# Mid-South CDRN Face-to-Face Meeting - Informatics

September, 16, 2016

8:45 – 9:45AM EST

Medical University of South Carolina

Charleston, SC

mid-south clinical data research network

VANDERBILT UNIVERSITY MEDICAL CENTER

# PCORnet Data Model

mid-south clinical data research network

# Data Characterization – Cycle 1 Complete

- All data nodes have gone through Cycle 1 data characterization

**Prep-to-Research** – Data characterization of data node has been completed but interview with coordinating center was not completed prior to end of cycle 1. Sites will complete process in cycle 2.

- HSSC
- Meharry

**Research Ready** – Data characterization of data node has been completed, site has had interview with coordinating center to review data, and any updates have been made and reported back to coordinating center.

- Duke
- UNC
- VHAN
- VUMC
- Greenway*

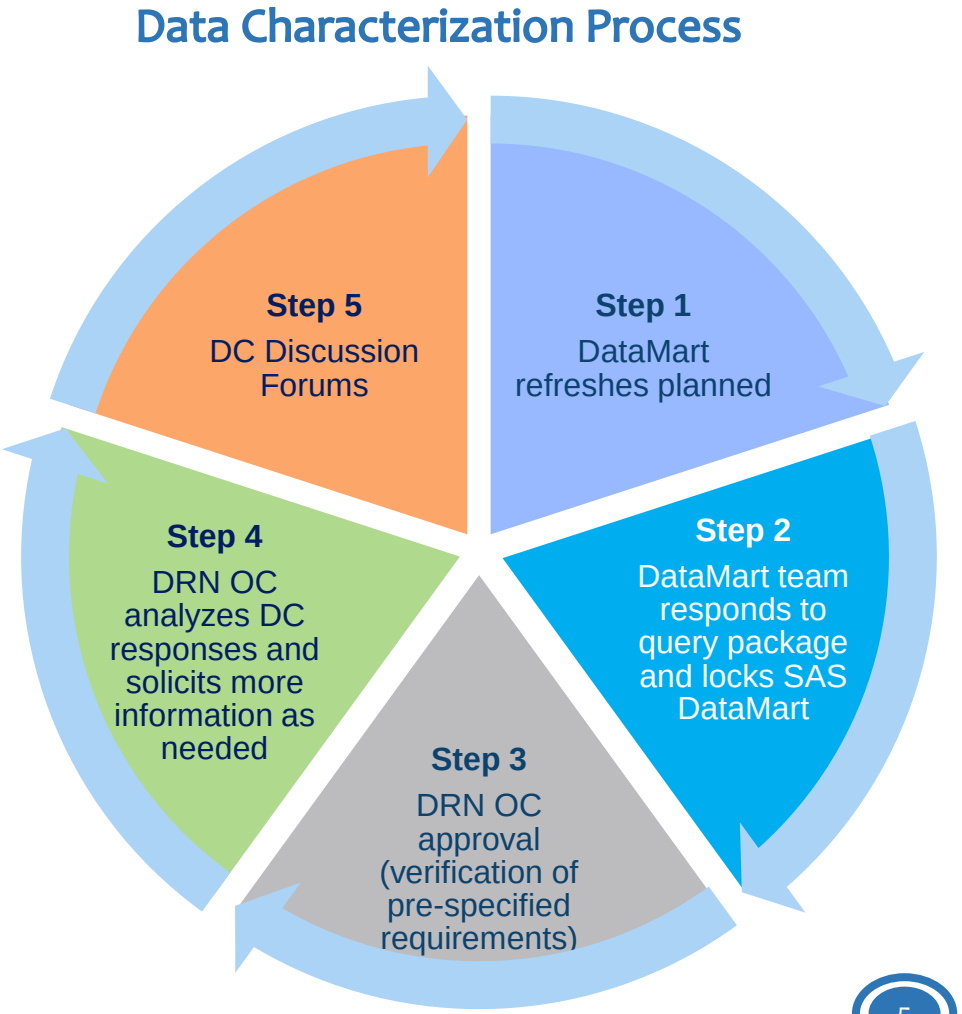
mid-south clinical data research network

