How Medical Informatics and HERON Can Help Your Research?

Russ Waitman
Division of Medical Informatics
Department of Biostatistics
November 3, 2011

This project has been supported in part by NIH grant UL1 RR033179-0
Outline for Today’s Presentation

- Core KUMC CTSA Informatics Aim: HERON
- Current Functionality
- Upcoming Clinical Data in HERON
  - Supporting both the CTSA and Cancer Center
- Future Functionality and Integration
  - 1.6 release versus current 1.4: visits and modifiers
- Open time to explore hypotheses
  - Cohort identification for prospective trials
  - Cohorts for observational health services research.
National Clinical and Translational Science Award (CTSA) Objectives:

The purpose of this initiative is to assist institutions to forge a uniquely transformative, novel, and integrative academic home for Clinical and Translational Science that has the consolidated resources to:

1) captivate, advance, and nurture a cadre of well-trained multi- and inter-disciplinary investigators and research teams;

2) create an incubator for innovative research tools and information technologies; and

3) synergize multi-disciplinary and inter-disciplinary clinical and translational research and researchers to catalyze the application of new knowledge and techniques to clinical practice at the front lines of patient care.
Frontiers Biomedical Informatics Aims

1. Provide a portal for investigators to access clinical and translational research resources, track usage and outcomes, and provide informatics consultative services.

2. Create a platform, HERON (Healthcare Enterprise Repository for Ontological Narration), to integrate clinical and biomedical data for translational research.

3. Advance medical innovation by linking biological tissues to clinical phenotype and the pharmacokinetic and pharmacodynamic data generated by research cores in phase I and II clinical trials (addressing T1 translational research).

4. Leverage an active, engaged statewide telemedicine and Health Information Exchange (HIE) effort to enable community based translational research (addressing T2 translational research).
Aim #2: Create a data “fishing” platform

- Develop business agreements, policies, data use agreements and oversight.
- Implement open source NIH funded (i.e. i2b2) initiatives for accessing data.
- Transform data into information using the NLM UMLS Metathesaurus as our vocabulary source.
- Link clinical data sources to enhance their research utility.
Develop business agreements, policies, data use agreements and oversight.

- September 2010 the hospital, clinics and university signed a master data sharing agreement to create the repository.
  - Executive Committee – decides organization/systems expansion
  - Data Request Oversight Committee – guides implementation and approves/monitors use.

- Use Cases:
  - After signing a system access agreement, cohort identification queries and view-only access is allowed but logged and audited
  - Requests for de-identified patient data, while not deemed human subjects research, are reviewed.
  - Identified data requests require approval by the Institutional Review Board prior to data request review.
  - Contact information from the Frontiers Participant Registry have their study request and contact letters reviewed by the Participant and Clinical Interactions Resources Program
Current Functionality

- Single sign-on using your email username
- Real-time check for current human subjects training
- System Access Agreements, Data Use Agreements and Review Processes implemented in HERON with web pages for monitoring system use

- Demonstration
  - i2b2 and HERON tools
HERON De-identification Decisions

- HIPAA Safe Harbor De-identification
  - Remove 18 identifiers and date shifting by 365 days back
  - Resulting in non-human subjects research data but treated as a limited data set from a system access perspective. System users and data recipients agree to treat as a limited data set (acknowledging re-identification risk)

- To be addressed:
  - For now, we won’t add free text such as progress notes with text scrubbers (DeID, MITRE Identification Scrubber toolkit)
  - Currently have “obfuscation” turned on. (turn off shortly)
    - No sets < 10 and sets randomly perturbed + 3 patients
  - While de-identified, access to timeline functionality provides individualized patient “signatures”
Data Sources for FY2012: Focus on Supporting Cancer Center Initiative

- HERON Executive Committee approval June 2011 for incorporating:
  - University Biospecimen Repository (Aim 3, Cancer Center)
  - Hospital Tumor registry (Aim 3, Cancer Center)
  - University REDCap and Velos Registries and Clinical Trials systems (Aim 3, Cancer Center)
  - Hospital billing ICD9, MS-DRG, Insurance Status
  - Social Security Death Master File (Aim 4, Cancer Center)
  - Cerner CoPath pathology system (Aim 3, Cancer Center)

- Also continue to extract and refine data from Epic EMR
Developing a Rich Description of our Population: Existing and Planned Data Sources for HERON. Existing sources shown in **bold underlined text** and planned in plain text.
An i2b2 query against HERON for currently supported cancer centric data sources

Any neoplasm ICD9 diagnosis (106,000 patients) and a WBC count (121,000) -> 44,000 distinct patients,
* require height (123,000) and weight (154,000) -> 35,000 patients,
• require Wong-Baker pain scale (84,000) -> 14468 patients,
• Body Temperature (158,000) -> 14463 patients,
• Surgical Pathology Procedures CPT (85,000) -> 12446 patients,

