HERON New 2012 Capabilities: Using RStudio to Explore Diagnoses, Orders, Medications, and Hospital Quality.

Russ Waitman, PhD
Director of Medical Informatics,
Associate Professor, Department of Biostatistics
Director, Frontiers Biomedical Informatics
Assistant Vice Chancellor, Enterprise Analytics
University of Kansas Medical Center
Kansas City, Kansas
and the HERON Team
Bhargav Adagarla, Dan Connolly, Nathan Graham,
Brandon Hamlin, Matthew Hoag, Tamara McMahon,
Mani Nair

This project is supported in part by NIH grant UL1TR000001.
A Fishing Story

How the team works, how is HERON constructed?
- Scaling data aggregation, cohort definition and retrieval across projects

New Capabilities review and demonstration
- Data Sources: I’ll cheat and add some brand new 2013 features!
- Technology + Data Structures

Next Steps and Collaboration Opportunities
- Projects: validate HERON data and cutting edge of clinical data questions
  - Data Structure, Technology, Ontological in support of projects that test the data we load.
- My research interests: hospital/care alignment opportunities
  - \( H_0 = \text{Clinical Data has no impact on understanding healthcare over administrative data} \)
  - \( H_0 = \text{The Electronic Medical Records and Decision Support have no impact on healthcare} \)
  - Example: UHC administrative data augmented by clinical data and preliminary analysis via RStudioServer (data never leaves HERON systems).
- Scaling improved data retrieval and preliminary analysis wrt above
- We’re hiring for an additional Medical Informatics Professor so get the word out on why Kansas City is a great place to be.
Biomedical Informatics Can Help Your Research

- We have tools and expertise to manage data and convert it into information
  - **REDCap** and **CRIS** – enter and manage data
  - **HERON** – *fish for data* from the hospital/clinic
  - **Biweekly** Frontiers Clinical Informatics **Clinics**
    - Tuesday 4-5 pm in 1028 Dykes Library.
    - Next session January 22nd, 2013.
You’re that fisherman: wanting to land data to answer your research hypothesis

Bennett Spring Trout Park, Lebanon Missouri
http://mdc.mo.gov/regions/southwest/bennett-spring
The Fish: Diagnoses, Demographics, Observations, Treatments
Why so many fish?

**Current Goal:** Build Hatchery, Manage the Fishery
Second Goal: If you need help fishing, get a guide

Photo Credit: HuntFishGuide.com
http://www.flickr.com/photos/huntfishguide/5883317106/
Prepare and Analyze Data

Photo Credit: S. Klathill
http://www.flickr.com/photos/sklathill/505464990/
Our shared goal: a tasty publication

Photo Credit: Steve Velo
http://www.flickr.com/photos/juniorvelo/259888572/
I want to go fishing, not fill a fish tank (REDCap)
Use HERON: a managed fishery

Bonneville Hatchery: Trout, Salmon, Sturgeon, Columbia River, Oregon
Central CTSA Informatics Aim: Create a data “fishing” platform: HERON, https://heron.kumc.edu

- **Get a License:** Develop business agreements, policies, data use agreements and oversight.

- **Get a Fishing Rod and Bass Boat:** Implement open source NIH funded (i.e. i2b2 https://www.i2b2.org/) initiatives for accessing data.

- **Know what your catching:** Transform data into information using the NLM UMLS Metathesaurus as our vocabulary source.

- **Stock Different Tasty Fish:** link clinical data sources to enhance their research utility.
• Fill out System Access Agreements to sponsor students/staff
• Fill out Data Use Agreement to request data export
• No Limit!!! IRB Protocol Not Required to view or pull de-identified data
• Must be on campus or use VPN or https://access.kumed.com
• Check http://frontiersresearch.org/frontiers/HERON-Introduction for more information, status, and training videos
The i2b2 “Fishing Rod”: build Diabetes cohort

<table>
<thead>
<tr>
<th>Types of “fish” in folders</th>
</tr>
</thead>
</table>

Drag concepts from upper left into panels on the right.
Dragging over the second condition
When you add a numeric concept, i2b2 asks if you want to set a constraint.
i2b2 Result: 497 patients in Cohort