# PCORI Common Data Model V 3.0

| Site | Sites in CDM | CDM Years | Data Refresh Cycles | Patients in CDM | Encounters in CDM |
|---|---|---|---|---|---|
| Vanderbilt | Vanderbilt University Health System | 2009-current | Dynamically updated version quarterly | 1,458,542 | 13,631,309 |
| VHAN | Williamson Medical Center, Maury Regional Medical Center, West TN Health | | Dynamically updated version quarterly | 346,241 | 968,254 |
| Greenway Health | 952 sites | 2010-current | Bi-Annual Update | 10,000,000 | 110,823,801 |
| UNC at Chapel Hill | UNC Hospitals , Rex Hospitals, UNC Physicians Network (more sites to come in June and July) | 2004-Present | Quarterly update-can do weekly if needed | 4,078,704 | 15,270,648 |
| Duke University | Duke University | 2005-2015 | Quarterly update | 2,062,439 | 34,172,582 |
| HSSC | Greenville Health System (GHS), MUSC Health (MUSC), Palmetto Health (PH), and Spartanburg Regional Healthcare System (SRHS) | GHS, MUSC, PH: 2007 - 2015<br>SRHS: 2011 - 2015 | Quarterly update | 2,879,835 | 39,068,523 |
| Meharry Medical Center | Meharry Medical College, Nashville General Hospital, and the Matthew Walker Comprehensive Health Center | 2011-current | Quarterly update | 18,874 | 18,874 |

# Data Characterization – Cycle 2 Beginning

- All Mid-South sites will run cycle 2 data characterizations which will include
  - Labs tables
  - Meds tables
  - Condisio
- Cycle 2 data characterization begins October 1




mid-south clinical data research network

# What to expect: Cycle 1 Vs. Cycle 2

| Aspect | Cycle 1 | Cycle 2 |
|---|---|---|
| Tables characterized | 7 tables - Demographic, Enrollment, Encounter, Diagnosis, Procedures, Vital, Harvest | 11 tables - Added Prescribing, Dispensing, Lab_Result_CM and Death |
| Query packages | Multi-Step Process - Diagnostic Query; Data Characterization Query; Empirical Data Characterization (EDC) reports | 1 integrated package |
| ETL Annotated Data Dictionary | Excel spreadsheets | REDCAP - Includes a section to comment on query results |
| DataMart Approvals | Step 1: Diagnostic; Step 2: Approved for PTR; Step 3: Approved for Research | Single approval based on pre-specified criteria. |
| DC Investigations | Discussions with ~50 DataMart teams; ~1/2 refreshed based on discussions | Discussion with all DataMart teams; opportunities to refresh based on discussions |
| Data Checks | 18 data checks | 20 data checks, classified as Required/Investigative |
| Duration and scope | Expected: 6 months; ~120 refreshes<br>Actual: 9 months; >170 refreshes | Expected: 4-5 months; ~100-160 refreshes |

# Data Characterization – Cycle 2 Beginning

- All Mid-South sites ( Research Ready and Prep-to-Research) will run cycle 2 data characterizations which will include”
  - Labs tables
  - Meds tables
- Cycle 2 data characterization begins October 1

# Additions to the Data Model: CDM 3.01, 4.0, ….

mid-south clinical data research network

# Soon to be released CDM V3.1

- v3.01 will not be an extensive update to the current CDM
- Updates encompasses “fixes” to CDM v3.0
  - Incorporated revised ETL Annotated Data Dictionary
  - Corrected truncation of some query results by increasing field lengths

# Mid-South Proposed Data Additions
## *CDM 4.0 and beyond*

| Data Element | Importance of Data Element |
|---|---|
| Cardiac Catheterization Data | Site is participating in initial pilots and the data would be helpful in heart disease studies and ADAPTABLE. |
| Echocardiogram Data | Site is participating in initial pilots and the data would be helpful in heart disease studies and ADAPTABLE. |
| Facility Type | Could help researchers limit analysis to specific facility types (i.e. hospitals, clinics, ets.) |
| Immunization Data | This is a data point that has been requested by researcher for studies. |
| Institutional Lab ID | Including the institutional LAB_ID would ensure lab data included in the CDM is more complete/identifiable within the source EHR data. |

| Data Element | Importance of Data Element |
|---|---|
| Payor/Insurance Status | Researchers are using this indicator of socio-economical status and has been requested by researchers. |
| All Lab Data | Current lab list is limited and not comprehensive enough for current CDM studies. |
| Mom-Baby Linkage | Site is participating in Antibiotic study and this linkage is a part of one of the aims of the study. |
| Patient's Current Zip | Beneficial for geospatial data analysis as well as enables researcher with the ability to perform better population-based research |
| Provider Type | This data element is often beneficial to researchers (i.e. distinguish between PCPs and specialists) |
| Referring Provider Type | This could be beneficial for capturing trends between providers |

# Discussion

- Are there any Issues/Concerns for completion of Data Characterization Cycle 2?
- Are there specific Data Additions as a Network we would want to prioritize internally?
- Other Questions/Topics/Issues?

