Information Exchange and Data Transformation (INFORMED) Initiative

Sean Khozin, MD, MPH
Senior Medical Officer
Office of Hematology and Oncology Products (OHOP)
Food and Drug Administration (FDA)

The opinions and information in this document are my own and do not necessarily reflect the views and policies of the FDA
Disclosures

None
About 3 quintillion bytes of data per day
The 4 v’s of #bigdata

**Volume**
- Large repositories

**Velocity**
- Increasing trajectory

**Variety**
- Clinical trials, omics, biometrics, EHRs, unstructured content

**Veracity**
- Uncertainty re: data quality/integrity
Velocity greater than anticipated

Growth of DNA Sequencing

- Recorded growth
- Double every 7 months (Historical growth rate)
- Double every 12 months (Illumina Estimate)
- Double every 18 months (Moore’s Law)

Cumulative Number of Human Genomes

Year

Worldwide Annual Sequencing Capacity

1e+00
1e+03
1e+06
1e+09
1 Zbp
1 Ebp
1 Pbp
1 Tbp

1st Sanger
IHGSC et al.
Venter et al.

1st PacBio
Chaisson et al.

1st Illumina
Bentley et al.
Wang et al.
Ley et al.

1st Personal Genome
Levy et al.

1st 454
Wheeler et al.

1000 Genomes

TCGA

Current Capacity

ExAC

Large amount of diversity

88M variants (84.7M single nucleotide polymorphisms, 3.6M short insertions and deletions, and 60K structural variants)

Leveraging #bigdata requires breaking siloes
Why?
We are nearing the limits of siloed approaches

“Driver” mutations in NSCLC:

<table>
<thead>
<tr>
<th>Mutations</th>
<th>Adenocarcinoma</th>
<th>Squamous-cell carcinoma</th>
</tr>
</thead>
<tbody>
<tr>
<td>EGFR</td>
<td>5-15%*</td>
<td>&lt;5%†</td>
</tr>
<tr>
<td>ALK</td>
<td>5-15%</td>
<td>&lt;5%</td>
</tr>
<tr>
<td>HER2</td>
<td>&lt;5%</td>
<td>0</td>
</tr>
<tr>
<td>BRAF</td>
<td>&lt;5%</td>
<td>0</td>
</tr>
<tr>
<td>KRAS</td>
<td>&gt;15%</td>
<td>&lt;5%</td>
</tr>
<tr>
<td>PIK3CA</td>
<td>&lt;5%</td>
<td>&lt;5%</td>
</tr>
<tr>
<td>AKT1</td>
<td>0</td>
<td>&lt;5%</td>
</tr>
<tr>
<td>MAP2K1</td>
<td>&lt;5%</td>
<td>0</td>
</tr>
<tr>
<td>MET</td>
<td>&lt;5%</td>
<td>&lt;5%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Mutations</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>KRAS</td>
<td>32.2%</td>
</tr>
<tr>
<td>EGFR</td>
<td>11.3%</td>
</tr>
<tr>
<td>NF1</td>
<td>8.3%</td>
</tr>
<tr>
<td>BRAF</td>
<td>7.0%</td>
</tr>
<tr>
<td>MET exon 14 skipping</td>
<td>4.3%</td>
</tr>
<tr>
<td>RIT1</td>
<td>2.2%</td>
</tr>
<tr>
<td>ERBB2</td>
<td>1.7%</td>
</tr>
<tr>
<td>HRAS, NRAS, MAP2K1</td>
<td>1.7%</td>
</tr>
</tbody>
</table>


Clear recognition

Key investments in President’s 2016 budget to launch the Precision Medicine initiative

<table>
<thead>
<tr>
<th>INVESTMENT</th>
<th>AGENCY</th>
<th>OBJECTIVES</th>
</tr>
</thead>
<tbody>
<tr>
<td>$130M</td>
<td>NIH</td>
<td>To develop, in collaboration with other agencies, a voluntary national research cohort of a million or more volunteers. Sources of information will include medical records, environmental and lifestyle data, patient-generated information, and personal device and biometric sensor data.</td>
</tr>
<tr>
<td>$70M</td>
<td>NCI</td>
<td>To scale up efforts to identify genomic drivers in cancer and apply that knowledge in the development of more effective approaches to cancer treatment.</td>
</tr>
<tr>
<td>$10M</td>
<td>FDA</td>
<td>To acquire additional expertise and advance the development of high-quality curated databases to support the regulatory structure needed to advance innovation in precision medicine and protect public health.</td>
</tr>
<tr>
<td>$5M</td>
<td>ONC</td>
<td>To develop interoperability standards and requirements that address privacy and enable secure exchange of data across systems for the voluntary national research cohort initiative.</td>
</tr>
</tbody>
</table>

M indicates million; NCI, National Cancer Institute; NIH, National Institutes of Health; ONC, Office of the National Coordinator for Health.


Khozin S, Blumenthal GM. AJMC Aug 2015
siloed data $\rightarrow$ #bigdata $\rightarrow$ #smartdata

Reductionist
- One-gene one-drug
- Trials with strict eligibility criteria
- Leap of faith clinical development

Holistic
- Pragmatic trials
- Multiomics
- Systems biology
- Predictive analytics
Patient

- Biometrics
- Clinical evaluation (physical exam, psycho-social assessment, etc.)
- Investigations (labs, etc.)
In the near future, #bigdata will simply be called data

Building capabilities and infrastructure to optimize data analysis, enable new data explorations, and serve as a platform for dialogue and stakeholder engagement to advance regulatory science and FDA’s mission of protecting and promoting the public health.
FDA’s strategic priorities for regulatory science

FDA’s regulatory scientists have a unique ability to facilitate development of knowledge and clinical evaluation tools needed for successful translation of discoveries into viable products.