Run the Query
Query took 4 seconds
497 patient in cohort
### I2b2: Explore Cohort, Visualize

#### Timeline

<table>
<thead>
<tr>
<th>Person</th>
<th>Status</th>
<th>Data</th>
<th>Observations</th>
<th>Measurements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Person #1774</td>
<td>f</td>
<td>Non-hispanic</td>
<td>08-031070 BMI (Calculated)</td>
<td>288,218 facts; 80,395 patients</td>
</tr>
<tr>
<td>Diabetes mellitus</td>
<td>376,016 facts; 31,703 patients</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frontiers Research Participant Registry</td>
<td>7,854 facts; 7,854 patients</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GLUCOSE (2011)</td>
<td>1,007,668 facts; 118,524 patients</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HEMOGLOBIN A1C (2003)</td>
<td>80,793 facts; 35,269 patients</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Person #1809</td>
<td>f</td>
<td>Non-hispanic</td>
<td>09-031070 BMI (Calculated)</td>
<td>269,216 facts; 80,395 patients</td>
</tr>
<tr>
<td>Diabetes mellitus</td>
<td>376,018 facts; 31,703 patients</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frontiers Research Participant Registry</td>
<td>7,854 facts; 7,854 patients</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GLUCOSE (2010)</td>
<td>872,692 facts; 100,192 patients</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HEMOGLOBIN A1C (2003)</td>
<td>80,793 facts; 35,269 patients</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Person #21253</td>
<td>f</td>
<td>Non-hispanic</td>
<td>03-031070 BMI (Calculated)</td>
<td>269,216 facts; 80,395 patients</td>
</tr>
<tr>
<td>Diabetes mellitus</td>
<td>376,018 facts; 31,703 patients</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frontiers Research Participant Registry</td>
<td>7,854 facts; 7,854 patients</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GLUCOSE (2011)</td>
<td>1,007,668 facts; 118,524 patients</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HEMOGLOBIN A1C (2003)</td>
<td>80,793 facts; 35,269 patients</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Person #13473</td>
<td>f</td>
<td>Non-hispanic</td>
<td>09-031070 BMI (Calculated)</td>
<td>269,216 facts; 80,395 patients</td>
</tr>
<tr>
<td>Diabetes mellitus</td>
<td>376,018 facts; 31,703 patients</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frontiers Research Participant Registry</td>
<td>7,854 facts; 7,854 patients</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GLUCOSE (2010)</td>
<td>872,692 facts; 100,192 patients</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HEMOGLOBIN A1C (2003)</td>
<td>80,793 facts; 35,269 patients</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Person #5138</td>
<td>m</td>
<td>Non-hispanic</td>
<td>09-031070 BMI (Calculated)</td>
<td>200,216 facts; 80,395 patients</td>
</tr>
<tr>
<td>Diabetes mellitus</td>
<td>376,016 facts; 31,703 patients</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frontiers Research Participant Registry</td>
<td>7,854 facts; 7,854 patients</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GLUCOSE (2010)</td>
<td>872,692 facts; 100,192 patients</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HEMOGLOBIN A1C (2003)</td>
<td>80,793 facts; 35,269 patients</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Person #25806</td>
<td>f</td>
<td>Non-hispanic</td>
<td>03-031070 BMI (Calculated)</td>
<td>269,216 facts; 80,395 patients</td>
</tr>
<tr>
<td>Diabetes mellitus</td>
<td>376,018 facts; 31,703 patients</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frontiers Research Participant Registry</td>
<td>7,854 facts; 7,854 patients</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GLUCOSE (2010)</td>
<td>872,692 facts; 100,192 patients</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HEMOGLOBIN A1C (2003)</td>
<td>80,793 facts; 35,269 patients</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The dream: landing the big one

http://www.oregon.com/columbia_gorge_attractions/bonneville_hatchery

Catch the data for JAMA, NEJM publication
Without getting bit

!!CAUTION!!

TROUT BITE

DO NOT PUT YOUR HANDS
OR FINGERS IN THE WATER
• Goal: stable monthly process, minimal downtime
  • Complete rebuild of the repository, not HL7 messaging update based.
  • Two databases: create new DB while old DB is in use.
  • When the new DB is ready, switch over i2b2 to serve customers fresh data.

