Administrative and Vital Records Data Housed in the Enterprise Data Warehouse (EDW)

CMS National Medicaid and CHIP Quality Conference
June 14-15, 2012

Gwen Smith, Project Director

CHIPRA Child Health Quality Demonstration Grant

Illinois Department of Healthcare and Family Services

Improving Care and Proving It

Administrative and Vital Records Data Housed in the EDW

Topics

- Background on EDW

- Data sources

- Matching process

- Matching constraints/challenges

- Data analysis/use

- Lessons learned

Background on EDW

- Legislatively mandated – proposed and funded in late 90's
- Construction began in 2000
- Outside contractor – Optum Government Solutions
  - Hardware and software
  - Support (development for data loading, matching, extraction)
  - Optum staff housed at HFS
- High level of expertise required due to complexity of data and scope/size of database
  - Extensive training for users
  - Tools under development to simplify EDW use
- Accommodates users from other state agencies

EDW Data Sources

- Medicaid program data
  - Financial
  - Services
  - Patient, provider and payee history

- Department of Public Health
  - Vital records – births and deaths
  - Lead screening
  - Adverse pregnancy outcomes
  - Genetic screening

- Department of Children and Family Services
  - Wards of the State

- University of Illinois at Chicago - Division of Specialized Care for Children
  - Children with specific chronic conditions

- Immunization registry information444

EDW Data Sources

- Department of Human Services - Cornerstone
  - WIC
  - Case management programs
    - Family Case Management
    - Healthy Start
    - Teen Parent Services
    - Targeted Intensive Prenatal Case Management
  - Early Intervention
  - Family planning

- Under construction
  - Early hearing detection and intervention
  - Hospital discharges
  - Pre-admission screening
  - Out-of-state vital records
  - Pregnancy Risk Assessment Monitoring System (PRAMS)

Handling Data from a Variety of Sources

- Each data source is drastically different
  - Different uniquely identifiable information
  - Different identifiable information
  - Significantly different quality
- Need to link data can create challenges
  - Keys to access individual data all different
  - Keys to access individual data may not be unique
- Demographics information can be significant challenge
  - Names (spelling, use of nicknames)
  - Addresses cannot reliably be used to match individuals
- Requires introduction of "probability" in some situations

Improving Care and Proving It

Data Matching Process

- Use deterministic and probabilistic matching

- Data preparation – Can't Rush!
  - Identify matchable data fields and assess data quality
  - Establish common names (Bob and Bobby become Robert)
  - Weight the match criteria
  - Establish scoring for each potential match

- Weights/scoring
  - Score each unique value for each field (weighting)
  - Score each potential field match based on nearness of data fields to unique values
  - Calculate weights and scores to determine final match score for the record
  - Determine "match" or "no match"
  - Evaluate matches

Improving Care and Proving It

Matching Constraints

- Data quality
  - Duplicate entries, missing or invalid information, multiple names

- Cannot guarantee 100% match
  - Different match rates for different sources
  - Decisions required on use of unmatched data for analysis purposes

- Access/security must be identified and controlled
  - Central management model best
  - Each user must have written approval/authorization to see data
  - Access/security limitations can affect ability to match data sources
  - Not HIPAA trained, no access to claims data

- Legal issues
  - Explicit data sharing agreements between data owners
    - Specify who can access data and how it can be used/shared
  - State/federal law limiting access to or use of data

Improving Care and Proving It

Linked Data Sets in Illinois

Data Sources

- **Moms and Babies Data Mart**

  - Links data between moms and their babies
  - Vital statistics
  - Adverse Pregnancy Outcome Reporting System (APORS)
  - Cornerstone (WIC/ Family Case Management)
  - Administrative data

- **Quality of Care**

  - Administrative data
  - Immunization registry
  - Cornerstone (WIC/Family Case Management)
  - Lead screening data

Improving Care and Proving It

Linked Data Sets in Illinois

Data Uses

- Moms and Babies Data Mart
  - Data analysis
    - Annual estimates
    - Risk identification
    - Case identification
    - Geographic distribution
    - Providers
    - Identify conditions associated with adverse outcomes
  - Predictive analytics
  - Measure outcomes (cost, quality, care coordination, system connectivity)
  - Family planning waiver reporting (birth spacing, fertility rate)

- Quality of Care
  - HEDIS quality measures
  - CHIPRA core measures
  - Pay-for-performance measures
  - Program monitoring

Improving Care and Proving It

How is Data Used?

- Providers
  - Practice change
  - Patient outreach/compliance
  - Quality improvement

- HFS
  - Program evaluation/monitoring
  - Policy/program changes
  - Public reporting
  - Federal reporting requirements
  - Quality improvement
  - Identification of high-risk populations

- Others
  - Other state agencies
  - Federal government
  - Advocacy groups
  - Patients
  - General public
  - Provider organization

Improving Care and Proving It

Lessons Learned

- A high quality back-end matching process can never replace high quality data input with a unified identification numbering system.

- Multiple state agencies and others can benefit from aggregating data and having access to data matched to Medicaid claims.

- Including non-clinical data adds richness to data analysis and affords opportunities for additional analysis.

- Aggregating data provides robust data analysis opportunities (CHIPRA core measures, quality measures, identification of high-risk populations).

- Data access and security is essential, but can also be prohibitive to working with the data.

- Legal aspects of data access must be addressed.

Conclusion

In spite of the constraints/challenges associated with data aggregation and matching, having aggregated and matched data provides richer data, more data points for more robust data analysis, more opportunities for using data for quality improvement, better measurement of outcomes, and increased likelihood of more accurately identifying trends and predictive characteristics to improve health outcomes.