Data Sharing Consortiums and Large Datasets to Inform Cancer Diagnosis

Improving Cancer Diagnosis and Care: Patient Access to Oncologic Imaging and Pathology Expertise and Technologies: An NCPF Workshop

Amy P. Abernethy, MD, PhD
CMO, CSO & SVP Oncology
Flatiron Health
Data sharing has incredible potential to strengthen academic research, the practice of medicine, and the integrity of the clinical trial system. Some benefits are obvious: when researchers have access to complete data, they can answer new questions, explore different lines of analysis, and more efficiently conduct large-scale analyses across trials. Other advantages, such as providing a guardrail against conflicts of interest in a clinical trial system in which external sponsorship of research is common and necessary, are less visible yet just as critical.
Data Sharing Consortiums

**Historical definition:** “the practice of making data used for scholarly research available to other investigators” (Wikipedia & NIH)

Aggregation of datasets (different variables, to generate critical mass)

Increasing focus on real-world data collected as a routine byproduct of care

Flatiron Health as an example, but there are many others
SEC. 505F. UTILIZING REAL WORLD EVIDENCE.

(a) In General.—The Secretary shall establish a program to evaluate the potential use of real world evidence—

(1) to help to support the approval of a new indication for a drug approved under section 505(c); and

(2) to help to support or satisfy postapproval study requirements.
Diagnosed with Stage II NSCLC

Undergoes surgery for early-stage disease

Develops metastatic disease

Tested for EGFR and ALK

Starts 1L therapy

Starts 2L therapy, deteriorates and is hospitalized

Progresses on 1L, tested for PD-L1 and re-tested for EGFR

Death

Large datasets can be used to inform cancer diagnosis and the interrelationship of diagnosis, treatment, and outcome, but it is important how they get curated.
The Promise of Data Sharing at Scale
As of January 2018

Patients in cohort: 43,697 (Community: 39,915 | Academic: 3,782)

### Histology

- **Non-squamous cell carcinoma**: 69.30%
- **Squamous cell carcinoma**: 24.80%
- **Not otherwise specified**: Unknown / not documented

### Smoking Status

- **History of smoking**: 85.50%
- **No history of smoking**: 12.40%
PDL1 Biomarker Testing

PDL1 Testing Rate Among Actively Treated Patients

PDL1 Status Among Tested Patients

- Positive: 19.5%
- Unknown: 43.2%
- Negative: 37.3%

PDL1 Testing Rate Over Time:

- Jan-16: 8%
- Feb-16: 9%
- Mar-16: 10%
- Apr-16: 11%
- May-16: 11%
- Jun-16: 12%
- Jul-16: 13%
- Aug-16: 14%
- Sep-16: 15%
- Oct-16: 17%
- Nov-16: 20%
- Dec-16: 24%
- Jan-17: 28%
- Feb-17: 32%
- Mar-17: 36%
- Apr-17: 40%
- May-17: 43%
- Jun-17: 47%
- Jul-17: 48%
- Aug-17: 51%
- Sep-17: 53%
- Oct-17: 55%
- Nov-17: 57%
- Dec-17: 58%
Patient Share by Therapy Class — PD1/PDL1
All Lines

Non-platinum-based chemotherapy combinations
Anti-VEGF-based therapies
Clinical study drug-based therapies
Platinum-based chemotherapy combinations
PD-1/PD-L1-based therapies
EGFR-antibody based therapies
Other therapies
ALK inhibitors
Single agent chemotherapies
EGFR TKIs

25% Jan 2016
45% Dec 2017
Patient Share by Therapy Class — PD1/PDL1

2nd or 3rd Line+

Jan 2016: 46%
Dec 2017: 50%
Patient Share by Therapy Class — PD1/PDL1

1st Line

- January 2016: 10%
- December 2017: 43%

© Flatiron Health 2017
Overall survival of PD-1 treated mNSCLC patients

**Findings:** Stratification by line of therapy in which patients received their first PD-1 inhibitor did not reveal significant differences in OS estimates.
PDL1 expression predicts survival

**Findings:** Patients who were PD-1 positive had a significantly longer median survival time (by ~5 months) and higher 1-year survival probability than those who were PD-1 negative.
Hinges on Confident Diagnosis