• Initial Files from Clinical Organizations
  • Export KUH Epic Clarity relational database instead of Cache/MUMPS.
  • Monthly file from UKP clinic billing system (GE IDX).
    • Demographics, services, diagnoses, procedures, and Frontiers research participant flag.

• Extract Transform Load (ELT) processes largely SQL (some Oracle PL/SQL)
  • Wrapped in python scripts.

• Goals for a **monthly release** (19 months in a row so far):
  – Fresh data. Example: another month of visits = millions of facts
  – New types of data. Example: family history
  – New functionality: Example: link data by encounter across clinical and financial sources; distinguish medication administration from prescription
Development Process Example: Last milestone

Milestone heron-council-grove-update

Due in 6 hours (01/17/13 18:00:00)

Number of tickets: closed: 23  active: 7  Total: 30  77%

Release Planning for Council-Grove.

On track:

- family and past medical history (#1515)
- ... see #1698 for a list of the goodies we're delivering in this release.

Coordination tasks from from HeronProjectTimeline#Sep2012Planning:

- pick #AHIA paper topics

Postponed to milestone:heron-neosho-update:

- 10% test data (#1385)
- Chronicles access (#20)

Note: See TracRoadmap for help on using the roadmap.
### Milestone equals tickets

<table>
<thead>
<tr>
<th>Ticket</th>
<th>Summary</th>
<th>Owner</th>
<th>Type</th>
<th>Priority</th>
<th>Component</th>
<th>Version</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1369</td>
<td>HERON timelines are really slow since the cedar_bluff upgrade</td>
<td>ngraham</td>
<td>problem</td>
<td>major</td>
<td>data-repository</td>
<td></td>
</tr>
</tbody>
</table>

