

FRI 9:00 - 10:30: Data Sharing and Record Linkage

- U-shaped table: Bob Greenlee@MCRF, Abel Kho, Mei, Li, Maryan Zirkle @PCORI, David, Alex, Alfredo, Mihail Popescu@MU, Nathan, Dan, Mike, Bill Barnett@IU
- outer table: Tim@MU, Brad Taylor, Paul Albright @Cerner, Lidia Bernik @New York CDRN, Coleen Maguire@ABOUT network, Rita Fothergill @CMH, Angela Bos@UTHSCSA, Matt Hoag @KUMC, ??@Colorado, Susan @UTSW
- Small table: Umberto@WISC, Brad P.

Russ lead-in with (@@IOU link) Data Sharing and Record Linkage slides:

- Explanation of ResDAC: CMS is the U.S. Center for Medicare and Medicaid Services; ResDAC is an interface to their claims data for research purposes (resdac.org)
- Plan on getting fresher claims data directly from CMS.
 - To be used with a three cohorts.
 - Provides an example of how to link with insurance company.
- Buccaneers serves as an intermediate to the CMS data.
 - If linked data is desired SSNs need to be provided to Buccaneers
 - Able has proposed as hashing method to provide non-PHI to Buccaneer in place of SSNs.
 - Plan is to update the GPC CMS data annually (or more frequently, as-needed, with additional cost to project).
- Brad: we may want to include Death Index data in the process
 - Russ: We (KUMC) are licensed to share death index data; sharing it is on our todo list ([#377](#) Distribute Social Security Death Master file to sites)
 - Abel: Is DI data up-to-date.
 - Russ/Brad: Updated every month.
- Russ: Another source of information we have coming in is geocoding data from Minnesota. (current work: [#350](#))
- Russ: CMS data can only be used in a not-for-profit organizations/projects.
 - Can be used to make a case to work with industry partners.
 - Umberto: Would we then be viewed as commercial?
 - Russ: No, we can isolate data for different use cases (?)
- Russ: HIE are another potential source of data... hands out to Abel (Russ mentioned our [Cross-Site Patient Record Linkage Demo](#) in [#299](#))
- Abel presents his slides...
 - Abel: Indiana has a great HIE
 - No HIE in Chicago...
 - Since 2009 we been trying to merge data across sites.
 - Replicate what was done in Indiana.
 - AK Explains using HASH to uniquely identify patients.

- The technology we need it there, the difficulties are organisational and cultural (with medical community).
- Used combinations of hash input to increase security of resulting hashes.
- Umberto: How is the performance for kids?
 - Abel: We haven't looked at that population yet. This is still a work in progress.
 - Abel: One question is how does this approach vary among populations.
 - Customize for different populations (?)
- AK: Since phone numbers have become more static they are a great feature.
 - Epidemiological study want to know how representative your data is of the total population.
 - As you get more sites, and you get more data, you get closer to a true representation of the population
- Overview of cluster ID generation process
 - This process is governed by an update to the Omnibus Rule
 - person, place, and time are re-identification risks.
 - geographic and temporal obfuscation methods are an interesting topic for additional research.
- AK: We got legal representative of all the relevant sites together; all but one were OK doing this without a BAA. Lawyers are still working through approving counts. Once that's accomplished we will work on additional payloads.
- Umberto: Are users of this service using i2b2?
 - Abel: We are primarily a PopMedNet site (?).
- AK: This been a challenge both technically and legally.
 - Louisiana is currently using this.
- BobG: Are you building central capacities, or is it network/site specific?
 - We've hedged and prepared for both options.
- Russ: so you're answering stuff such as "Can you check how many vets you have?"; is that using this structure? ("Cluster ID..." slide)
 - AK: no, that project predates this full architecture [?]
- Russ: Who is the key master? Chicago?
 - Abel: Software has the potential to be decentralized (?)
- Russ/Abel: Describe Chicago Community Trust.
- Russ: Would like afternoon session to workout GPC use of Abel's software.
- Blue Button: In the VA you can use this blue button to download to download your data.
- Brad: How can this be used for prospective trials?
 - Abel: you add data as you go.
 - Brad/Abel: Hash can be aggregated between sources to identify patient
 - Abel: Desegregated by study, so you can associate hashes between studies (?).
- Umberto: Could you use a method similar to BitCoin
 - Abel: What we're doing is essentially like bitcoin

