EHRs, BIG DATA and Explorys driving Research

Anil Jain, MD, FACP
Senior VP & CMIO, Explorys, Inc.
Consulting Staff, Cleveland Clinic

December 14, 2012
Clinical & Translational Science Collaborative
Relevant Disclosures – Anil Jain, MD, FACP

- Salaried officer of Explorys, Inc.
- Deferred compensation, Explorys, Inc.
- Co-sponsor of an Explorys-CTSC pilot grant program
- Explorys has subsidized a subscription to the Explorys Explore application for the CTSC affiliates, MetroHealth and University Hospitals.
Agenda

- Describe how the evolution in health care delivery is fueling BIG DATA
- Discuss the opportunities and challenges of leveraging health care data in supporting clinical research
- Demonstrate how the Explorys platform facilitates hypothesis generation/validation and cohort identification
- Review the governance challenges of leveraging clinical data
About Explorys

Cloud-based healthcare analytics company formed in October 2009 facilitating clinical, operational and financial performance management.

Original technology developed by Dr. Anil Jain in eResearch at Cleveland Clinic for rapid standardization and search of EMR data

Based in Ohio, spin-off of Cleveland Clinic Innovations

Formed by leaders in BIG DATA, security, and healthcare
Explorys “Explore”

Statistics

- 17.9 M patients
- More than 120 hospitals
- Hundreds of practices
- Nearly 100,000 providers
- 85 Billion data points

Explorys Support

- 9 Data Scientists & Researchers
- 3 implementation (medical homes, ACO) consultants
- 3 data integration specialists
Report of the Institute of Medicine, 2001:

- As many as 98,000 people die in hospitals each year due to preventable errors.
- Errors result in total annual cost of $17 billion to $29 billion.
- U.S. health care is poorly organized, overly complex, inefficient and wasteful, operating in silos and problematic for those with chronic illness.
- Requires fundamental sweeping redesign of the entire system
By 2011, not much had changed…

The Commonwealth Fund recently reported:

- U.S. ranks last out of 19 countries on measures of “mortality amenable to health care.”
- This is down from 15th in past five years as other nations raised the bar on performance.
- Up to 101,000 deaths could be prevented each year if U.S. raised standards of care to benchmark levels abroad.
- U.S. spends two times what other nations spend on health, but there’s overwhelming evidence of inappropriate care, missed opportunities, and waste.
Safer Care

Effective Care Coordination

Person-and-Family Centered Care

Prevention and Treatment of Leading Causes of Mortality

Supporting Better Health in Communities

Making Care more Affordable

Are EHRs associated with improved quality?

- Electronic Health Records have shown to improve certain aspects of quality and patient safety and perhaps in some cases to reduce cost.
- If they work, why don’t more doctors use them?

Cebul, Jain, et al. NEJM. September 2011.
Electronic health record use was associated with significantly higher quality of care for four of the measures:
- hemoglobin A1c testing in diabetes,
- breast cancer screening,
- chlamydia screening, and
- colorectal cancer screening.

Effect sizes ranged from 3 to 13 percentage points per measure.

When all nine measures were combined into a composite, EHR use was associated with higher quality of care (sd 0.4, p=0.008).

http://www.springerlink.com/content/p62lv68333315162/fulltext.pdf
Slow EHR Adoption and the “Meaningful Use” Program

- DesRoches et al surveyed 2758 docs in late 2007 and early 2008
  - ~ 4% have extensive, fully functional systems
  - ~ 13% have a basic system

- Jha et al – survey study conducted between 3/08 and 9/08 with AHA acute care hospitals
  - ~ 2% of acute care hospitals have comprehensive electronic-records
  - ~ 8-12% of hospitals have a basic electronic-records

DesRoches et al. NEJM 359 (1): 50
Jha et al. NEJM 360 (16): 1628
Use of Electronic Medical Records

Percent of primary care physicians using electronic medical records

<table>
<thead>
<tr>
<th>Year</th>
<th>United States</th>
<th>International Comparison, 2009</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000</td>
<td>17</td>
<td>17 (NETH)</td>
</tr>
<tr>
<td>2006</td>
<td>28</td>
<td>97 (NZ)</td>
</tr>
<tr>
<td>2009</td>
<td>46</td>
<td>97 (NOR)</td>
</tr>
</tbody>
</table>

