



PopMedNet: Data Characterization Tools in Distributed Networks

Melanie Davies, HPHCI; Jessica Malenfant, MPH, HPHCI; Bruce Swan, LPP; Jeffrey Brown, PhD, HPHCI
 HPHCI: Harvard Pilgrim Health Care Institute; LPP: Lincoln Peak Partners



Background

- The PopMedNet™ (PMN) platform was implemented for the Mini-Sentinel distributed research network in 2010 and continues to be enhanced with new functionality each year. Mini-Sentinel uses PMN to facilitate medical product safety and related public health surveillance activities.¹⁻⁴
- The Mini-Sentinel Operations Center (MSOC) conducts stringent data quality review and characterization checks for each data partner's Mini-Sentinel Distributed Database. After each data refresh, the MSOC distributes program packages for local execution by the data partner. Aggregate output tables are returned to MSOC for review.⁵
- To improve network learning and make network data more readily available across Mini-Sentinel investigators, in 2014 the Mini-Sentinel project supported development of a new PMN feature, the network Data Checker tool.

Objectives

Allow querying of aggregate data characterization data to quickly assess key characteristics such as missingness and distribution of select variables. These queries are executed against the aggregate data characterization data returned as part of the routing data quality assurance process.

Data Checker Criteria	
Data Partner	Race
<input type="checkbox"/> Organization A (AAAA)	<input type="checkbox"/> Unknown
<input type="checkbox"/> Organization B (BBBB)	<input type="checkbox"/> American Indian/Alaska Native
<input type="checkbox"/> Organization C (CCCC)	<input type="checkbox"/> Asian
<input type="checkbox"/> Organization D (DDDD)	<input type="checkbox"/> Black/African American
<input type="checkbox"/> Organization E (EEEE)	<input type="checkbox"/> Native Hawaiian/Pacific Islander
<input type="checkbox"/> Organization F (FFFF)	<input type="checkbox"/> White
	<input type="checkbox"/> Missing

Data Checker Race Query Interface

Methods

- The PopMedNet and MSOC teams collaborated to identify high priority data characterization queries and define the requirements for the new querying functionality.
- A new data characterization data model was defined and populated to enable querying.
- The teams assessed the requirements to ensure that the enhancements would meet the specific needs of MSOC while being general and extensible enough to be seamlessly implemented for other distributed research networks using PMN.
- Nine new PopMedNet Data Checker request types were designed, developed, and implemented within the Mini-Sentinel network. Users can run data checker queries using a point-and-click interface incorporated into the PMN platform. The nine request types focus on the following:
 - Race
 - Ethnicity
 - Data completeness by table
 - Diagnosis codes
 - Discharge diagnosis types by encounter type
 - Presence of National Drug Codes
 - Pharmacy dispensing amount
 - Pharmacy dispensing days supply
 - Procedure codes
- Query results can be displayed across all sites, within a specific site, or stratified by each site. Data are presented overall and by data partner in tables, graphs, and charts.

Race	n	%
Unknown	88648115	64.24
American Indian or Alaska Native	247859	0.18
Asian	2851445	2.07
Black or African American	6345641	4.6
Native Hawaiian or Other Pacific Islander	353943	0.26
White	39538240	28.65
Missing	101	0

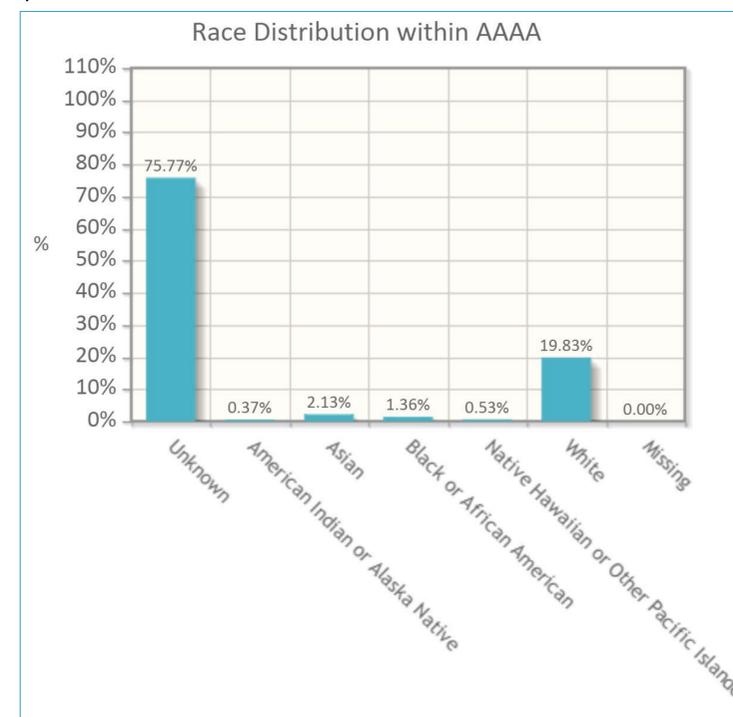
Data Checker Results Across All Data Partners Displayed as Table