**Histology**
- Squamous cell carcinoma: 69.30%
- Non-squamous cell carcinoma: 24.80%
- Not otherwise specified: 6.90%

**PDL1 Status**
- Positive: 19.5%
- Unknown: 43.2%
- Negative: 37.3%
Diagnosed with Stage II NSCLC

Undergoes surgery for early-stage disease

Develops metastatic disease

Tested for EGFR and ALK

Starts 1L therapy

Starts 2L therapy, deteriorates and is hospitalized

Progresses on 1L, tested for PD-L1 and re-tested for EGFR

Death

Diagnostic events are a combination of clinical, pathological, radiological, & biomarker data - *in context*
Diagnosed with Stage II NSCLC

Undergoes surgery for early-stage disease

Tested for EGFR and ALK

Develops metastatic disease

Starts 1L therapy

Starts 2L therapy, deteriorates and is hospitalized

Progresses on 1L, tested for PD-L1 and re-tested for EGFR

Death

Gross Description
The specimen is received in formalin labeled with the patient’s name. It consists of a 1 x 0.3 x 0.1 cm aggregate of pink-tan to red-pink soft tissue cores and fragments entirely submitted in one block.

Signed by: GREGORY W SMITH PA
Entered: 02/06/12 - 3544 JAM

Microscopic Description
The specimen consists of a well differentiated adenocarcinoma, favor lung primary. CK7 and TTF are positive. CK5/6 is negative. A colleague agrees with this malignant diagnosis.

Signed by: THOMAS J GRIFFIN MD
Entered: 05/07/12 - 1429 SMU

Diagnosis
Specimen submitted as TURBOT BIOPSY LEFT LUNG MIDDLE:
- WELL DIFFERENTIATED ADENOCARCINOMA, LUNG PRIMARY.
- SEE ABOVE.
Diagnosed with Stage II NSCLC

Undergoes surgery for early-stage disease

Tested for EGFR and ALK

Develops metastatic disease

Starts 1L therapy

Starts 2L therapy, deteriorates and is hospitalized

Progresses on 1L, tested for PD-L1 and re-tested for EGFR

Death
Diagnosed with Stage II NSCLC

Undergoes surgery for early-stage disease

Tested for EGFR and ALK

Develops metastatic disease

Starts 1L therapy

Starts 2L therapy, deteriorates and is hospitalized

Progresses on 1L, tested for PD-L1 and re-tested for EGFR

Time to progression is dependent on when patient is evaluated

Starts 1L therapy
Data Management Challenges

- Need for a common data model
- Scaled curation
- Access to identified source documents
- Differences in timing of assessments
- Data provenance
- Traceability
- Standard policies & procedures to synthesize complex clinical data
- Rigorous quality controls

Here’s what that looks like at Flatiron...
Data source and curation

EHR

- Diagnosis
- Visits
- Demographics
- Labs
- Therapies
- Pathology
- Discharge Notes
- Physician Notes
- Radiology Report

Structured Data Processing
Unstructured Data Processing

Common Database
Data Linkage

Hospital
Reports
Diagnosed with Stage II NSCLC

Starts 1L therapy

Develops metastatic disease

Starts 2L therapy, deteriorates and is hospitalized

Starts 1L therapy

Death

Gross Description
The specimen is received in formalin labeled with the patient’s name. It consists of a 1 x 0.3 x 0.1 cm aggregate of pink-tan to red-pink soft tissue cores and fragments entirely submitted in one block.

Signed by: Gregory V Smith PA
Entered: 02/06/12 - 2564 JAM

Microscopic Description
The specimen consists of a well differentiated adenocarcinoma, favor lung primary. CK7 and TTF are positive. CK5/6 is negative. A colleague agrees with this malignant diagnosis.

Signed by: Thomas J Gifford MD
Entered: 02/09/12 - 14235 ADM

Diagnosis
Specimen submitted as TBCUT 10/28 Left Lung Nodule:
- Well differentiated adenocarcinoma, favor lung primary.
- See above.
Consistent approach to curating unstructured data

### PD-L1 IHC Report Scanned into EHR
Contains Rich Data:

- Test status
- Test result
- Date biopsy collected
- Date biopsy received by laboratory
- Date result received by provider
- Lab name
- Sample type
- Tissue collection site
- Type of test (e.g., FISH)
- Assay / kit (e.g., Dako 22C3)
- Percent staining & staining intensity

... for every test the patient receives
Technology Enabled Abstraction

**Expert abstractors**

A network of abstractors comprised of oncology nurses, certified tumor registrars, and oncology clinical research professionals.