Expansion and improvement of the existing IT infrastructure would enhance and augment these ongoing activities.

FDA will collaborate with others to help develop the new tools and approaches needed to catalyze the development of personalized medicine.

INFORMED: 3 components

1. Transformation of FDA’s existing clinical trial datasets into a common standard;

2. Development of a big data environment for storage and mining of transformed datasets; and

3. Incorporation of diverse pipelines of data (e.g. electronic health records, biometric monitoring devices, unstructured content [e.g. social media], omics) into the big data environment
INFORMED: Framework

Sponsor

Formal submission

Transformation* as needed

Real world & -omics

Direct

Clinical trials

Transformation*

Information Exchange and Data Transformation

I N F O R M E D

Data exchange/visualization/analytics*

Data exported for further analysis if needed

*Technology and software development

Real world data working group

... and others
High-Performance Integrated Virtual Environment (HIVE)
Project examples

Building on previous experience and developing new hypotheses
Efficacy endpoints in non-small cell lung cancer (NSCLC)

Multi-dimensional model to capture tumor kinetics

Depth of tumor response

Pazdur index

- Response (depth, velocity)
- Time: pre-specified landmarks ($t_1$, $t_2$, ...)
- Fidelity: % patients on treatment at $t_x$
- Other (work in progress)

Exploratory pooled analysis of two single arm trials in advanced NSCLC treated with next generation TKI

Patients with >0% decrease in tumor size from baseline based on independent radiology review
Patient- and biometrically-captured experience in oncology

Biometrics/wearables
- Weight
- Mobility
- Sleep pattern
- Heart rate
- Pulse Ox

Mobile app (patient-reported)
- Cough
- Dyspnea
- Chest pain
- Abdominal pain
- Diarrhea
- Fatigue
- Appetite
- Jaundice
- Gas and bloating
- Steatorrhea
“Real world” data

• FDA Adverse Event Reporting System (FAERS)
• Sentinal Initiative
• Medicare Claims
• Patient registries
  – Usually via formal submissions
FDA Adverse Event Reporting System (FAERS)

A database that contains postmarket information on adverse event and medication error reports submitted to FDA

Reports are evaluated by clinical reviewers to monitor the safety of products after they are approved by FDA

If a potential safety concern is identified, further evaluation is performed

May include conducting studies using other large databases, e.g. the Sentinel System
Section 905 of the Food and Drug Administration Amendments Act (FDAAA), which became law in September 2007, mandates FDA to develop an enhanced ability to monitor the safety of drugs after these products reach the market using “active surveillance”

- In May 2008, HHS announced the launch of FDA’s Sentinel Initiative
- Based primarily on billing codes
  - International Classification of Disease codes (ICD)
A. Only those academic institutions with electronic healthcare data will receive safety questions for evaluation.
B. Data partners will provide summary results from analyses conducted within their secure data environments. Those summary results will not include directly identifiable health information.
Challenges with claims-based data

Clinical relevance

Missing data

• Information on disease characteristics (e.g., stage, histology, molecular profile) and clinical outcomes (e.g., toxicity, response to treatment)
Real world data: beyond postmarket surveillance?

• Data on natural history of disease, available therapy definitions

• Incorporation of randomization into systems collecting population health data
  – Cluster-randomized studies
    • Individuals grouped into “clusters”
    • Can be a factorial design: multiple clusters and interventions
    • Inference made to the individual
Challenges
More than just technology

Present difficulty

Technical
- Computation
- Accuracy

Sociopolitical
- Engineering
- Regulatory

Data linkage and sharing

Data standards

Adapted with modification from Kohane IS. Science. Vol. 349 no. 6243 pp. 37-38
Avoid data standards proliferation
Data standards harmonization

Clinical Trials
• Clinical Data Interchange Standards Consortium (CDISC)
• Coalition For Accelerating Standards and Therapies (CFAST)
  – Therapeutic area standards development with input from FDA and other stakeholders
    • Lung, breast, prostate, brain tumors

Real World
• Health Level Seven (HL7)
• Biomedical Research Integrated Domain Group (BRIDG) HL7 Work Group is working on developing a shared information model
• BRIDG includes CDISC, HL7, FDA, and NIH among its stakeholders
## Harmonization

<table>
<thead>
<tr>
<th>Inputs</th>
<th>Outputs</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Clinical trials</strong></td>
<td></td>
</tr>
<tr>
<td>CRF</td>
<td>CDISC</td>
</tr>
<tr>
<td>Genomics</td>
<td>MedDRA</td>
</tr>
<tr>
<td>Laboratory</td>
<td>Legacy</td>
</tr>
<tr>
<td><strong>Biometrics</strong></td>
<td></td>
</tr>
</tbody>
</table>

| **Real World**                 |                               |
| EHR                            | ICD                           |
| PHR                            | CPT                           |
| Genomics                       | NDC                           |
| Laboratory                     | SNOMED                         |
| **Biometrics**                 | LOINC                         |
#bigpicture: product lifecycle

- **Early decisions**
  - Dose, delivery, and manufacturing process optimization

- **Clinical trials**
  - PK/PD
  - Patients
  - Omics
  - Biosensors
  - Outcomes
  - Active surveillance

- **Real world**
  - Patients
  - Omics
  - Outcomes
  - Biosensors
  - Unstructured content

- **Predictive analytics**
  - Validation
The learning health system (IOM)

A system where science, informatics, incentives, and culture are aligned for continuous improvement and innovation

Discovery as a product of the healthcare delivery experience
#thankyou