Results

- The new data characterization tools were implemented for users in the Mini-Sentinel Operations Center. The MSOC teams use the Data Checker to help inform data requests and investigate data resources in preparation for requests.
- PCORnet has adopted the data characterization data model and will use the Data Checker query across the PCORnet network of over 50 partners.

Conclusions

- The Data Checker tool allows networks to easily query commonly generated data characterization data (e.g., rate of missingness, proportions). This helps investigators better assess data availability before developing analytic code or distributing queries to data partners.
- The Data Checker tool and data model are generalizable for use in other distributed research networks. Any network that utilizes this new data characterization data model based on their own data checking and quality assurance processes can take advantage of this data checking tool in PopMedNet.

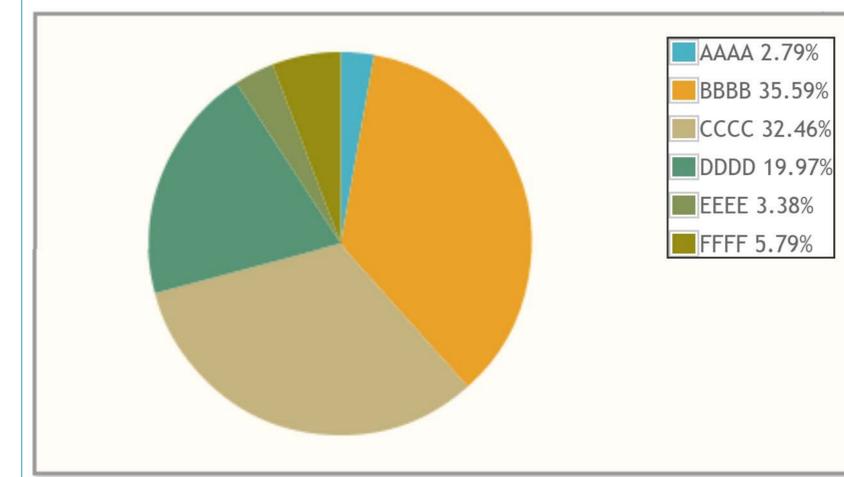


Data Checker Results Within Data Partner Displayed as Bar Chart

Recommendations

- The following functionality should be considered for future work:
- Enhance the data characterization data model to include additional data domains from other common data models (e.g., clinical and laboratory data).
 - Incorporate additional tables or data elements (e.g., incorporate data from multiple data refreshes for each data partner/site) to assess trends.
 - Extend access to the tool while maintaining data partner privacy by refining permissions to query all or specified sites, view site-specific or aggregate results, and view identified or masked site IDs.
 - Transition from a query-based approach to a reporting tool that allows the users to customize their searches and the report output.

Data Partner Contribution to Race: Unknown



Data Checker Results Stratified by Data Partner Displayed as Pie Chart

References

- Behrman RE, *et al.*, Developing the Sentinel System – a national resource for evidence development. *New England Journal of Medicine* 2011; 364(6): 498-99.
- Brown J, *et al.*, Distributed Health Data Networks: A Practical and Preferred Approach to Multi-Institutional Evaluations of Comparative Effectiveness, Safety, and Quality of Care. *Medical Care* 2010; 48 Suppl: S45-S51.
- Maro JC, *et al.*, Design of a national distributed health data network. *Annals of Internal Medicine* 2009; 151:341-4.
- PopMedNet: <https://www.popmednet.org>
- Mini-Sentinel Operations Center, Mini-Sentinel Common Data Model Data Quality Review and Characterization Process and Programs v3.2 – Overview. May 2014.