Data: Commonwealth Fund International Health Policy Survey of Physicians.
Source: Commonwealth Fund National Scorecard on U.S. Health System Performance, 2011.
HITECH — Stimulus Act of 2009 (ARRA)

EHR Incentive Program requires “Meaningful Use” requirements with measures that are tied to national goals of:

- Improving quality, safety and efficiency
- Engaging patients in their care
- Increasing coordination of care
- Improving the health status of the population
- Ensuring privacy and security
Adoption of Basic vs Any EHR over 10 years

- Use of any EHR increased from 17.0 % in 2002 to 57.0 % in 2011
- The rate of adoption of a basic EHR system increased more quickly among primary care physicians (17.1% to 40.2%) as compared to specialists (12.4% to 30.9%)
- 51.3 percent of all physicians reported that they intended to apply for meaningful use incentives


NCHS survey of ~ 6,000 physicians in 2011
The Affordable Care Act: Four Key Strategies

- Insurance Reform
- Coverage Expansion
- Delivery System Redesign
- Payment Reform

Healthcare Reform
Perfect Storm Driving More and More Data for CER

- 2010 “Affordable Care Act”
  - Value-based care rather than Fee-For-Service requires DATA
  - Transformative delivery models focused on outcomes and coordination requires DATA

- 2009 ARRA/HITECH – “ Meaningful Use” and Data
  - Promoting standardized EMR adoption through incentives
  - Connecting providers and hospitals (HIEs)

- Health 2.0 – Consumer/Patient devices and Personalized Medicine
  - Smart Devices and Patient Home-Based Devices
  - Imaging
  - “omics”
Clinical Research Cycle

- Site Selection
- Protocol Assessment
- Protocol Development
- Research Hypothesis
- Investigator Recruitment
- Subject Recruitment
- Data Collection
- Data Analysis
Explorys “Explore” Current Use Cases

- **Preparatory to Research**
  - Hypothesis Generation and Validation
  - Protocol Feasibility and Development
    - Signal Detection
    - Disease/Condition Profiling
  - Grant Proposals
  - Disease Modeling

- **IRB Approved studies**
  - Pilot Data Collection
  - Retrospective Cohort Studies
  - Prospective Clinical Trials
    - e-Recruitment
    - EHR Data Collection
    - Site and Potential investigator identification
  - Comparative Effectiveness
Research Informatics

Data Collection, Standardization and Validation
Complexity of Typical Relational Databases without Explorys!

Sample tables
Socio-Technical Challenges of Leveraging Data

Data Governance
Privacy & Security related regulatory considerations
Standardization and Harmonization
Data Quality & Validation
Interpretation and Analysis
Sustainability

Performance goals and measures reflect the continuum of care for a patient or a population.

- Preventive,
- Acute,
- Post-Acute

**Aspects of Care:**
- Population Health
- Patient and family engagement
- Safety
- Care Coordination
- Palliative and End-of-life Care
- Overuse

National Quality Forum Integrated Measurement Framework
Informatics Partnerships

Ontologies & Tools

Content & Best Practices

Academics & Research
Harmonizing disparate data by using national standards…

Curation Engine
Searchable Attributes
- SNOMED Diagnosis
- LOINC Observation
- Vital Signs
- Intake
- Implants
- SNOMED Procedure
- ISO Patient Profile*
- Providers
- Encounters

SNOMED/RxNorm
- Pharma Class
- Trade Name
- Ingredients
- Dispensable Product

Diagnoses
- SNOED Diagnosis
- ICD-9
- ICD-10
- Diagnosis & Med History

Procedures
- CPT
- HCP-HCPCS
- Procedure
- Surgical History

Drugs
- NDC
- Drug

Observations / Lab Tests
- Observation or Test
- Encounter Measure

Devices
- Implants
- Prosthetics

Other
- Patient Profile
- Social History

Providers

Encounters

SNOMED

HCPCS

LOINC

Pharma Class

Trade Name

Ingredients

Dispensable Product

Standard Code

Non-Standard Code

Mapped Code
How good is extracted data from EHRs?
Challenges of EHR-based data collection for Disease Registries

