Identification, Remediation and Knowledge Management of Data Quality Issues in PCORnet: Experiences in the Obesity Demonstration Projects

Casie E. Horgan, MPH[1], Jessica L. Sturtevant, MS[1], Jane Anau[2], Juliane Reynolds, MPH[1], David Arterburn, MD, MPH[2], Jason P. Block, MD, MPH[1], and Jeffrey S. Brown, PhD[1]

[1]Harvard Medical School and Harvard Pilgrim Health Care Institute, Boston, MA; [2]Kaiser Permanente Washington Health Research Institute, Seattle, WA

Background

- The PCORnet Distributed Research Network (DRN) uses a Common Data Model (CDM) to allow for development of efficient, reusable analytic tools.
- The PCORnet Antibiotics (ABX) and Bariatric (PBS) Demonstration Projects assessed heterogeneity of CDM implementation and available data at sites via a process called Study Specific Data Characterization (SSDC).
- Conducting SSDC identified numerous issues in network data quality or related network policy/procedures.

Methods

- Issues were initially identified via execution of distributed SAS SSDC programs, evaluation of SSDC program logs, and detailed examination of returned SSDC data. Additional issues were identified via communication with Network Partners (NPs).
- Issues were tracked and analyzed to identify recurring themes and most effective methods of remediation.
- Areas of concern and recommendations for remediation were summarized and shared with PCORnet stakeholders to help inform future querying and development efforts.

Results

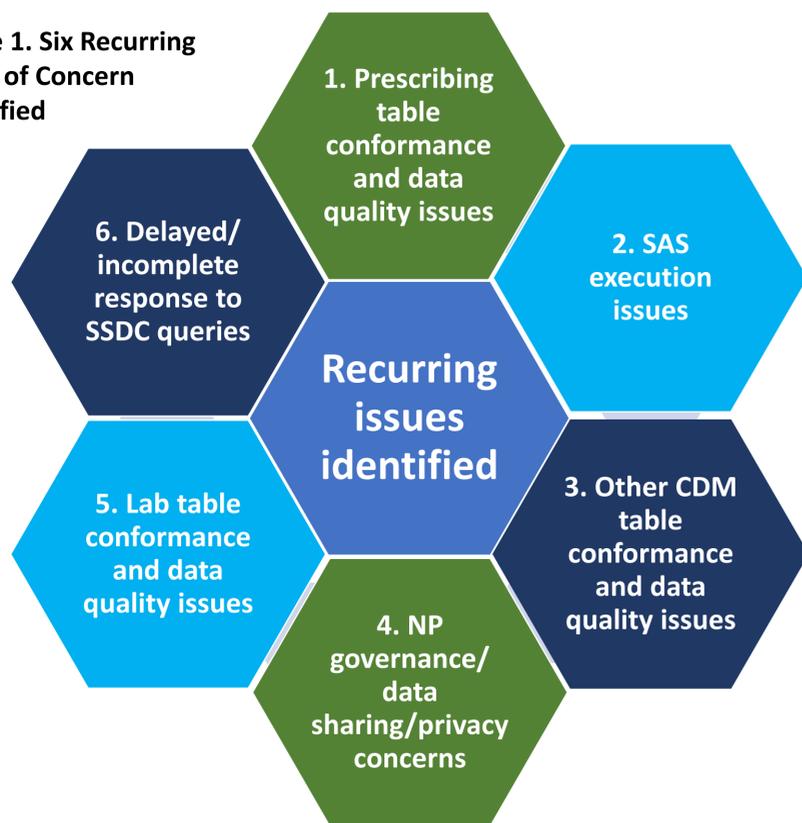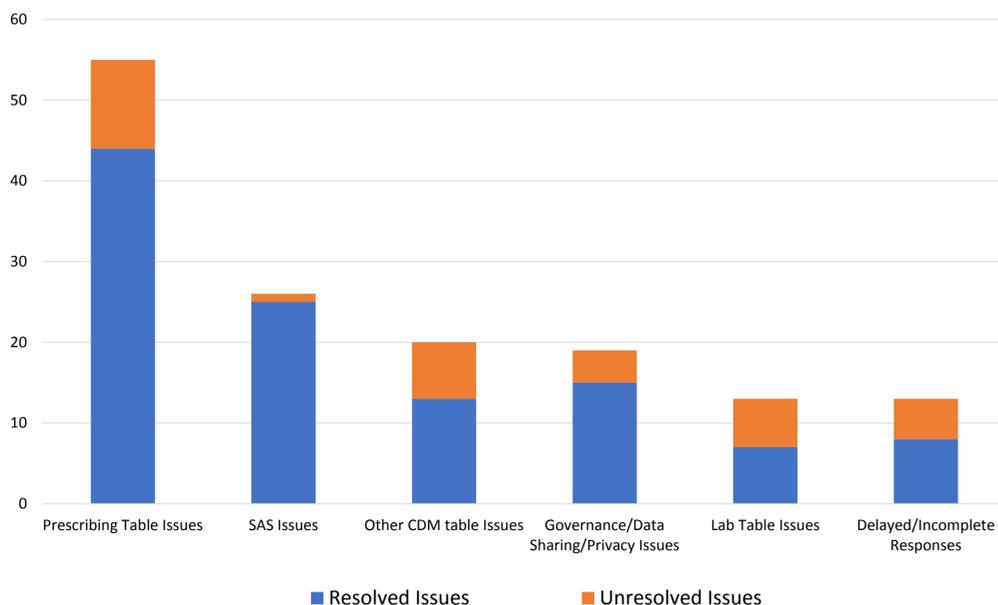Figure 1. Six Recurring Areas of Concern Identified

Recurring issues identified
1. Prescribing table conformance and data quality issues
2. SAS execution issues
3. Other CDM table conformance and data quality issues
4. NP governance/ data sharing/privacy concerns
5. Lab table conformance and data quality issues
6. Delayed/ incomplete response to SSDC queries

Recommendations for Studies in PCORnet

1. Require SSDC of prescribing table prior to research. Examine medication codes of interest in detail.

2. Use standard program format and programming so NPs are familiar with execution.

3. Require SSDC of relevant tables prior to research. Maintain PCORnet data issue repository for easier error identification and remediation.

4. Maintain open communications about types of data to be utilized throughout the study and the future uses of said data.

5. Require SSDC of lab table prior to research. Examine lab codes of interest in detail.

6. Build time into timelines to allow for SSDC, remediation and subsequent re-assessment. NPs must have data sharing policy in place before participation.

Chart 1: Issue Count by Resolution Status

Prescribing Table Issues, SAS Issues, Other CDM table Issues, Governance/Data Sharing/Privacy Issues, Lab Table Issues, Delayed/Incomplete Responses

■ Resolved Issues   ■ Unresolved Issues

Conclusions

- Tracking and analysis of data issues across network projects allows for coordinated knowledge management of lessons learned.
- Issue topics and resolutions identified by the Demonstration Projects contribute to the knowledge repository to inform future studies.
- Knowledge repository helps create a virtuous learning feedback loop so future studies have a better understanding of data available and can be implemented more efficiently.
- Highlighting recurring problems allows the Network to focus resources on areas requiring improvement.
- Development of a formal issues tracking process and knowledge management increases Network efficiency while allowing for broader sharing of issues and remediation techniques.

Contact Information
Casie Horgan, MPH
Harvard Medical School, Harvard Pilgrim Health Care Institute;
Casie_Horgan@harvardpilgrim.org

AMIA 2018 Informatics Summit March 12th-15th

pcornet®