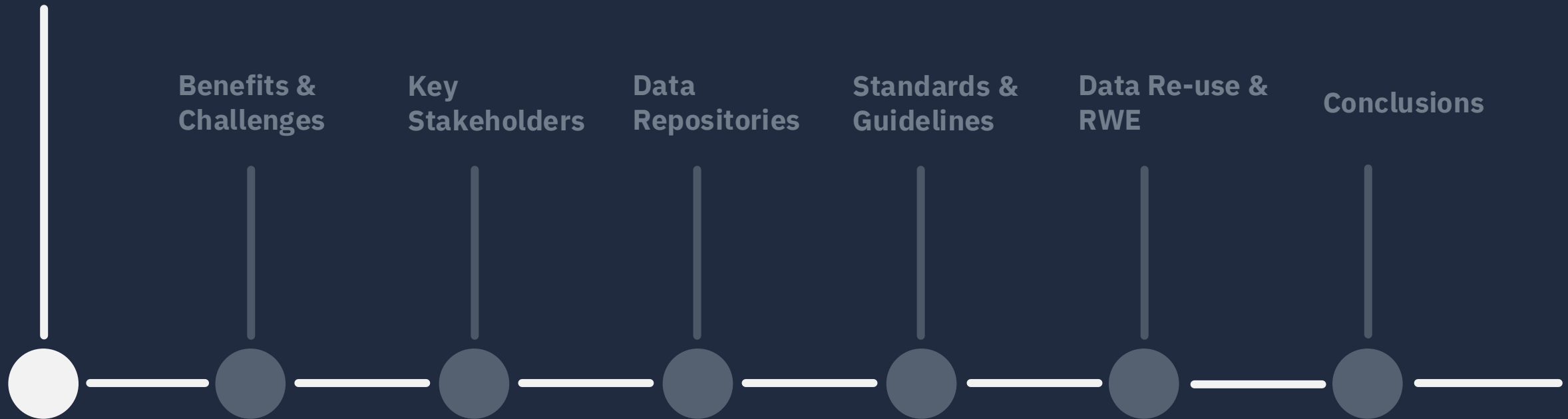
Data Re-use & RWE

7

Conclusions



Scenarios for Data Reuse



Scenario 1:

Evidence Synthesis

Definition:

Brings together evidence to answer a specific research question

Optimal timing of umbilical cord clamping in preterm infants

- **The global collaboration:** over 100 researchers, pooled IPD from more than 60 trials, including over 9,000 preterm infants worldwide
- **Conclusion:** delayed cord clamping significantly reduced mortality compared with immediate clamping
- **Result:** direct informing international clinical guidelines and improving neonatal outcomes



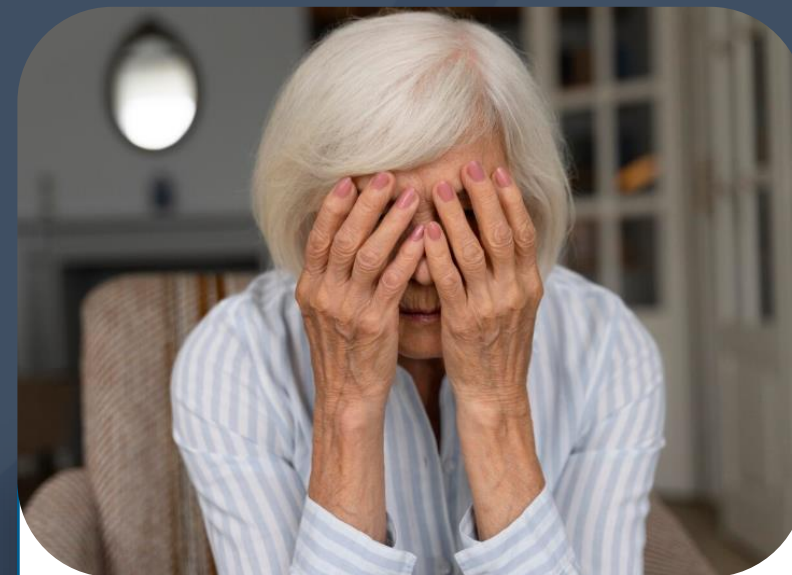
Scenario 1:

Evidence Synthesis



Antimalarial treatment safety in early pregnancy

- An IPD meta-analysis pooling data from 12 cohorts and 34,178 pregnancies
- Proved the safety and effectiveness of antimalarial treatments during early pregnancy
- Informed updates to WHO guidelines