### Status: closed

<table>
<thead>
<tr>
<th>Ticket</th>
<th>Summary</th>
<th>Owner</th>
<th>Type</th>
<th>Priority</th>
<th>Component</th>
<th>Version</th>
</tr>
</thead>
<tbody>
<tr>
<td>#478</td>
<td>Query for encounter vitals in HERON</td>
<td>badaqarl</td>
<td>enhancement</td>
<td>major</td>
<td>data-repository</td>
<td></td>
</tr>
<tr>
<td>#986</td>
<td>some flowsheet observations are not connected to the relevant encounter</td>
<td>badaqarl</td>
<td>defect</td>
<td>major</td>
<td>data-repository</td>
<td></td>
</tr>
<tr>
<td>#1482</td>
<td>Identified request - check that BSR is not in the query</td>
<td>tscmahn</td>
<td>task</td>
<td>major</td>
<td>data-repository</td>
<td></td>
</tr>
<tr>
<td>#1478</td>
<td>Query multiple REDCap projects with multiple events from HERON</td>
<td>badaqarl</td>
<td>enhancement</td>
<td>major</td>
<td>data-repository</td>
<td></td>
</tr>
<tr>
<td>#1515</td>
<td>Query past medical and family history from Epic for ambulatory and inpatient</td>
<td>mhoa</td>
<td>enhancement</td>
<td>major</td>
<td>data-repository</td>
<td></td>
</tr>
<tr>
<td>#1516</td>
<td>query smart data note concepts from the ambulatory environment</td>
<td>mmair</td>
<td>enhancement</td>
<td>major</td>
<td>data-repository</td>
<td></td>
</tr>
<tr>
<td>#1614</td>
<td>Student Access off campus to HERON</td>
<td>tscmahn</td>
<td>enhancement</td>
<td>major</td>
<td>data-repository</td>
<td></td>
</tr>
<tr>
<td>#1650</td>
<td>hipaa-adjusted birthdates not re-adjusted for death date from social security DMF</td>
<td>ngraham</td>
<td>defect</td>
<td>major</td>
<td>data-repository</td>
<td></td>
</tr>
<tr>
<td>#1666</td>
<td>Create hierarchy for family history</td>
<td>tscmahn</td>
<td>task</td>
<td>major</td>
<td>data-repository</td>
<td></td>
</tr>
<tr>
<td>#1668</td>
<td>REDCap production data for BRCA registry has weird data</td>
<td>badaqarl</td>
<td>design-issue</td>
<td>major</td>
<td>data-repository</td>
<td></td>
</tr>
<tr>
<td>#1669</td>
<td>Age at visit incorrectly set to age now</td>
<td>ngraham</td>
<td>defect</td>
<td>major</td>
<td>data-repository</td>
<td></td>
</tr>
<tr>
<td>#1688</td>
<td>Update medication tests for new modifier hierarchy</td>
<td>ngraham</td>
<td>task</td>
<td>major</td>
<td>data-repository</td>
<td></td>
</tr>
<tr>
<td>#1693</td>
<td>Survival Plugins give error pop-up due to version-skew between client and gate back-end</td>
<td>dconnolly</td>
<td>problem</td>
<td>major</td>
<td>data-repository</td>
<td></td>
</tr>
<tr>
<td>#1693</td>
<td>Stage December data for HERON Council-Grove Update</td>
<td>bhamlin</td>
<td>task</td>
<td>major</td>
<td>data-repository</td>
<td></td>
</tr>
<tr>
<td>#1695</td>
<td>UHC erroneously records &quot;discharge status&quot; with the start of the case not end</td>
<td>ngraham</td>
<td>defect</td>
<td>major</td>
<td>data-repository</td>
<td></td>
</tr>
<tr>
<td>#1696</td>
<td>Run ETL for HERON Council-Grove Update</td>
<td>ngraham</td>
<td>task</td>
<td>major</td>
<td>data-repository</td>
<td></td>
</tr>
<tr>
<td>#1697</td>
<td>HERON Council-Grove: Prepare new data for testing</td>
<td>ngraham</td>
<td>task</td>
<td>major</td>
<td>data-repository</td>
<td></td>
</tr>
<tr>
<td>#1702</td>
<td>Identified DB server: syslog flooding central logging due to impdp log</td>
<td>dconnolly</td>
<td>problem</td>
<td>major</td>
<td>data-repository</td>
<td></td>
</tr>
<tr>
<td>#1740</td>
<td>Don’t drop tree_crd table in concept_stats since it’s needed for data request code</td>
<td>ngraham</td>
<td>task</td>
<td>major</td>
<td>data-repository</td>
<td></td>
</tr>
<tr>
<td>#1766</td>
<td>Query for social history (e.g. tobacco use) in HERON</td>
<td>mhoa</td>
<td>enhancement</td>
<td>minor</td>
<td>data-repository</td>
<td></td>
</tr>
<tr>
<td>#1850</td>
<td>review HERON user session, database command execution logs</td>
<td>bhamlin</td>
<td>task</td>
<td>minor</td>
<td>data-repository</td>
<td></td>
</tr>
<tr>
<td>#1620</td>
<td>Include units in Rstudio Data Builder data sets</td>
<td>dconnolly</td>
<td>task</td>
<td>minor</td>
<td>data-repository</td>
<td></td>
</tr>
<tr>
<td>#1639</td>
<td>Remove Obsolete IP_MAR_OLD Table from Clarity Extraction Scripts</td>
<td>ngraham</td>
<td>task</td>
<td>minor</td>
<td>data-repository</td>
<td></td>
</tr>
</tbody>
</table>

### Status: new

<table>
<thead>
<tr>
<th>Ticket</th>
<th>Summary</th>
<th>Owner</th>
<th>Type</th>
<th>Priority</th>
<th>Component</th>
<th>Version</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1698</td>
<td>Validate and Release HERON Council-Grove Update</td>
<td>ngraham</td>
<td>task</td>
<td>major</td>
<td>data-repository</td>
<td></td>
</tr>
<tr>
<td>#1737</td>
<td>Can’t build tooltips</td>
<td>ngraham</td>
<td>defect</td>
<td>major</td>
<td>data-repository</td>
<td></td>
</tr>
<tr>
<td>#1123</td>
<td>UHC: evaluate Emergency Room visits and sameday/surgical procedures</td>
<td>mnair</td>
<td>task</td>
<td>minor</td>
<td>data-repository</td>
<td></td>
</tr>
<tr>
<td>#1323</td>
<td>integrated Demographics with search by UHC/Epic source</td>
<td>mnair</td>
<td>enhancement</td>
<td>minor</td>
<td>data-repository</td>
<td></td>
</tr>
</tbody>
</table>
Tickets Equal Code using Version Control