- Rand Corporation review of the Quality Assessment Tools system
- Approximately only a third of the clinical indicators for outcomes or baseline characteristics would be readily accessible from EHR data.
- Barriers include:
  - complexity of required data elements
  - provider documentation habits
  - EHR variability
  - Accurately identifying eligible cases for quality assessment

Roth et al, 2009.
Examples of EHR Data Reliability Issues

- Death status is not always reliable in an EHR derived data set or registry.
- Documentation of exclusions are generally not captured in structured data.
- Certain allergies & intolerances (used for exclusions to the numerator) may not be captured.
- Data entry errors that are mentally corrected at POC, e.g., patient 6” rather 6’ tall.
- Documentation of over-the-counter (OTC) medications prescribed/given/taken, etc.
Validating diagnoses by chart review

Table 4. Agreement rates between EHR and chart review for ICD-9 codes related to nephrology diagnosis and selected comorbid conditions

<table>
<thead>
<tr>
<th>Condition</th>
<th>$\kappa$ Statistic (95% CI)</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Positive Predictive Value</th>
<th>Negative Predictive Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diabetic nephropathy</td>
<td>0.75 (0.55 to 0.95)</td>
<td>85%</td>
<td>90%</td>
<td>90%</td>
<td>85%</td>
</tr>
<tr>
<td>Glomerulonephritis</td>
<td>0.85 (0.69 to 1.00)</td>
<td>95%</td>
<td>90%</td>
<td>90%</td>
<td>85%</td>
</tr>
<tr>
<td>Polycystic kidney disease</td>
<td>0.90 (0.77 to 1.00)</td>
<td>100%</td>
<td>90%</td>
<td>90%</td>
<td>100%</td>
</tr>
<tr>
<td>Diabetes</td>
<td>0.90 (0.77 to 1.00)</td>
<td>91%</td>
<td>100%</td>
<td>100%</td>
<td>90%</td>
</tr>
<tr>
<td>Hypertension</td>
<td>0.45 (0.20 to 0.70)</td>
<td>65%</td>
<td>90%</td>
<td>95%</td>
<td>50%</td>
</tr>
<tr>
<td>Hyperlipidemia</td>
<td>0.85 (0.69 to 1.00)</td>
<td>95%</td>
<td>90%</td>
<td>90%</td>
<td>95%</td>
</tr>
<tr>
<td>Coronary artery disease</td>
<td>0.60 (0.35 to 0.85)</td>
<td>80%</td>
<td>80%</td>
<td>80%</td>
<td>80%</td>
</tr>
<tr>
<td>Congestive heart failure</td>
<td>0.85 (0.69 to 1.00)</td>
<td>87%</td>
<td>100%</td>
<td>100%</td>
<td>85%</td>
</tr>
<tr>
<td>Cerebrovascular disease</td>
<td>0.90 (0.76 to 1.00)</td>
<td>95%</td>
<td>95%</td>
<td>95%</td>
<td>95%</td>
</tr>
</tbody>
</table>

# Traditional Criteria for Diabetes from Administrative Data

## Inclusion

<table>
<thead>
<tr>
<th>Description</th>
<th>ICD-9-CM Codes</th>
<th>UB-92 Revenue Codes</th>
<th>CPT Codes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acute inpatient/ED</td>
<td>10X-16X, 20X-22X, 450, 451, 452, 459, 72X, 80X, 981, 987</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

## Exclusion

<table>
<thead>
<tr>
<th>Description</th>
<th>ICD-9-CM Codes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Polycystic Ovaries</td>
<td>256.4</td>
</tr>
<tr>
<td>Steroid Induced</td>
<td>251.8, 962.0</td>
</tr>
<tr>
<td>Gestational Diabetes</td>
<td>648.8</td>
</tr>
</tbody>
</table>
A Bayesian approach by combining clinical data

- In a Marshfield Clinic biobank-based study, identifying diabetics (phenotype) via diabetic diagnostic codes alone produced false positives...
  - 3.0% (in subjects with the highest A1c values)
  - 44.4% (in subjects with the lowest A1c values)

- Investigators can add additional requirements (constraints) to reduce the number of false-positives at the cost of false-negatives.
  - For example, in the same study, investigators combined clinical laboratory data and medication history
  - This “algorithm” yielded a cohort more amenable to the specific research design

How good are traditional research databases?