Genetic susceptibility loci for Alzheimer's disease

- A large meta-analysis of genome-wide association studies, combining summary-level data from multiple studies (74,046 individuals)
- 11 novel genetic susceptibility loci for Alzheimer's disease were identified

Scenario 2:

Secondary Analysis

Definition:

The reuse of existing datasets to address research questions that are different from the original studies



Systolic Blood Pressure Intervention Trial (SPRINT) was reused for numerous secondary analyses:

- Subgroup analysis in older adults and patients with chronic kidney disease
- Benefit-risk assessments of intensive vs standard blood pressure control
- The development of clinical prediction tools



Ovarian cancer relapse monitoring

- Pooled secondary analysis, involving over 1,200 patients
- Comparing the tumor marker CA-125 with imaging-based assessments of disease progression
- **Result:** the poor concordance between marker-based and imaging-based detection, leading to changes in surveillance recommendations

Scenario 3:

Reproducibility, Replication and Validation Studies

Definition:

Focus on
confirming, testing,
or contextualizing
previously reported
findings



Reanalysis of GSK Study 329

- The original 2001 publication reported paroxetine as effective and well-tolerated for adolescents with depression
- The **reanalysis found no evidence of efficacy** on prespecified primary outcomes **and identified increased risks of adverse events**, including suicidal ideation



Validation of cardiovascular risk prediction models

- Prediction tools have been repeatedly replicated and externally validated across different countries
- Those validation studies revealed systematic miscalibration in certain populations, leading to model recalibration and the development of region-specific tools

Scenario 4:

Education and Methods Development

Definition:

Using real clinical trials and observational data for educational purposes



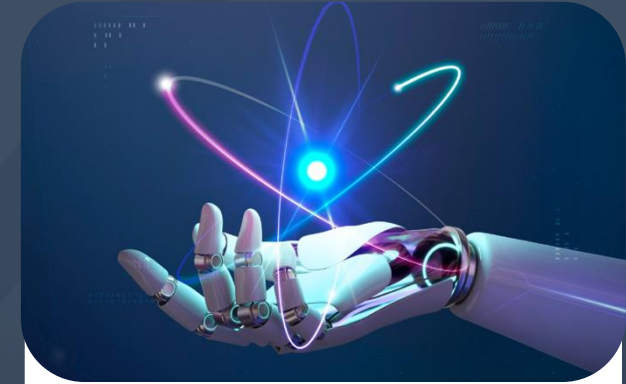
Training using shared clinical trial datasets

- Clinical trials repositories provide access to real data that are widely used in postgraduate education and professional training
- Learners address real-world challenges while avoiding unnecessary new data collection



Methods development for causal inference

- Developing and testing new causal inference methods and approaches
- Supporting methodological innovation without exposing participants to additional risk



Machine learning using health data

Access to shared data allows transparent benchmarking and bias assessment, supporting responsible and reproducible AI development in healthcare