**Flatiron Technology**

Software helps trained human abstractors efficiently organize and review unstructured documents to capture key data elements in predetermined forms.
Diagnosed with Stage II NSCLC

Undergoes surgery for early-stage disease

Tests for EGFR and ALK

Develops metastatic disease

Starts 1L therapy

Starts 2L therapy, deteriorates and is hospitalized

Progresses on 1L, tested for PD-L1 and re-tested for EGFR

Death

Documentation of source, quality and provenance.
Diagnosed with Stage II NSCLC

Undergoes surgery for early-stage disease

Tested for EGFR and ALK

Develops metastatic disease

Starts 1L therapy

Progresses on 1L, tested for PD-L1 and re-tested for EGFR

Starts 2L therapy, deteriorates and is hospitalized

<table>
<thead>
<tr>
<th>Patient</th>
<th>Stage at Dx</th>
<th>Biomarkers</th>
<th>2L Treatment</th>
<th>Progression</th>
<th>Date of Death</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jane Doe</td>
<td>II</td>
<td>EGFR-, ALK-, PD-L1-</td>
<td>nivolumab</td>
<td>2017-03-08</td>
<td>2017-04-12</td>
</tr>
</tbody>
</table>
Diagnosed with Stage II NSCLC

Undergoes surgery for early-stage disease

Tested for EGFR and ALK

Develops metastatic disease

Progresses on 1L, tested for PD-L1 and re-tested for EGFR

Starts 1L therapy

Starts 2L therapy, deteriorates and is hospitalized

Starts 1L therapy

> Abstracted by Sue Smith on 4/31/17 at 10:10am
> Physician notes and scan interpretation reviewed
> Medical record from West Florida Cancer Clinic

Quality of Progression abstraction

> Completeness: 99%
> Sue Smith is 96% accurate at last testing
> Inter-abstractor agreement: 97%
> Kappa: 0.93

> Audit trail for any changes
> Dataset freeze and storage

<table>
<thead>
<tr>
<th>Patient</th>
<th>Stage at Dx</th>
<th>Biomarkers</th>
<th>1L Treatment</th>
<th>Progression</th>
<th>Date of Death</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jane Doe</td>
<td>II</td>
<td>EGFR-, ALK-, PD-L1-</td>
<td>nivolumab</td>
<td>2017-03-08</td>
<td>2017-04-12</td>
</tr>
</tbody>
</table>
Quality assurance & quality control

Centralized Controlled Environment

Upfront
- Feasibility
- Policies & Procedures
- Training & Testing

Ongoing
- Auditing & Monitoring
- Performance Management
- Review Panel

Dataset QA
- Cohort QA
- Data Alignment
- Clinical Assertions
## Resulting clinical data quality and completeness

### Completeness of technology-enabled abstraction

*Example: Advanced NSCLC*

<table>
<thead>
<tr>
<th>Variable</th>
<th>Structured data only</th>
<th>Flatiron data completeness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metastatic diagnosis</td>
<td>26%</td>
<td>100%</td>
</tr>
<tr>
<td>Smoking status</td>
<td>0%&lt;sup&gt;1&lt;/sup&gt;</td>
<td>94%</td>
</tr>
<tr>
<td>Histology</td>
<td>37%</td>
<td>99%&lt;sup&gt;2&lt;/sup&gt;</td>
</tr>
<tr>
<td>Stage</td>
<td>61%</td>
<td>95%</td>
</tr>
<tr>
<td>ALK results (of those tested)</td>
<td>9%</td>
<td>100%&lt;sup&gt;3&lt;/sup&gt;</td>
</tr>
<tr>
<td>EGFR results (of those tested)</td>
<td>11%</td>
<td>99%&lt;sup&gt;3&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

1 58% are free text in dedicated field in EHR (requiring hand abstraction)
2 Including 8% of patients with results pending or unsuccessful test
3 Including 6% of patients with results pending or unsuccessful test