- Umberto: But isn't BitCoin entirely de-centralized?
- Abel: I believe they still use central stores
- Nathan: BitCoin uses a "ledger" ...
- Russ: How long does hashing take?
 - Abel: A million records took us a few hours
 - This is not for clinical purposes, it's intended for research.
 - Obviously it's hardware dependent.
 - Russ: how many records do you expect to do?
 - Abel: Currently at 8 million records in Chicago, 5 million de-duped.
 - Bill: and how often? not every day, right?
 - AK: plan is monthly
- Umberto: How do you handle redundancy from an infrastructural viewpoint.
 - Abel: This data center dependent (not application dependent) [?]
- Abel: The long term goal was to show hotspots in Chicago to advocate for an HIE.
- Abel: Are patients willing to sign a universal consent? (to David)
 - "I consent to have my information released by organization X"
 - Russ: the legal hurdles are larger at times than the technical hurdles.
- Russ: Have there been discussion on how to partners with CDRNs? (to Coleen)
 - Coleen: The discussion has begun, but no absolutes yet [?]
 - Russ: ABOUT patients have consented centrally to have their data released?
 - Coleen: ABOUT 70%, yes (?)
 - Russ: If there is a benefit to all parties involved then everyone is incentivised to participate.
- Alex: How do you reliably know where you patients are receiving their care?
 - Coleen: That's why we're looking into the hashing methods.
- Abel: There has to be PPRNs that have answer to these consent and patient care questions (?)
- Abel: Capricorn and GCP working together with funding from the CDC to determine whether Medicaid initiatives have improved diabetes outcomes.
 - Maryan: PCORI will be funding additional network which did not receive funding from the CDC in this effort.
- Mei: Thanks everyone for their participation.

FRI 1:00 - 2:30: Record Linkage Continuation

- Mei begins session and gives overview of morning session for those that were not at the morning session.
 - ResDAC
 - Linking patients between sites : Capricorn network's hashing approach.
 - Use hash id to link to CMS data as well.
 - Laurel: Is there a central repos of CMS data?
 - CMS wants to have one custodian per CDRN.

- Laurel: CMS data for a single pilot study, or general use (?)
 - Mei: For general use (?)
- Mei: Are there technical or institutional issues that are obvious?
 - Abu Mosa (MU): I don't know if CMS?ResDAC has been discussed in-depth with all site PIs.
 - What data gets shared, with who, and how.
 - Would like to see the proposal.
 - Mei: While the proposal has been submitted, we would like to wait until it's accepted to share, as edit/revision may need to be made before acceptance.
 - Abu: Were the data living could be an institutional issue?
 - Abu: CMS is probably the toughest organization to work with, with regards to data use. Governance will need to be more streamline.
- Abu: Plan to integrate data into i2b2?
 - Mei: Yes, that is what is in our application.
- Abu: Sees a potential issue with integrating CSM data into i2b2, as it may conflict with other CSM offering.
- Abu: Is there ability to reuse data?
 - Mei/Abel: Yes, the is a process (and re-use fee)
- Abu: Do we have a hashing tool?
 - Mei: Yes, Abel has worked on the development of one.
 - Abel: Yes, it is in use. A caveat is that it's still in active development.
 - Abu: Is there a fee?
 - Abel: Yes, a nominal fee.
 - Abel: We would work with you to balance specificity vs. sensitivity for you dataset.
 - Our primary focus has been from security and organization risk standpoints:
 - IRB (?)
 - data-use
 - VAA (?)
 - Mei: Who acts as the fourth party?
 - Abel: in our case the Community Trust handles nothing but the keys.
 - Abel: Benefit is that our solution is currently up and running in with two networks.
- Abel: It would be great to see per state counts of the DI data ...
- ...
- Laurel: What is you date shift method... we shift up to 365 days
 - Abel: We do the same.
- Abu: We need to know the entire workflow for this process. What is the timeline, etc.
 - Abel: Louisiana would be a good case study.

- Session broke into smaller discussions.