Benefits & Challenges



Benefits of Clinical Trial Data Reuse



Validation and
Replication of
Previously
Published Findings



Novel Inferences
from Existing
Data



Sensitivity
Analysis, Testing
Different Methods
and Assumptions



Deeper Safety
Evaluations



Development of
Patient-Centered
Prediction Models



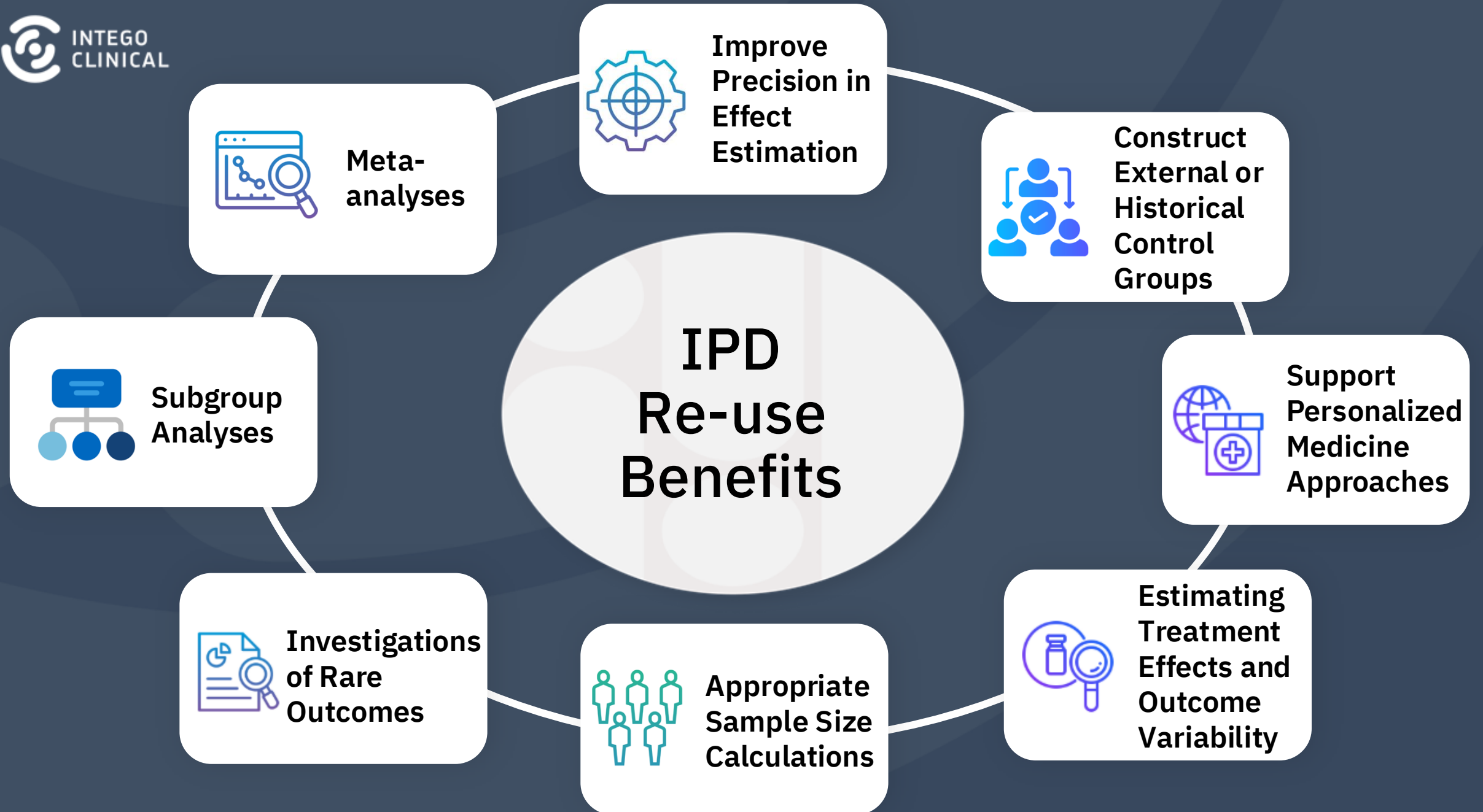
Return on
Investment for
Publicly Funded
Research