Finally selective serotonin 5-HT3 antagonist antiemetics -> 8517 patients

With our improved hardware (Fusionio memory cards), the cohort size is returned in 15 seconds for this 8 group query.
Biospecimen Shared Resource Integration

i2b2 Query & Analysis Tool

Navigate Terms

- Heron Cottonwood (Contains July 2011 data)
  - Demographics [15,937,969 facts; 1,888,481 patients]
  - Diagnoses [16,413,722 facts; 651,339 patients]
  - Flowcharts [374,869,347 facts]
  - KUMC BSR Specimens [20,412 facts; 2,247 patients]
    - fluid [12,043 facts; 1,836 patients]
    - blood [12,692 facts; 1,933 patients]
    - other [116 facts; 11 patients]
  - tissue [7,583 facts; 1,522 patients]
    - ? [47 facts]
    - blood, connective tissue, skin, breast [938 facts; 232 patients]
    - digestive systems and peritoneum [1,701 facts; 230 patients]
- GU/GU/Genitourinary [3,733 facts; 932 patients]
  - Bladder [124 facts; 28 patients]
  - Cerebrum [127 facts; 20 patients]
  - Endometrium [116 facts; 56 patients]
  - Fallopian Tube [28 facts; 11 patients]
  - Kidney [956 facts; 85 patients]
  - Lash [<10 facts]
  - Myometrium [<10 facts]
  - Ovary [346 facts; 83 patients]
  - Prostate [2,283 facts; 538 patients]
  - Spermatic Vesiile [<10 facts]
  - Uterus [119 facts; 23 patients]
  - Vaginal Wall [<10 facts]
  - Vulva [<10 facts]
  - Lip, oral cavity, pharynx [301 facts; 45 patients]
  - Respiratory and intrathoracic organs [371 facts; 104 patients]
  - unspecified sites [347 facts; 74 patients]
  - Laboratory Tests [39,974,387 facts; 157,954 patients]
  - Medications [24,366,367 facts; 68,062 patients]
  - Procedures [8,361,830 facts; 511,233 patients]

Find Terms

- Deformities
  - Limb [115 facts; 128 patients]

Query Tool

- Group 2
  - Diagnoses
  - Prostate
- Group 3
  - Malignant neoplasm of prostate
- Group 4
  - Surgical Procedures on the prostate

Query Status

Finished Query: "PRO-Prostate@21:99:14"
Patient Set: 262 Patients - 262 patients
Patient Count: 262 patients
Finish Time: 1.6 seconds
KUH Tumor Registry

- Validated Outcomes and Observations
  - Tumors, Nodes, Metastasis (TNM) on complete cases
  - Untapped investment: 7 cancer registrars (Tim Metcalf)
  - ~65,000 cases, data since 1950s

- North American Association of Central Cancer Registries (NAACCR) file format
  - Will build on work at other NCI designated i2b2 users
    (Group Health Cooperative in Seattle, Kimmel Cancer Center in Philadelphia have shared their code/metadata with us)
  - John Keighley providing invaluable expertise

- Later, supplement with additional treatment information not in NAACCR file
Adding Social Security Death Master File

- Have Death status on approximately 90 million people.
  - Contains Social Security Number, Name, Date of Birth, Date of Death, Place of Death
  - Monthly update file from ntis; will sync with releases
- Initial match on SSN and Date of Birth
  - Should be released with August data this month
  - ~200,000 of our 1.8 million patients are noted as dead in the SS DMF.
Future Functionality: IRB and i2b2 1.6

- Moving beyond counting to line item data review
  - In August, Karen Blackwell Privacy officials agreed to allow timeline access under current system access agreement
  - Need to address some technical issues before release feature
    - Goal: available in October 2011

- i2b2 version 1.5: Visit enabled queries

- i2b2 Modifiers with i2b2 version 1.6
  - Will have to redo ELT to take advantage
Example: Prostate Cancer and PSA tests
What do Visits and Modifiers Offer?

- **Visits:**
  - I want to know the patient had the lab and the medication in the same episode of care.
  - Conceptually, i2b2 has had a table for the visit dimension but the software never exploited the data.

- **Modifiers:**
  - Is it a billing diagnosis or from the problem list? Is it a primary or secondary?
  - How do I represent all parts of a medication order (dose, route, frequency)?
Constrain observations to the same visit

Murphy SN et al, https://www.i2b2.org/events/slides/i2b2_OpeningTalk_20110628_Murphy.pdf
i2b2 Modifiers in the User Interface

Murphy SN et al, https://www.i2b2.org/events/slides/i2b2_OpeningTalk_20110628_Murphy.pdf
i2b2 Modifiers in the User Interface

Murphy SN et al, https://www.i2b2.org/events/slides/i2b2_OpeningTalk_20110628_Murphy.pdf
i2b2 Modifiers in the User Interface

Murphy SN et al, https://www.i2b2.org/events/slides/i2b2_OpeningTalk_20110628_Murphy.pdf
i2b2 Modifiers in the User Interface

Murphy SN et al, https://www.i2b2.org/events/slides/i2b2_OpeningTalk_20110628_Murphy.pdf
Questions and Hypothesis Generation

C.1: HICTR Single Signon Portal
AIM 1
- Search for Funding & Pilots Requests
- CREDO Regulatory & Support
- eIRB
- Biostatistics Services
- Informatics Resources & Consult Service
- Data Request Oversight
- Educational Resources
- Translational Technologies Resource Center
- Clinical & Translational Science Unit (PCIRP)

C.2: Healthcare Enterprise Repository for Ontological Narration
- HERON C2
- IDX, SMS administrative systems

C.3: Advance Innovation T1
AIM 3
- Pharmacokinetics/dynamics results
- Comprehensive Research Information System (CRIS)
- Research Labs
- Biological Tissue Repository
- Clinical Pathology
- K-INBRE Bioinformatics Core
- "wire" hospital systems for T2 research

C.4: KDEP leadership,
AIM 4
- Clinical Trial Support in Rural/Critical Access Hospitals
- Evidence Based Medicine
- Evidence Based Medicine
- ePRISM Modeling 4.6
- Tertiary Care Clinician
- Community Clinician

C.4.1 Kansas & Bistate Health Information Exchanges
- State Agencies Healthcare Policy
- Researcher

C.4.2 Telemedicine & CRIS
- KPHA Medicaid Databases

C.4.3: "wire" hospital systems for T2 research

C.4.4: community HIT as a research platform
C.4.5 Database Resources Integration T2

C.4.6: deploy novel decision support in new environments T2