# Study Specific Informatics Support

*Data Sharing*

mid-south clinical data research network

# Streamlining Data Sharing for our CDRN

DUA Working Group met August 17

- Defined a plan for the Mid-South data agreements for sharing within our CDRN and sharing to collaborators outside the CDRN
- A Draft Global Agreement and  External Agreement will be circulated for the WG to review **by October 7**

mid-south clinical data research network

# Global Mid-South CDRN Data Use Agreement

Global Agreements between Mid-South CDRN sites-An agreement would be between Mid-South CDRN sites after the infrastructure contract ends and would allow us to share:

- De-Identified Data
- HIPPA Limited-Data Sets
- Identified Data

➢ Establish a defined data sharing relationship amongst Mid-South CDRN sites to facilitate quicker, easier data sharing across sites.

mid-south clinical data research network

# External Investigator DUA

Create a uniform Mid-South Agreement covering:

- De-Identified
- HIPPA Limited
- Identified data

➢ Standardize data use expectations for investigators wanting to engage in research with Mid-South CDRN

mid-south clinical data research network

# Study Specific Informatics Support

## *Study Specific Data Support*

mid-south clinical data research network

# Types of Study Specific Queries via PopMedNet

## Pre-research

- Feasibility queries
- Engagement
- Match-making

**Prep-to-Research Ready**

## Interventional studies

- Clinical trials
- Pragmatic randomized clinical trials
  - e-Identification
  - e-Consent
  - e-Randomization
  - e-Follow-up
- Cluster randomization

## Observational studies

- Cross-sectional
- Epidemiology
- Health services
- Comparative effectiveness or safety

**Research Ready**

mid-south clinical data research network

# Study Specific Data Characterizations

- PBS – PCORnet Bariatric Study (Greenway, UNC, VUMC)
  - SSDC Complete!
  - Next  Study Deliverable: Aim 1 queries anticipated Winter 2017
- ABX – Antibiotic Study (Greenway, UNC, and VUMC)
  - SSDC sent to sites September 6, Responses due **September 20**
- ADAPTABLE
  - SSDC To be determined at a later date

# PopMedNet Query Tools

## PCORnet Modular Program 1 (PMP1)

- SAS based routine querying tool
  - Allows rapid query specification and distribution
  - Managed like software with routine updates and releases

## Menu Driven Queries

- Constructed in PopMedNet web portal (drag and drop)
- DataMart Client (DMC) communicates directly with your database
- Requires a specific adapter for your database type (provided by PCORI/Lincoln Peak)
  - PopMedNet Configuration Test completed by **September 9**

mid-south clinical data research network

# Discussion

- Are there additional areas to be considered for creating a unified Data Sharing/Use Agreements?
- Are there any questions or technical concerns with implementing Study Specific Queries?
- Are there any questions or technical concerns with supporting PopMedNet?
- Other Questions/Topics/Issues?