Reducing
Unnecessary
Duplication of
Trials



Minimization of
Participant
Burden and
Research Waste



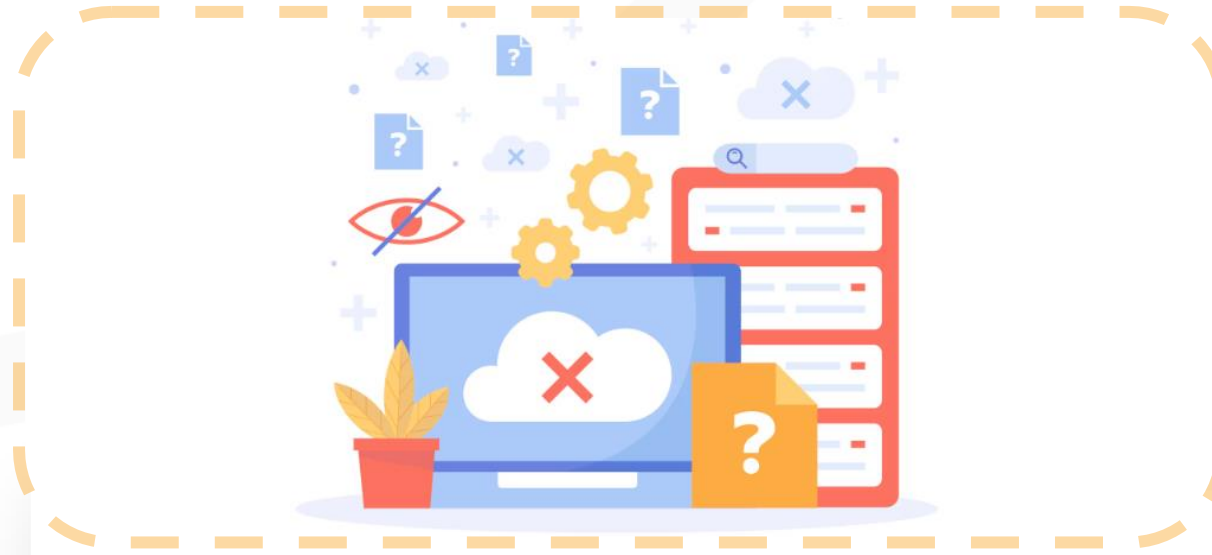
Challenges in Clinical Trial Data Reuse



Data Acquisition Process

- Limited data availability & restricted access
- Required the approval from all authors
- Remote platforms may limit data merging and analytical flexibility
- Varying national regulations on data protection
- Data collected in multiple languages

Challenges in Clinical Trial Data Reuse



Data quality

- High level of heterogeneity
- Shared datasets could be incomplete or not properly cleaned
- Critical metadata is missing or insufficiently detailed
- Lack of interoperability
- Different coding systems, data collection, and conversion standards lead to information loss

Challenges in Clinical Trial Data Reuse



Operational & Technical

- Required components:
 - High capacity repositories
 - Data sharing policies and infrastructure
 - Workforce with the necessary skills and knowledge
- Additional pre-processing is vital for unstructured data
- Advanced statistical techniques are essential to support the appropriate adjustment of pooled data

Challenges in Clinical Trial Data Reuse



Ethical and Incentive-related

- Re-identification risks and patients' distrust
- Potential misinterpretation and misuse of data
- Insufficient incentives for data sharing
- Concerns about loss of academic credit and reputational risks for original investigators
- Commercialization and sale of clinical data
- How to distribute the costs of data sharing?