- Using a “double entry” method, Goldberg and colleagues detected data error rates that ranged from 2.3 to 26.9%.

- A literature review of 42 articles that used the source to database method, which compares case report forms to database entries, found the average error rate was 9.76%.


Data from a PHR or patient portal

- Flow Sheets designed in a patient portal that send data into an EHR must identify such data as having been patient entered and then validated by a provider.

- Patient portals should have high levels of validation to ensure that only acceptable information is entered.

- Patient portal based questions and/or forms should be sensitive and specific to the patient using it. (e.g., don’t ask a diabetic when was their last diabetes screening test)

- Even with best of intentions, the quality may vary.

- Device integration introduces new challenges (who does quality control checks at home – e.g., BP cuff)?
What happens when we start to bring in medical information from the external patient health portals?

What do we do with the results of questionnaires that we administer to our patients?
HIEs: A new source of data and a new source of challenges

- Health Information Exchange (HIE) is mandated in ONC Meaningful Use
- It is unclear how the receiving organization will ensure data quality control from the source system
- Patients will selectively release and block parts of their medical record creating the potential for important gaps
- Variable use and adoption by providers
- When “it” does occur, “it” occurs at the POC – not very suitable for population health
- Consent policies across the U.S. vary and it’s unclear whether quality measures will adapt...
Technical issues of leveraging EMR data

- EMR vendors generally provide reports and access to the data repository to facilitate reporting on *model* systems

- Standard performance measurement reports may or may not consider the *actual* implemented workflows

- Technical understanding and experience with EMR data repository is essential to fully leverage available data

- Variable workflow, multiple implementations create confounders for interpretation especially in large systems

- Non-standard ontology and dynamic and inconsistent data dictionaries can make rapid response and reproducible processes challenging
Potential Solutions to overcome challenges

- Link to external data sources when possible, e.g., social security death index, regional health information exchange, payor data, etc.

- Use chart reviews to validate electronic data points

- Adjust for workflow differences of documentation when examining outcomes rates for internal buy-in by docs

- Identify EHR tools that can be **seamlessly** integrated into clinical workflow and capture structured data

- Consider the use of NLP to augment clinical concepts

- Compare results with published benchmarks!

- Embed performance measurement expertise into the EMR implementation process
Explore BIG DATA and the Value in Clinical Research

- Unlock significant value by securely and safely making patient information usable for hypothesis generation and validation.
- Securely utilize healthcare data with HIPAA/HITECH compliant data governance workflows for data sets and recruitment.
- Lower barriers for investigators to obtain de-identified aggregate pilot data for studies and grant proposals.
- Safely “share” data across organizations for the purposes of conducting larger studies than possible any single organization.
- Data sets for Comparative Effectiveness Research
2012 Recent Explorys BIG DATA/Clinical Research related citations

Manuscripts & Abstracts


• Kaelber D, Foster W, Gilder JR, Love TE, Jain A. “Patient characteristics associated with Venous Thromboembolic Events (VTEs) - A cohort study using pooled Electronic Health Record (EHR) data.” JAMIA. 2012.


• Tyler C. Clinical & Translational Science Collaborative Grant for “A Multi-Level Exploration of Cancer Control Disparities in A Medically Underserved Population.” 2011.

Invited Talks


• Jain A. “Big Data Analytics: Optimizing Clinical Research and Improving Patient Outcomes”, Invited Speaker & Panelist. Informatics: Driving Discovery, Improving Health (IDD 2012), School of Medicine, CWRU, Cleveland, OH, April 6, 2012.


Dr. Anil Jain is Senior VP and Chief Medical Information Officer of Explorys, Inc., a Cleveland-based BIG DATA healthcare analytics company formed in 2009 based on innovations that he developed while at the Cleveland Clinic.

Dr. Jain spent 16 years at the Cleveland Clinic, most recently as Senior Executive Director of IT where he led several Health IT innovations, including programs to support research and quality informatics and created interactive dashboards to monitor the “meaningful use” of the Electronic Health Record.

He has authored more than 100 publications and abstracts and has given numerous talks at national and international meetings on the benefits of Health IT and how BIG DATA analytics can support quality improvement and biomedical research.