Ticket #1680 (closed defect: fixed)

Age at visit incorrectly set to age now

Priority: major
Milestone: heron-council-grave-update

Component: data-repository
Version:
Keywords:
CC: 
Blocking:
Estimated Number of Hours: 0
Total Hours: 0
Add Hours to Ticket: 0
Backlog:

Description
In the de-identification step (source:heron_load/12b2_facts_deid.sql), age at visit is incorrectly set to the HIPAA age. It looks like this was introduced in [262:057534a].

Attachments

Change History

1680:4a91c554837df

Changset 1928:a91c554837df
Timestamp: 01/03/13 16:59:51 (9 days ago)

Author: Nathan Graham <ngraham@kumc.edu>
Branch: default
Children: 1931:8cf33beae7214, 1937:bl36cb39e3e36

Message:
Optionally use end date for the death fact in the death analysis script (#1655)
- Use the end date for UHC data
- Replaced misleading comments
- Put each fact count on its own line in the subtitle

File: 0 edited
- heron_analysis/death.Rmd (3 diffs)

Changeset 1956:e988930f5fde
Timestamp: 01/11/13 14:49:53 (6 days ago)

Author: Nathan Graham <ngraham@kumc.edu>
Branch: default
Parents: 1954:e888930f5fde (diff), 1955:0e98930f5fde (diff)

Message:
Merge with limit_age_1655
(No files)
HERON Council Grove update includes family and social histories

The Council Grove release introduces family and social histories within the HERON tool. Family history can be searched by all or by a particular family member type (e.g., Maternal Grandmother). Other data changes within HERON include encounter vitals, which are searchable under the Visit Detail folder, and smart data notes, which are expanded to include ambulatory in addition to the previously released inpatient notes.

HERON Council Grove Contents Summary

This month, our tour of Rivers and Lakes in Kansas honors Council Grove Lake.

The HERON repository contains approximately 1 billion real observations from the hospital, clinics, and research systems:

