



Limited Data Sets for the MA APCD

Special User Group Meeting
September 9, 2015

Current Data Request Process



- Requests made at the element level
- Requests reviewed at the data element level for privacy and minimum use concerns
- CHIA IT fulfills customized extract for each request
- CHIA analysts support customized extracts

Issues with Current Process



- Completing request is time intensive – for requestors and CHIA
- Reviewing and fulfilling data requests are resource intensive
- Data elements with a significant amount of missing data forces users to unexpectedly revise their analytic plans
 - B and C elements with low thresholds, many variances
- Level 2 data elements currently in release may pose re-identification risk
- Amendments (additional elements) often requested

Goals for Limited Data Set (LDS)



- Protect patient privacy
- Serve analytic needs of the non-gov't users as is possible
 - Gov't users would get access to all MA APCD data as needed
- Streamline request and review processes

LDS as Defined by HIPAA



- Excludes a specific set of direct identifiers, such as the following which appear in MA APCD:
 - Name
 - Postal Address
 - SSI
 - Medical Record numbers
 - Health plan beneficiary numbers
- May include dates (of admission, discharge, service, birth, death)
- May include age, city or town, state, ZIP

CMS LDS's



- Limited sets of patient-level PHI in which selected variables are encrypted, blanked or ranged.
 - Excludes SSI
 - Excludes ZIP, but includes county and state
 - Excludes date of birth, but includes either age in years or 5-year age range
 - Includes encrypted beneficiary identifiers on claims and enrollment files
 - Includes encrypted NPIs in provider files (1999-2012) but includes real NPIs beginning in 2013
 - Includes claim file dates

MA APCD Proposed Limited Data Set



- MA APCD LDS is a hybrid
 - Contains only information that is permitted for inclusion in a HIPAA LDS
 - Incorporates additional privacy protections, such as ranging and encryption, from CMS LDS

CHIA Methodology



- Determined what must be excluded:
 - HIPAA-defined direct identifiers with the exception of carrier specific subscriber and members IDs which are hashed
 - All Level 3 MA APCD data elements
- Excluded elements due to significant amount of missing data:
 - Examples: inpatient DRGs, outpatient APCs, hours of admit/discharge
- Investigated elements that potentially should be excluded due to patient privacy concerns
 - Free text fields, including carrier-defined/non-standard lookup tables, names of drugs, street addresses

CHIA Methodology – con't



- Exclude certain quasi-identifiers:
 - Which make individuals unique in the population and thus possibly used for indirect re-identification
 - Examples: Race, Ethnicity, Member Language, Disability Indicator, Family Planning Indicator, Member SIC code
- Retained, but ranged, the following:
 - Individual relationship code
 - Gender
- Maintained substance abuse (Part 2) filter

Examples of Level 2 Data Elements Not in Proposed LDS



- Member and subscriber birth month
- Service provider name, city, state, ZIP
- Date service approved
- Admission/discharge hour
- DRG, APC
- Product ID number
- Denial reason
- Family planning indicator, EPSDT indicator,
- Denial reason
- Race, ethnicity, language preference
- NUBC codes – condition, value, occurrence,
- Accepting new patients, EHR used, offers e-visits

CHIA Methodology

Transforming Geography and Age Information to Reduce Risk of Re-identification



- HIPAA LDS: Excludes postal address, but may retain city or town, state and ZIP
- CMS LDSs: Allow for county and state info only
- CHIA team discussed pros and cons of providing more or less granularity in the MA APCD LDS

Geography / Beneficiary Age Transformations



- Re-coded (ranged) all out of state information to “non-MA”
- Calculated age at end of year
- Ranged ages 65-74 and 75+ due to the fact that seniors are under-represented in the MA APCD available to non-gov’t users
 - CHIA’s DUA with CMS only allows CHIA to share Medicare files with gov’t agencies
- Standardized MA county and municipality data

Two Approaches to Geography Granularity



One Approach – 3 levels

1. State
2. State, MA County*
3. State, MA County*, “Large” Municipalities**

Second Approach – 2 levels

1. 3 digit ZIP
2. 5 digit ZIP

* Dukes, Nantucket and Barnstable counties combined

** Boston, Worcester, Springfield, Lowell, Cambridge, New Bedford, Brockton, Quincy, Lynn, Fall River, Newton, Lawrence, Somerville, Framingham, Haverhill, Waltham, Malden, Brookline, Plymouth, Medford, Taunton, Chicopee, Weymouth, Revere, Peabody, Methuen, Barnstable, Pittsfield, Leominster

Provider Identifiers – Proposed Two Levels



- Option 1: Hashed NPIs and CMS Provider Type (from NPPES)
Allows users to track physicians across payers
- Option 2: Unhashed NPIs
Users link to external sources (such as NPPES) for name, address, etc.

LDS for MA APCD



- Would apply to non-government users only
- Simplified request form. Requestors would need to justify:
 - Geo breakout
 - Unencrypted NPI
 - LDS files needed – not elements
- DUAs and Data Management Plans would still be required
- MassHealth would review requests for MassHealth data

Input Sought



- Feedback on approaches for MA geographic breakouts
 - County/Muni
 - 3 digit/5 digit
- Recoding to non-MA for the states contiguous to MA
- Impact on usefulness of MA APCD

If you prefer to send written comments: apcd.data@state.ma.us.