He is a Diplomate of the American Board of Internal Medicine (ABIM), a Fellow of the American College of Physicians (ACP), and an active member of both the Health Information Management and Systems Society (HIMSS) and the American Medical Informatics Association (AMIA).

He continues to practice medicine and teach medical residents as Consulting Staff at Cleveland Clinic’s Department of Internal Medicine.
Questions?

Anil Jain, MD, FACP
Senior VP & CMIO, Explorys, Inc.
Anil.Jain@Explorys.com
Research Informatics

eRecruitment (Data-Driven subject recruitment)
Effect of a Clinical Trial Alert System on Physician Participation in Trial Recruitment

Peter J. Embi, MD, MS; Anil Jain, MD; Jeffrey Clark, ES; Susan Bizjak, MSN; Richard Hornung, DrPH; C. Martin Harris, MD, MBA

Background: Failure to recruit a sufficient number of eligible subjects in a timely manner represents a major impediment to the success of clinical trials. Physician participation is vital to trial recruitment but is often limited.

Methods: After 12 months of traditional recruitment to a clinical trial, we activated our electronic health record (EHR)-based clinical trial alert (CTA) system in selected outpatient clinics of a large, US academic health care system. When a patient’s EHR data met selected trial criteria during the subsequent 4-month intervention period, the CTA prompted physician consideration of the patient’s eligibility and facilitated secure messaging to the trial’s coordinator. Subjects were the 114 physicians practicing at selected EHR-equipped clinics throughout our study. We compared differences in the number of physicians participating in recruitment and their recruitment rates before and after CTA activation.

Results: The CTA intervention was associated with significant increases in the number of physicians generating referrals (5 before and 42 after; $P < .001$) and enrollments (5 before and 11 after; $P = .03$), a 10-fold increase in those physicians’ referral rate (5.7/mo before and 59.3/mo after; rate ratio, 10.44; 95% confidence interval, 7.98–13.68; $P < .001$), and a doubling of their enrollment rate (2.9/mo before and 6.0/mo after; rate ratio, 2.06; 95% confidence interval, 1.22–3.40; $P = .007$).

Conclusions: Use of an EHR-based CTA led to significant increases in physicians’ participation in and recruitment rates to an ongoing clinical trial. Given the trend toward the EHR implementation in health care centers engaged in clinical research, this approach may represent a much-needed solution to the common problem of inadequate trial recruitment.

Arch Intern Med. 2005;165:2272-2277

- Point of care alert based on inclusion and exclusion criteria of a clinical trial utilizing the electronic health record’s (EHR) clinical decision support system (CDS)
- Built alerts to recruit for a diabetes clinical trial at Cleveland Clinic
- Method repeated in numerous studies since
EHR based point-of-care recruitment - Results

- **False Positives**
  - Many referrals made for each enrollee – excessive false positives...
  - EHR may not capture key criteria.
  - Chart review may not validate “computable” criteria
  - Patients may not necessarily be good candidates or willing to consent

- **False Negatives**
  - Documentation gap
  - Time lag between presentation and documentation
  - Only patients who have come in for a visit are present in EHRs
EHR-data based recruitment lists work!

- Cleveland Clinic involved in multi-site clinical trial for safety of NIH - H1N1 vaccine among children with severe asthma.

- Comparison of the EHR-identified subjects to volunteers in the Severe Asthma Research Program (SARP) network

- EHR-approach led to accelerated enrollment with 93/540 (17.2%) vs. 24/109 (22%) for SARP.
  - Performance was similar to the volunteer registry without significant increase in costly screen failures
  - Diversity in terms of race/ethnicity of the subjects was increased using EHR-based identification

Increasing participation and diversity... (U Pitt)

- Over a 22-month period, EMR-prompted recruitment:
  - PCPs referred 794 patients via EMR-prompt, 176 (22%) met study inclusion criteria and enrolled,
  - 8,095 patients were approached by wait room-based recruiters of whom 193 (2.4%) enrolled.
  - Subjects enrolled by EMR-prompted PCPs were more
    - non-white (23% vs 5%; \( P < 0.001 \))
    - male (28% vs 18%; \( P = 0.03 \))