Key Stakeholders

Scenarios for
Data Reuse

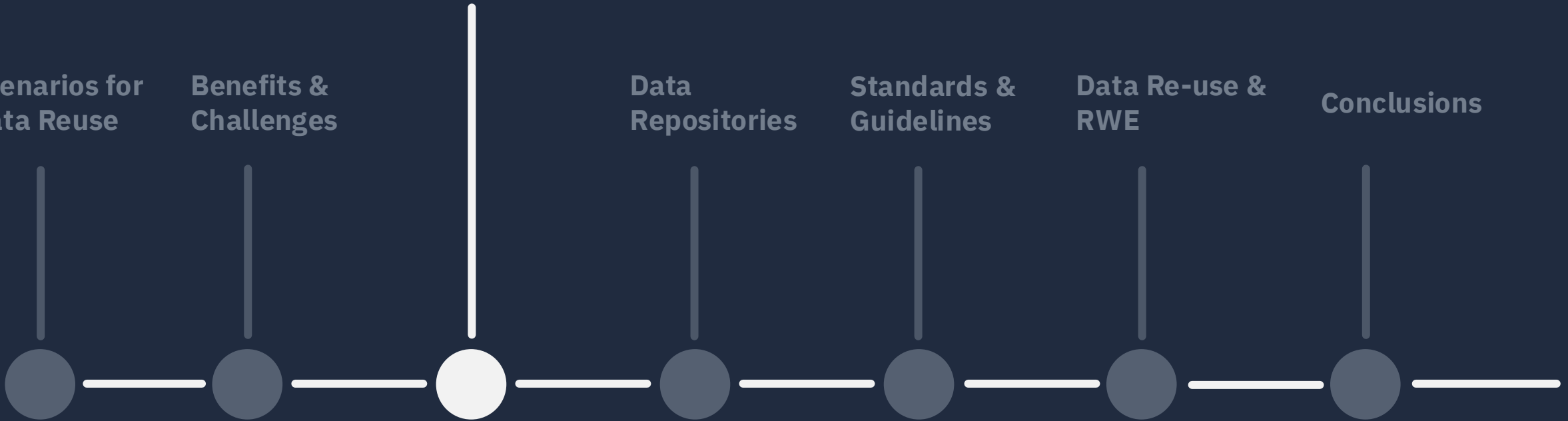
Benefits &
Challenges

Data
Repositories

Standards &
Guidelines

Data Re-use &
RWE

Conclusions



Key Stakeholders

Involved in
Clinical Trial Data
Sharing



**Industry
sponsors**



**Public and
nonprofit
funders/sponsors**

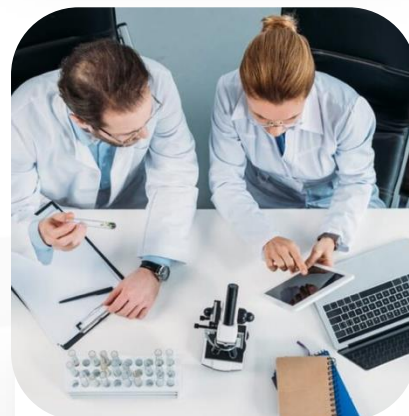


**DMCs & Disease
advocacy
organizations**



EUROPEAN MEDICINES AGENCY
SCIENCE MEDICINES HEALTH

**Regulatory
agencies**



**Clinical trialists &
Secondary users**



**Research
institutions &
universities**



Journals



**Professional
societies**

Data Repositories

Scenarios for
Data Reuse

Benefits &
Challenges

Key
Stakeholders

Standards &
Guidelines

Data Re-use &
RWE

Conclusions



Where can you get the data?

Clinical trials
repositories

Study
Sponsors or
Authors

Public Clinical Trials Registries

ClinicalTrials.gov



Clinical Trials



International Clinical Trials
Registry Platform
Search Portal

Open repositories



kaggle

Controlled-Access Repositories



Standards & Guidelines

Scenarios for
Data Reuse

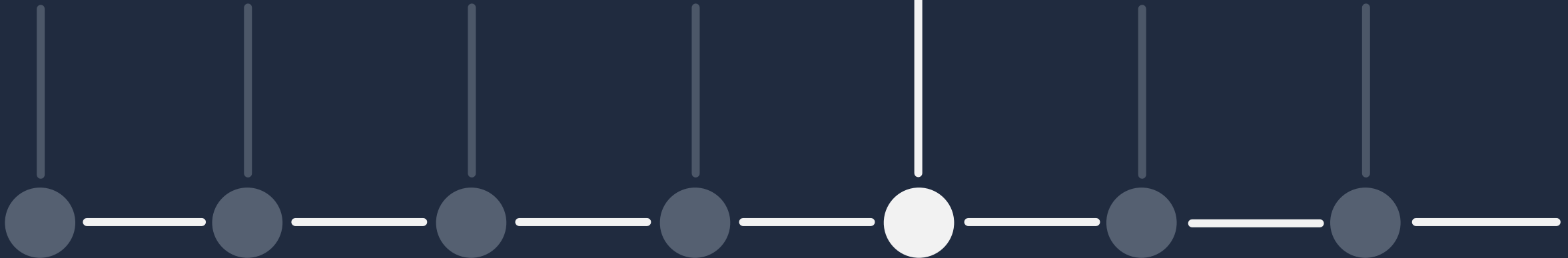
Benefits &
Challenges

Key
Stakeholders

Data
Repositories

Data Re-use &
RWE

Conclusions



General regulations & principles

GDPR General Data Protection Regulation



EU's data privacy law

- Put into effect on May 25, 2018
- Applying globally to any organization processing data of EU residents
- Fines for violating the GDPR are very high

THREE RULES OF HIPAA



HIPAA
Privacy Rule



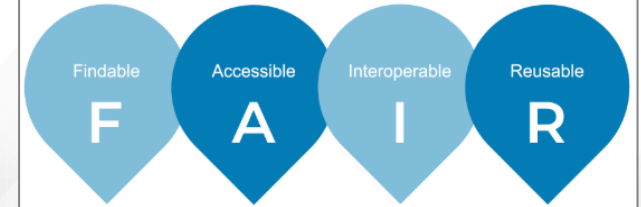
HIPAA
Security Rule



Breach
Notification Rule

Data privacy rule in the USA

- National standards to protect individuals' medical records
- Limiting the use and disclosure of Protected Health Information (PHI) without patient authorization



Guiding principles

- Introduced in 2016
- Have significantly increased the standardization and accessibility of data

Standards & Initiatives

for secondary use
of health data



- Joint action, carried out by 30 European countries
- Ongoing, May 2024 - Dec 2026
- Develops common guidelines and technical specifications