### Accuracy of technology-enabled abstraction

*Example: Sites of metastases*

<table>
<thead>
<tr>
<th>Site of met</th>
<th>Inter-abstractor agreement</th>
<th>Kappa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bone</td>
<td>97%</td>
<td>0.93</td>
</tr>
<tr>
<td>Brain</td>
<td>96%</td>
<td>0.91</td>
</tr>
<tr>
<td>Liver</td>
<td>92%</td>
<td>0.83</td>
</tr>
<tr>
<td>Lung</td>
<td>94%</td>
<td>0.87</td>
</tr>
</tbody>
</table>
## Link Datasets

<table>
<thead>
<tr>
<th>Flatiron</th>
<th>External</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Demographics</td>
<td>- Genomics</td>
</tr>
<tr>
<td>- Diagnosis</td>
<td>- Admin Claims</td>
</tr>
<tr>
<td>- Visits</td>
<td>- Sensors &amp; PROs</td>
</tr>
<tr>
<td>- Therapies</td>
<td>- Mortality</td>
</tr>
<tr>
<td>- Physicians Notes</td>
<td>- Other EHRs</td>
</tr>
<tr>
<td>- Discharge Notes</td>
<td></td>
</tr>
<tr>
<td>- Pathology Reports</td>
<td></td>
</tr>
<tr>
<td>- Radiology Reports</td>
<td></td>
</tr>
<tr>
<td>- Mortality*</td>
<td></td>
</tr>
</tbody>
</table>

- **Core**
- **Linked**
Collaboration with the FDA

- Determine how and when Flatiron data can be used to document the effectiveness and safety of cancer drugs
- Define optimal ways to analyze Flatiron data
- Document endpoints (e.g., evidence of tumor burden or survival) that are clinically meaningful in the Flatiron data
- Best describe the reliability and quality of our data for end-users
Clinical Pharmacology & Therapeutics

Harnessing the Power of Real-World Evidence (RWE): A Checklist to Ensure Regulatory-Grade Data Quality

Rebecca A. Miksad, Amy P. Abernethy

First published: 6 December 2017  Full publication history

DOI: 10.1002/cpt.946 View/save citation

Cited by (CrossRef): 0 articles  Check for updates

Abstract

The role of real-world evidence (RWE) in regulatory, drug development, and healthcare decision-making is rapidly expanding. Recent advances have increased the complexity of cancer care and widened the gap between randomized clinical trial (RCT) results and the evidence needed for real-world clinical decisions.[1] Instead of remaining invisible, data from the >99% of cancer patients treated outside of clinical trials can help fill this void.

DEFINING RWE

RWE is generated from high-quality data that are 1) derived from relevant RWD sources, 2) cleaned, harmonized, and validated to fill in gaps, and 3) include endpoints. Quality control need to encompass the entire process to generate RWE, from data sources and processing to defining appropriate use case.

Figure 1. Open in figure viewer | Download Powerpoint slide

The journey from data to evidence. Real-world data (RWD) are data that are routinely collected in the form of electronic health records (EHRs), patient disease registries, wearables, genomic datasets, medical claims registries, and others. These data can be aggregated, linked, and processed to produce key conclusions in the form of real-world evidence (RWE). The proposed checklist can be used to assess the quality of the RWD is regulatory-grade.
Meta-characteristics of RWD and RWE
Regulatory grade RWE, a potential checklist

- **Clinical Depth**
  Data granularity to enable appropriate interpretation and contextualization of patient information.

- **Completeness**
  Inclusion of both structured and unstructured information supports a thorough understanding of patient clinical experience.

- **Longitudinal Follow-up**
  Ability to review treatment history and track patient journey going forward over time.

- **Quality Monitoring**
  Systematic processes implemented to ensure data accuracy and quality.

- **Timeliness / Recency**
  Timely monitoring of treatment patterns and trends in the market to derive relevant insights.

- **Scalability**
  Efficient processing of information with data model that evolves with standard of care.

- **Generalizability**
  Representativeness of the data cohorts to the broader patient population.