# Data Linkage Projects

# Additional Linkage for "Complete" Data

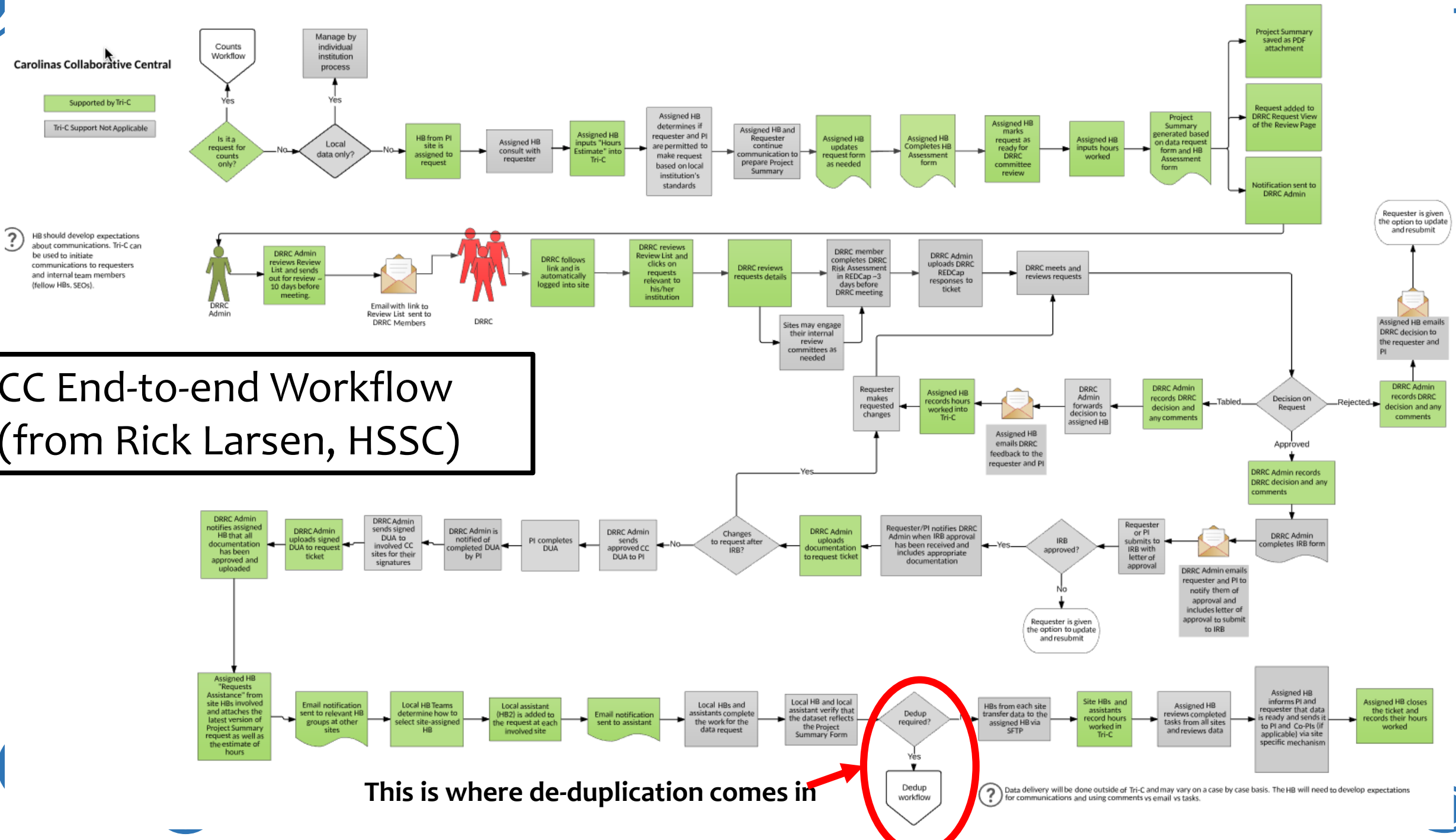
| Data Source | Description | Years Available | Identified (Y/N) | De-Identified (Y/N) | Linked to CDM (Y/N) |
|---|---|---|---|---|---|
| Vanderbilt Health Plan (Atena) | Universal Medical/ Dental Plan: Service, ICD, CPT, Admission, DRG Codes and Provider information | January 2014-September 2015 | Y | Y | Y |
| | Pharmacy Plan: dispensing data, dose, strength duration, refill info, pharmacy location | | | | |
| TennCare | Demographics, cause of death, death, diagnosis, dispensing, encounter, enrollment, and procedure | 2000-2014 | Y | Y | Y |
| TDOH | Statewide hospital/emergency discharge claims, birth/death certificates, demographics, diagnosis, dispensing, encounter, enrollment, and procedure | 2011-2013 | Y | Y | Y |
| North Carolina Blue Cross/Blue Shield | All inpatient and outpatient claims, including pharmacy | 2008-2015 | Y | Y | N |
| North Carolina Medicaid | DUA in process, not available yet. Will be all inpatient and outpatient claims | 2008-2015 (requested) | N | Y | N |
| South Carolina Claims | Discharge database including inpatient, outpatient surgery, emergency, ambulatory free-standing clinic, home health encounters. Demographic, encounter, diagnosis, procedure, and charges. | 2000-2015 | Y | Y | N |
| CMS | **Data acquisition underway** | - | - | - | - |