- Data sharing is now a requirement for research funding awarded by WHO
- Policy and practice on the reuse and onward sharing for research purposes



- Multi-stakeholder consortium
- Develops Standards for Clinical Trial Data Reuse
- Builds innovative data governance approaches

NIH HEAL INITIATIVE

- NIH intends to maximize the availability of publications and the sharing of underlying data
- Guidance for Appropriate Public Access and Data Sharing Plans

Data Re-use & RWE

Scenarios for
Data Reuse

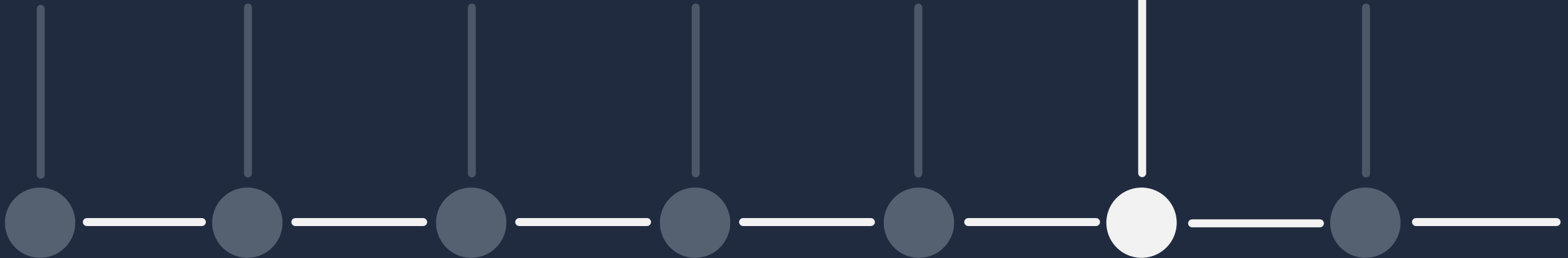
Benefits &
Challenges

Key
Stakeholders

Data
Repositories

Standards &
Guidelines

Conclusions



How can data re-use enhance RWE generation and inform clinical decision-making?



Extending evidence
beyond the constraints of
randomized controlled
trials populations



Supporting post-
authorisation safety
and effectiveness
monitoring



Informing personalised
treatment decisions in
routine practice



Enabling predictive
analytics and clinical
decision support



Training statistical and
machine-learning
models

Conclusions

Scenarios for
Data Reuse

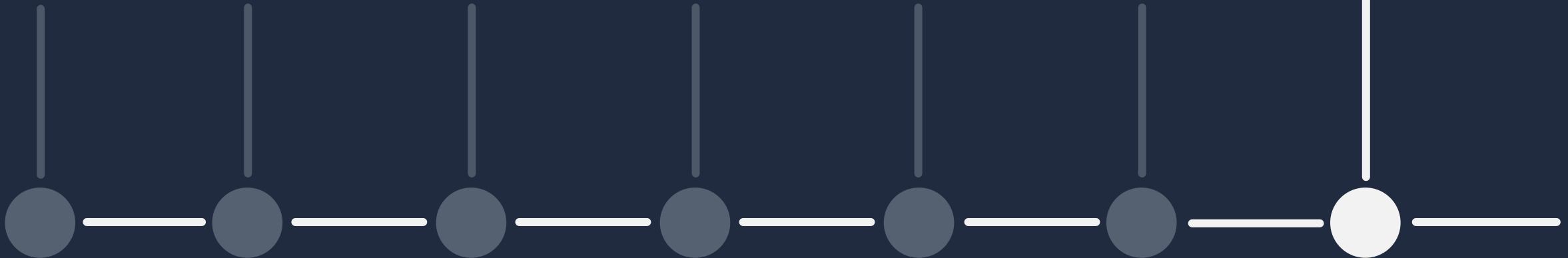
Benefits &
Challenges

Key
Stakeholders

Data
Repositories

Standards &
Guidelines

Data re-use &
RWE



Conclusions



1

The secondary use of clinical trial data holds substantial scientific, ethical, and societal promise

2

Secondary use is constrained by persistent practical, technical, and ethical challenges

3

Investments in interoperable data infrastructures, harmonised standards, privacy-preserving analytics, and data sharing rewards are necessary

4

In the future, the proactive sharing and reuse of data should become the industry standard

Thank you!
Any questions?

