



PCOR Building Data Capacity Workshop 3

Allison Oelschlaeger
CMS Chief Data Officer
Director, Office of Enterprise Data & Analytics

Meeting PCOR Research and Data Capacity Challenges

- Over **130 million Americans** receive healthcare coverage through programs administered by CMS
 - Medicare
 - Medicaid and the Children's Health Insurance Program (CHIP)
 - Health Insurance Marketplaces
- CMS data is an invaluable resource for comparative clinical effectiveness research
 - Enrollment information and patient characteristics
 - Utilization and cost of health care services, including treatments/therapies
 - Diagnoses

Limitations of CMS Data

- CMS data is often not sufficient on it's own for clinical effectiveness research
 - Cause of death data
 - Clinical data (e.g., lab results)
 - Certain patient demographic data (e.g., income, high quality race/ethnicity)
 - Social determinants of health-related data
 - Health behaviors data
 - Patient generated health data
 - Human services data
 - And more
- Options for enhancing CMS data include: new data collection, imputation, or linkages to other data sources

PCOR-TF Projects and Other Data Linkages

- PCOR-TF has funded a number of projects that leverage CMS data:
 - Blue Button 2.0 Application Programming Interface
 - Data linkages
 - Augmenting the National Hospital Care Survey Data through Linkages with Administrative Records
 - A National COVID-19 Longitudinal Research Database Linked to CMS Data
 - National Death Index – Medicare Enrollment Data Linkage
- Other data linkages:
 - Surveillance, Epidemiology, and End Results (SEER)
 - US Renal Data System (USRDS)
 - Health and Retirement Survey/National Health and Aging Trends Study

Data Linkage Barriers

- Collection of sufficient identifiers (e.g., only last 4 digits of SSN) to facilitate linkage while also protecting patient privacy
- Limitations on sharing of identifiers in private/proprietary data
- Conducting linkage may require access to a larger set of patient records than is needed once the linkage is complete
- Multiple tokenization solutions
- Probabilistic matching consistency

Data Linkage Solutions: Virtual Research Data Center

- The CMS Virtual Research Data Center (VRDC) is a secure and efficient means for researchers to virtually access and analyze the vast store of CMS data
 - Researchers access and perform their own analyses and manipulation of CMS data virtually from their independent workstations
 - Researchers can download aggregate results from their analyses
- Benefits for researchers include:
 - Faster access to data and lower cost for large cohorts
 - CMS data is already linked by a universal beneficiary ID
 - The ability to upload **other data into the VRDC for linkage**