# Data Linkage Projects

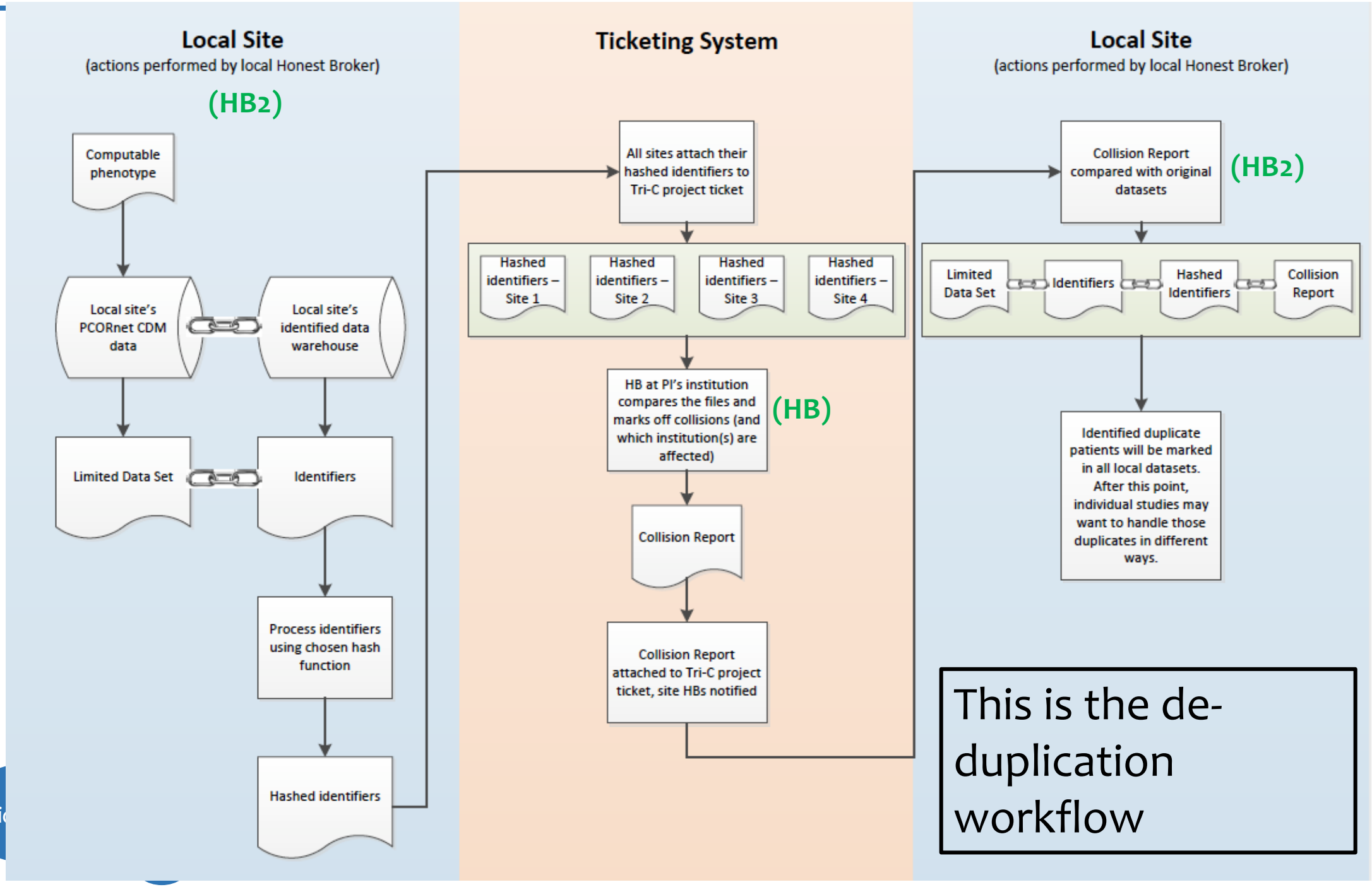
## *Carolinas Collaborative De-duplication*