<table>
<thead>
<tr>
<th>Demographics</th>
<th>Observation</th>
<th>Patients</th>
<th>Source</th>
<th>Go-Live</th>
<th>Snapshot</th>
<th>Issues</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>18.5M</td>
<td>1.95M</td>
<td>KUH Billing (O2 via SMS)</td>
<td>1980s</td>
<td>Dec 2012</td>
<td>various*</td>
</tr>
<tr>
<td></td>
<td>15K</td>
<td>15K</td>
<td>Frontiers participant registry</td>
<td>Jun 2009</td>
<td>Dec 2012</td>
<td></td>
</tr>
<tr>
<td>Diagnoses (ICD9)</td>
<td>42.4M</td>
<td>657K</td>
<td>KUH/O2/Epic</td>
<td>Nov 2007</td>
<td>Dec 2012</td>
<td>various*</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>University HealthSystem Consortium (UHC)</td>
<td>Q4 2008</td>
<td>Sept 2012</td>
<td></td>
</tr>
<tr>
<td>Family and Social History</td>
<td>10.9M</td>
<td></td>
<td>KUH/O2/Epic</td>
<td>Nov 2007</td>
<td>Dec 2012</td>
<td></td>
</tr>
<tr>
<td>Medications</td>
<td>78.7M</td>
<td>302K</td>
<td>KUH/O2/Epic (Organized by VA Class)</td>
<td>Nov 2007</td>
<td>Dec 2012</td>
<td>various*</td>
</tr>
<tr>
<td>Nursing Observations</td>
<td>563M</td>
<td>?</td>
<td>KUH/O2/Epic</td>
<td>Nov 2007</td>
<td>Dec 2012</td>
<td>various*</td>
</tr>
<tr>
<td>Lab Results</td>
<td>83M</td>
<td>289K</td>
<td>KUH/O2/Epic</td>
<td>Nov 2007</td>
<td>Dec 2012</td>
<td>various*</td>
</tr>
<tr>
<td>Procedure Orders</td>
<td>113M</td>
<td>441K</td>
<td>KUH/O2/Epic</td>
<td>Nov 2007</td>
<td>Dec 2012</td>
<td>various*</td>
</tr>
<tr>
<td>Reports/Notes</td>
<td>35.7M</td>
<td>278K</td>
<td>UKP Billing</td>
<td>2000</td>
<td>Dec 2012</td>
<td></td>
</tr>
<tr>
<td>Specimens</td>
<td>44.3K</td>
<td>3.79K</td>
<td>KUH/O2/Epic</td>
<td>?</td>
<td>Dec 2012</td>
<td></td>
</tr>
<tr>
<td>Visit Details</td>
<td>27M</td>
<td>527K</td>
<td>KUMC Biospecimen Repository</td>
<td>2003 (?)</td>
<td>Dec 2012</td>
<td>various*</td>
</tr>
<tr>
<td>Cancer Cases</td>
<td>9.8M</td>
<td>66.4K</td>
<td>KUH/O2/Epic</td>
<td>Nov 2007</td>
<td>Dec 2012</td>
<td></td>
</tr>
<tr>
<td>Hospital Quality Metrics</td>
<td>4.43M</td>
<td>65.2K</td>
<td>KUH Cancer Registry</td>
<td>1950s</td>
<td>Dec 2012</td>
<td>labels*</td>
</tr>
<tr>
<td>Triple Negative Breast Cancer Registry (BRCA)</td>
<td>20.9K</td>
<td>122</td>
<td>University HealthSystem Consortium (UHC)</td>
<td>Q4 2008</td>
<td>Sept 2012</td>
<td></td>
</tr>
</tbody>
</table>

https://informatics.kumc.edu/work/blog
HERON De-identification Decisions

• HIPAA Safe Harbor De-identification
  – Remove 18 identifiers and **randomly date shifting by up to 365 days back in time**
    • Downside: can’t do seasonal studies without IRB approval to go back and get actual dates
    • In general, tack on 7 months when wanting volume for the last year.
  – 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)
  – While de-identified, access to timeline functionality provides individualized patient “signatures”

• **Date Shift example:**
  – Patient was *born August 13, 1968*, had their *blood pressure measured* on November 28, 2012.
  – Each month dates shifted, ex: to **-15** for January release: *New birthday* is **July 29, 1968** and the blood pressure measurement occurred on **November 13, 2012**.
    • For another patient, their offset might be **-278**. Next month the Aug 13th patient’s offset might be **-192**.
Not just Software: Engagement and Governance

- Dedicated Coordinator/Honest Broker: Tamara McMahon. Informatics Clinics held biweekly and one-to-one trainings and consultations offered.

- Integrating HERON’s use into other research workflows:
  - Finding patients for prospective trials: combining the Frontiers Participant Registry with the EMR data to find willing participants that meet study criteria.
  - Searching for samples: Biospecimen Repository combined with EMR to find tissues that meet research criteria.

Auditing small queries
"Who’s Using HERON" and collaboration approaches

Find a colleague
- Talk with hospital, clinic to understand workflow
- Attend bi-weekly clinics
- Watch the videos: [http://frontiersresearch.org/frontiers/informatics-training-videos](http://frontiersresearch.org/frontiers/informatics-training-videos)
- Request a consult [http://frontiersresearch.org/frontiers/biomedical-informatics](http://frontiersresearch.org/frontiers/biomedical-informatics)

If you don’t see what you want, or you really like things, let us know:
[https://redcap.kumc.edu/surveys/?s=3SBkPg&tool=1](https://redcap.kumc.edu/surveys/?s=3SBkPg&tool=1)
HERON’s Data Sources, Types of Data, Demo

