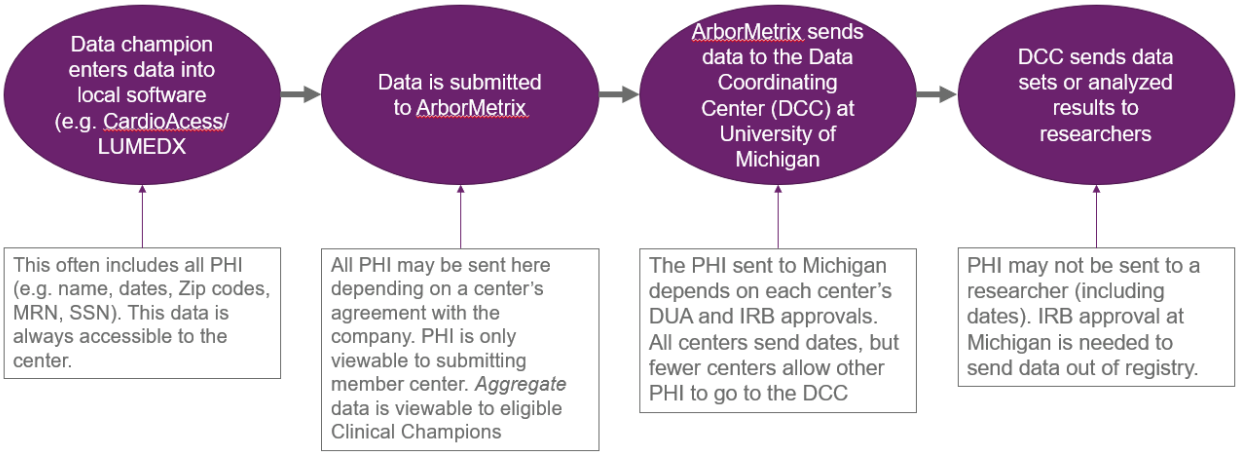
Data Flow for Research Projects

Data flow in the PC4/PAC^3 registry for research projects



What is PHI?

PHI is any health information that can be tied to an individual. PHI that is captured in the PC4/PAC^3 registry includes:

- *Name – only in ArborMetrix; not at DCC*
- All geographical identifiers smaller than a state, except for the initial three digits of a zip code
- Dates (other than year) directly related to an individual
- *Social Security numbers – not at DCC*
- *Medical record numbers – not at DCC*

But dates and zip codes are essential for my research project… what can I do?

If the dataset is being analyzed by the DCC at Michigan and researchers outside of Michigan will only see the aggregate results of the analysis, all PHI in the registry can be used in this analysis. If the analysis is done at another center, there are a few suggestions…

Suggestions for dates

- Dates can be shifted by the DCC by undisclosed multiples of seven, thus maintaining the original day of the week and accurately calculating date-dependent outcomes such as length of stay.
- Date of birth is not sent by all centers. *However*, the DCC receives the patient's age in days at hospital admission, so the date of birth is deduced by subtracting hospital admit date from the age in days at hospital admission.

Suggestions for zip codes

- Drop patients that do not have a zip code on file altogether or can remain in the project but are excluded from analyses that include demographic variables.
- Have analysis related to zip code conducted at the DCC and request results of that analysis (e.g. census poverty score rather than zip code)
- If collecting data prospectively, consider using a geocoding tool (e.g. DeGAUSS. https://degauss.org/). Each institution participating in a project must use this tool on a computer at their hospital and then email the output to the DCC.

Menu of quotes to include in the IRB application to Michigan: *(Select all that apply)*

- "No identifiable data is leaving the DCC at University of Michigan."
- "Data is being analyzed at the DCC at the University of Michigan" **or** "Only aggregate, de-identified data will be sent to [center] for analysis."
- If the data is being analyzed by a third party (e.g. Dr. Patel at Colorado): "Aggregate, de-identified data is going to [center] for analysis and the results of that analysis will be sent to [your center]."
- If date shifting or other methods are being used to obscure identification, include that as well.