mid-south clinical data research network

# CC End-to-end Workflow (from Rick Larsen, HSSC)




**This is where de-duplication comes in**


Local Site
(actions performed by local Honest Broker)
(HB2)
Computable phenotype
Local site's PCORnet CDM data
Local site's identified data warehouse
Limited Data Set
Identifiers
Process identifiers using chosen hash function
Hashed identifiers
Ticketing System
All sites attach their hashed identifiers to Tri-C project ticket
Hashed identifiers – Site 1
Hashed identifiers – Site 2
Hashed identifiers – Site 3
Hashed identifiers – Site 4
HB at PI's institution compares the files and marks off collisions (and which institution(s) are affected)
(HB)
Collision Report
Collision Report attached to Tri-C project ticket, site HBs notified
Local Site
(actions performed by local Honest Broker)
Collision Report compared with original datasets
(HB2)
Limited Data Set
Identifiers
Hashed Identifiers
Collision Report
Identified duplicate patients will be marked in all local datasets. After this point, individual studies may want to handle those duplicates in different ways.
This is the de-duplication workflow

# Data Linkage Projects

## *CMS Data Linkage*

mid-south clinical data research network

# Goals: Data Completeness

CMS claims and Mid-South CDRN site data to characterize the increase in data completeness and comprehensiveness provided through claims integration.

1. Quantifying completeness of the health system-derived data repositories
2. Using CMS claims data to enhance quality control processes for aggregating health system-derived data and establishing correlations with CMS claims data for health system-derived data to support trial recruitment  and observational studies.

mid-south clinical data research network

# CMS Data:  Possible Data Requests

Proposal 1: Create a cohort  to obtain claims data for all patients residing in Mid-South CDRN regions including the following states:

- Tennessee
- North Carolina
- South Carolina

Proposal 2: Obtain claims data for patients in our Healthy Weight cohort which includes patients in TN, NC, and SC.

- Mid-South CDRN CMS Workgroup met September 2 to review project and assess next steps

mid-south clinical data research network

# Decision Points

- Is it institutionally feasible for a combined submission?
- What will be the method of data management at each participating site?
- How will each participating site fund the up-front-costs?
- What will the support model be for on-going usage?

# Project Cost

## Up-front Costs

| Single Request | | | |
|---|---|---|---|
| | (>40 million, $60,500/yr) | Assuming all sites participate | Assuming 5 sites participate |
| 2010-2015 | $363,000 | $51,857.14 | $72,600.00 |
| 2011-2015 | $302,500 | $43,214.29 | $60,500.00 |
| 2012-2015 | $242,000 | $34,571.43 | $48,400.00 |
| 2013-2015 | $181,500 | $25,928.57 | $36,300.00 |
| 2014-2015 | $121,000 | $17,285.71 | $24,200.00 |

| Study Specific Request | | | |
|---|---|---|---|
| | (>40 million, $60,500/yr) | Assuming all sites participate | Assuming 5 sites participate |
| 2010-2015 | $291,000 | $41,571.43 | $58,200 |
| 2011-2015 | $242,500 | $34,642.86 | $48,500 |
| 2012-2015 | $194,000 | $27,714.29 | $38,800 |
| 2013-2015 | $145,500 | $20,785.71 | $29,100 |
| 2014-2015 | $97,000 | $13,857.14 | $19,400 |

## Additional Costs:

- Programmer Support for Data Management/Data Extracts
- Programmer Support for Data Linkage
- RA support for application submissions

mid-south clinical data research network

# Return on Investment

How many projects do we need to have a "reasonable cost" per project?

- Without incorporating additional cost and assuming a single requests, cost could be:

| | | | Number of Projects | | | | |
|---|---|---|---|---|---|---|---|
| Years | Single Request | Assuming all sites participate | 1 | 6 | 12 | 24 | 36 |
| 2010-2015 | $363,000 | $51,857.14 | $51,857 | $8,643 | $4,321 | $2,161 | $1,440 |
| 2011-2015 | $302,500 | $43,214.29 | $43,214 | $7,202 | $3,601 | $1,801 | $1,200 |
| 2012-2015 | $242,000 | $34,571.43 | $34,571 | $5,762 | $2,881 | $1,440 | $960 |
| 2013-2015 | $181,500 | $25,928.57 | $25,929 | $4,321 | $2,161 | $1,080 | $720 |
| 2014-2015 | $121,000 | $17,285.71 | $17,286 | $2,881 | $1,440 | $720 | $480 |

mid-south clinical data research network

# Discussion

- Are there other CMS Decision Points needed?
- Is it feasible to distribute cost to study users?  Or are there other funding streams that can be used to support this?
- Who from each site can be the identified point person for the CMS project?