- **Complete Provenance**
  Robust traceability throughout the chain of evidence.
Data Quality Concerns

These will always be there, so, we need to figure out a way to address them
Challenges with Images

Images captured as part of routine care differ in meaningful ways from those captured within clinical trials:

- Raw images are not available for central review
- Imaging reports generated outside of clinical trials often lack info needed for RECIST
- RECIST requires comparison to prior charts. Radiologists interpreting imaging may not have access to prior images.
Accounting for Changing Interpretations Over Time

Impression:
1. Large left upper lobe bronchogenic carcinoma extending to the left hilum and markedly increased in size since prior study.
2. Mediastinal adenopathy increased in size particularly subcarinal space. The adenopathy in the AP window has undergone partial necrosis since previous exam.
3. Stable right apical nodularity possibly scar.
4. Emphysema.

Currently restaging study showed no soft tissue disease. Bone scan showed stable, bony metastasis, with the exception of 1 new lesion in the left superior pubic rami. The nature of this new left superior pubic rami lesion is not clear even though it is possible this is new metastatic lesion; however, it is unusual to have an isolated progression yet rest of the bony metastasis are stable. The patient is completely asymptomatic. His tumor markers are stable; therefore, we decided to continue the current management and we will get followup studies in 3-4 months.
Need a consistent approach to documenting quality

---

### Appendix B: Flatiron Health PD-L1: Inter-rater agreement and kappas on abstracted variable

**Project:**
- **FDA**
- **PD-1 inhibitors in aNSCLC**

<table>
<thead>
<tr>
<th>Kappas scale</th>
<th>0.8 to 1.0</th>
<th>0.6 to 0.8</th>
<th>0.4 to 0.6</th>
<th>0.2 to 0.4</th>
<th>0.0 to 0.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Almost perfect</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Substantial</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Moderate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fair</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Slight</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Note:** For questions where a high percentage of patients have a common answer (e.g., PD-L1 testing status), kappa may be significantly lower than inter-rater agreement. In these cases, it may be more accurate to use inter-rater agreement to measure reliability of the data.

### Table: Enhanced_AdvancedaNSCLC

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description of variable</th>
<th>Corresponding question(s) on abstraction form</th>
<th>Question type</th>
<th>Inter-rater agreement (exact day for dates)</th>
<th>Kappa (exact agreement)</th>
<th>Kappa (30-day window for dates)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DiagnosisDate</td>
<td>Date of diagnosis of initial disease</td>
<td>Enter the date of initial diagnosis</td>
<td>Date</td>
<td>0.795</td>
<td>0.794</td>
<td>0.807</td>
</tr>
<tr>
<td>AdvancedDiagnosisDate</td>
<td>Date of diagnosis of advanced disease: first recurrence or metastasis</td>
<td>Enter the date of the first diagnosis of metastatic or advanced NSCLC</td>
<td>Date</td>
<td>0.655</td>
<td>0.659</td>
<td>0.790</td>
</tr>
<tr>
<td>MetastaticDiagnosisDate</td>
<td>Date of diagnosis of metastatic disease</td>
<td>Enter the date of initial diagnosis (for ~56% of patients in the cohort who are diagnosed metastatic)</td>
<td>Date</td>
<td>0.795</td>
<td>0.794</td>
<td>0.807</td>
</tr>
<tr>
<td>Enter the date of distant metastatic diagnosis (for ~45% of patients in the cohort who are diagnosed non-metastatic)</td>
<td>Date</td>
<td>0.527</td>
<td>0.476</td>
<td>0.557</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Histology</td>
<td>Histology</td>
<td>Select the histology</td>
<td>drop down</td>
<td>0.847</td>
<td>0.894</td>
<td></td>
</tr>
<tr>
<td>GroupStage</td>
<td>Group stage at time of initial diagnosis</td>
<td>Select the group stage</td>
<td>drop down</td>
<td>0.848</td>
<td>0.748</td>
<td></td>
</tr>
<tr>
<td>SmokingStatus</td>
<td>Documented history of smoking</td>
<td>Smoking status</td>
<td>drop down</td>
<td>0.534</td>
<td>0.555</td>
<td></td>
</tr>
<tr>
<td>EgrTested</td>
<td>Indicator of whether the tumor was tested for an EGFR mutation</td>
<td>Was the tumor tested for an EGFR mutation?</td>
<td>boolean</td>
<td>0.327</td>
<td>0.44</td>
<td></td>
</tr>
<tr>
<td>AlkTested</td>
<td>Indicator of whether the tumor was tested for an ALK rearrangement</td>
<td>Was the tumor tested for an ALK rearrangement?</td>
<td>boolean</td>
<td>0.817</td>
<td>0.791</td>
<td></td>
</tr>
<tr>
<td>PolLITested</td>
<td>Indicator of whether the tumor was tested for PD-L1 expression</td>
<td>Was the tumor tested for PD-L1 expression?</td>
<td>boolean</td>
<td>0.817</td>
<td>0.547</td>
<td></td>
</tr>
<tr>
<td>KrasTested</td>
<td>Indicator of whether the tumor was tested for a KRAS mutation</td>
<td>Was the tumor tested for a KRAS mutation?</td>
<td>boolean</td>
<td>0.854</td>
<td>0.736</td>
<td></td>
</tr>
<tr>
<td>RissITested</td>
<td>Indicator of whether the tumor was tested for a ROS1 rearrangement</td>
<td>Was the tumor tested for a ROS1 rearrangement?</td>
<td>boolean</td>
<td>0.817</td>
<td>0.725</td>
<td></td>
</tr>
</tbody>
</table>