Incorporate Clinical, Administrative, Research Datasources

- Inpatient and outpatient electronic medical records (Epic)
- Professional Services Billing and Scheduling (GE IDX)
- KUCC Biospecimen Shared Resource Samples Database
- Hospital (KUH) Tumor Registry (NAACCR format)
- Social Security Death Master File (NIST format)
- Technical Charges from hospital and clinics (UHC validated format)
- Research Data Capture (REDCap)
- Clinical Research Information System (RedCap)

HERON’s current contents:

- Demographics (master patient index)
- Race/Ethnicity
- Laboratory Results
- Nursing observations/vital signs
- Clinical Diagnoses (ICD9)
- Medications (dispensed, ordered, home meds, administered)
- Physician Orders
- Procedure charges (CPT)
- Outpatient Billing diagnoses (ICD9)
- Inpatient visit/provider service
- Family/Social/Past Medical History
- Specimen collected
- Tumor Staging and Grade
- Diagnosis and Treatment
- Survival and Progression
- Site Specific Factors (e.g. ER positive)
- Death per Social Security Administration
- MDSRG, APDRG, LOS, Readmissions
- Technical Charge Diagnoses ICD9
- Service line, AHRQ quality and JCAHO core measures

- Triple Negative Breast Cancer Registry initial pilot completed

Status as of January 17, 2013

https://informatics.kumc.edu/work/wiki/HeronProjectTimeline#Sep2012Planning
- contains current plan for next several monthly releases
Demo Areas

- Sliding workspace, minimize/maximize windows
- Existing: Frontiers, Tumor Registry, Biospecimens.
- Diagnoses Modifiers
- UHC
- Medications, Procedure Orders, REDCap
- Family History: Breast Cancer mom vs dad, link to Tumor Registry
- Same Financial Encounter: but Performance Struggles
- Demographics plugin, cancer survival analysis
- **New** Two way survival:
  - lung cancer **plus** smoke > 1 pack versus never
  - DRG CABG w/cath versus w/o cath: death in hospital, death per SSA
- Go to test server for RStudioServer “R Data Builder Demo”
  - Build a cohort with UHC data and then add a clinical feature. See the impact on mean LOS and the histograms.
Ideal Projects: validate HERON data and enable us to answer new clinical questions
  - Data Structure, Technology, Ontological in support of projects that test the data we load

Past: Flowsheets paper published on new data source
  - Challenges, what’s it tell us about nursing practice, future directions

Present: Medications with KUH Pharmacy (Sawyer) and Preventive Medicine (Shireman)
  - Can HERON medication representation measure time from order to administration for critical antibiotics in patients with Sepsis?
    - Tests modifiers (admin versus order), medication timing to the hour and minute (most i2b2 usage is at the day/month/year), and uses UHC DRG for Sepsis
    - Once accomplished, could look at clinical events suggesting Sepsis until clinical diagnosis entered

Diagnoses:
  - Internal Medicine example: Are patients’ “Do Not Resuscitate” preferences recorded in the O2 EMR problem list, reflected in inpatient Code Status orders?
  - Orders, Social History (smoking), Family History (ex: cancer Dx and Tumor Registry), Medication Delivery, Meds and Labs,
    - Data unused.... rots.
Hypothesis #1: Admin + Clinical -> Better Knowledge?

Hypothesis #2: Computer + Clinical Process -> Better Health?

Medical Informatics Environment and Hypotheses
From Data Aggregation to Prelim Analysis

• Motivation: Build a way to go beyond counting and obtain insight before you need a Data Use Agreement and risk disclosure of patient data.

• This is largely Dan Connolly’s work
  – grows out of work to create survival analysis tools for NCI site visit
  – R Data Builder plugin in i2b2 by Dan and integration with RStudio
  – This is being tested, not deployed for production yet.
  – RStudio Server (http://www.rstudio.com/ide/docs/server/getting_started)

• Very simple crude example at this point:
  – Hip Fractures, Coronary Artery Bypass Grafts, Hospital Length of Stay (O vs E)
  – Histograms, means, KnitR perhaps

• Looking for collaboration
  – Older models of administrative data based on linear and logistic regression, ex: lm(). Want to reproduce this for our data, UHC
  – Then develop systematic method to evaluate utility of clinical data
    • Perhaps applicability of newer statistical methods (ex: glmnet() and bootstrapping)
Questions?