Industry Trial Data: Mountains To Mine For AI Gold

Gregory Goldmacher, MD, PhD, MBA
Executive Director, Head of Clinical Imaging
Merck Research Laboratories

#FDAPDSsymp | #AIinHealth
Outline

● Ecosystem and flow
● Data available
● AI opportunities
● Challenges
The Players

- Pharma companies ("sponsors")
- Hospitals ("sites")
- Independent review facilities ("iCROs")
- Regulators
Data Flow

Site → Protocol → Sponsor → CSR → Regulators

Site → Scans → iCRO → Manual → Sponsor

Site → Scans → BICR → Responses

Sponsor → Responses → Stats Magic → Endpoints

Sponsor → Responses
Analysis: Response Criteria

Endpoints
Date of progression → PFS
Best response → ORR etc…
Read Process

● Quantitative
  ● Choose tumors to measure (baseline)
  ● Outline each on one slice (largest)

● Qualitative
  ● Judgment about “non-target” tumors
  ● Whether/when new tumors have appeared

● Synthesis
  ● Math and logic with human override

● “2+1” adjudication
Data Stored

- **Images**
  - Scans
  - Single slice outlines: every tumor on every scan

- **Assessments**
  - Tumor locations and categories
  - Measurements and qualitative judgments
  - Calculated responses for every scan
Use In Training AI

● Human judgment
  ● Finding tumors
  ● Choosing what to measure
  ● Segmentation
  ● Response

● Non-imaging data
  ● Tissue/molecular
  ● Survival
Example 1 - Learning From Humans

- Manual segmentation is laborious
- 10,000 samples of single slice tumor outlines
- Trained CNN to segment tumors
Example 2 – Training on other data

- Gene expression profiling – inflammation signature
Obstacles To Sharing

● Technical
  ● iCRO holds the images, pharma owns them
  ● Older non-compatible data formats

● Sponsor concerns
  ● Re-analysis of efficacy
  ● New safety questions
  ● Imposed restriction of eligibility
Collaborate To Overcome Challenges

- Design smart datasets
- Trusted third parties
- Standard contracts / agreements
- Can regulators help “de-risk”? 
Thank you!