© Flatiron Health 2017
## Validation of Endpoints

<table>
<thead>
<tr>
<th>Data Quality &amp; Validation Framework</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Face Validity</strong></td>
</tr>
<tr>
<td>● Oncologist agreement with definition &amp; approach</td>
</tr>
<tr>
<td>● Regulator and other stakeholder agreement with definition &amp; approach</td>
</tr>
<tr>
<td><strong>Feasibility &amp; Quality of Variables (structured &amp; abstracted)</strong></td>
</tr>
<tr>
<td>● Completeness of collected data</td>
</tr>
<tr>
<td>● Inter-rater agreement on progression dates for duplicate abstracted patients</td>
</tr>
<tr>
<td>● Qualitative feedback from abstractors reviewing the medical records</td>
</tr>
<tr>
<td><strong>Validity of Outputs</strong></td>
</tr>
<tr>
<td>● Likelihood of predicting a downstream event (e.g., overall survival)</td>
</tr>
<tr>
<td>● Association between OS and PFS/TTP</td>
</tr>
<tr>
<td>○ Patient-level correlation</td>
</tr>
<tr>
<td>○ Responsiveness of endpoint to treatment effects</td>
</tr>
</tbody>
</table>
Evaluate data against a reference standard

E.g., gold standard = National Death Index
Diagnosed with Stage II NSCLC

Undergoes surgery for early-stage disease

Tested for EGFR and ALK

Progresses on 1L, tested for PD-L1 and re-tested for EGFR

Starts 1L therapy

Starts 2L therapy, deteriorates and is hospitalized

Death

Data in context reinforces diagnostic accuracy

Overall Survival by PD-L1 expression

Log rank p = 0.0005

Number at risk

Survival probability

Survival Summary of 1L treatment and the Median Survival Time

1 Year Survival Probability: 0.47

Survival Summary Probability of surviving for 1 year
Labeled data helps with AI / machine learning

Increasing importance

1. Find labeled data
2. Extract good features
3. Train a model

> 500k patients processed

Clinically informed features

Technical expertise
Problems to be Solved

● Ambiguous interpretation
● Differing issues in scan timing
● Data curation standards with appropriate labeling to support AI/ML
● Data linkage - privacy while maintaining adequate information for context
● Consistent approach to documenting data quality
● Consistent endpoint definitions
● Appropriate use of AI/ML
Policy Context for Data Sharing

● Enabling policy (21st Century Cures, PDUFA VI)

● Privacy, security and governance
  ○ Need approaches to maintain privacy while ensuring adequate contextual information

● Incentives for data sharing

● Regulatory policy that drives standards
  ○ Consistent approach to documenting data quality
  ○ Consistent endpoint definitions
  ○ Incorporating machine learning and AI
Take Home Summary

• Data sharing consortiums offer the opportunity to observe the interrelationship between diagnosis, treatment, and outcomes at scale

• To achieve this we must solve the challenges of data aggregation, curation, and confident assessment of data quality - this can be achieved

• With respect to diagnostic data quality, we must resolve issues related to ambiguous interpretation, differences in scan timing, data curation standards, evaluation of data in context, and the roles of machine learning and artificial intelligence