Data Improvement Efforts: Centers for Medicare & Medicaid Services

Thomas Reilly, Ph.D.
February 10, 2009
Today’s Discussion

- Data improvements at CMS that may be helpful for the NHQR and NHDR
  - Chronic condition data warehouse (CCW)
  - Part D prescription drug event data
  - Improved coding of race/ethnicity in CMS administrative data
As a byproduct of administering the Medicare program, CMS has a wealth of data on program enrollment, health care utilization, and cost.

Historically CMS data has been in “stovepipes” and can be very difficult to use.

Section 723 of the MMA mandated that CMS create a data base to better support research on chronic conditions.

The CCW is an initiative to integrate our administrative data systems to facilitate research applications across the continuum of care.
Chronic Condition Data Warehouse

Content (for all Medicare beneficiaries)

- Enrollment/eligibility data
- Claims data
  - Medicare fee-for-service institutional claims (inpatient, outpatient, skilled nursing facilities, home health, hospice)
  - Medicare fee-for-service non-institutional claims (physician/supplier, durable medical equipment)
  - Medicare prescription drug event data (new)
  - Medicaid claims for dual eligibles (future)
Chronic Condition Data Warehouse

Content, continued

- Health assessment data
  - Minimum Data Set (MDS) – nursing homes
  - Outcome and Assessment Information Set (OASIS) – home health
  - Inpatient Rehabilitation Facility Patient Assessment Instrument (IRF-PAI)

- Flags to identify beneficiaries with 21 selected chronic conditions
Chronic Condition Data Warehouse

For a given project researchers can:

- Request 100% files (subject to minimum data necessary restrictions)
- Request data for the CMS 5% beneficiary sample
- Define custom cohorts or submit finder files for extraction

All requests are subject to standard CMS privacy protection and data release protocols
Chronic Condition Data Warehouse

Our intent is that the CCW will provide an improved infrastructure to support research and measurement efforts using Medicare data

- Cross-sectional and longitudinal analyses
- Linkage across the continuum of care
- Large enough to support the study of small populations and relatively rare events
Prescription Drug Event Data

- Part D program makes available prescription drug coverage to Medicare beneficiaries
- Notice and rule-making process was followed to define how prescription drug event (PDE) data generated through the Part D program can be released
- Part D Data Rule became effective in June 2008
Prescription Drug Event Data

- Data include a record for each Part D covered prescription filled for beneficiaries enrolled in a Part D plan.

- This includes data for about 25 million of the 44 million Medicare beneficiaries.
  - Many other beneficiaries are covered by the retiree drug subsidy or have other forms of creditable coverage. We do not have PDE records for them.

- About a billion PDE records per year.
Prescription Drug Event Data

- PDE record content (37 elements)
  - Beneficiary link key
  - National Drug Code (NDC) of drug dispensed
  - Script information
    - Date dispensed, quantity dispensed, days supply
  - Cost and payment information
  - Identifiers (encrypted)
    - Prescriber, dispenser, drug plan
  - Linkable “characteristics” files
Prescription Drug Event Data

These PDE data allow the development of measures of drug utilization, cost, safety, etc for the Medicare population that would not be possible previously.
Improved Coding of Race/Ethnicity

Because of the history of how information has been captured, there are significant limitations in the coding of race/ethnicity in current CMS administrative data.

E.g., Studies have shown that only about 30% of Hispanic beneficiaries are identified in the current race variable in the Medicare Enrollment Database
Improved Coding of Race/Ethnicity

- Long-term improvement
  - The Social Security Administration is looking into ways to begin capturing race/ethnicity using the OMB standard

- Short-term improvement
  - Imputation approach based on surname
Improved Coding of Race/Ethnicity

Imputation approach

- Using census data, the Census Bureau has identified the probability that people with a given surname will identify themselves with a given race/ethnic category.
- We have beneficiary name in our administrative data.
- Using the data from the Census Bureau, we associated the probability of being Hispanic with every Medicare beneficiary.
- We imputed beneficiaries as being Hispanic if the probability associated with their surname was above a high threshold.
- Also used data from other sources to help identify Hispanics (e.g., if they requested the Medicare & You Handbook in Spanish).
### Improved Coding of Race/Ethnicity

<table>
<thead>
<tr>
<th></th>
<th>Sensitivity</th>
<th>PPV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current data</td>
<td>0.30</td>
<td>0.95</td>
</tr>
<tr>
<td>Imputation</td>
<td>0.77</td>
<td>0.84</td>
</tr>
</tbody>
</table>
Summary

Over the last couple of years, CMS has made improvements in its Medicare data infrastructure that may be helpful for the future development of the NHQR and NHDR:

- Construction of the Chronic Condition Data Warehouse
- Availability of Part D prescription drug event data
- Improvement in coding of race/ethnicity

We’d be glad to provide whatever additional information about our data that the Committee needs for its deliberations.
Contact Information

Thomas Reilly, Ph.D.
Deputy Director
Office of Research, Development, and Information
Centers for Medicare & Medicaid Services
Phone: 410-786-0631
Email: Thomas.Reilly2@cms.hhs.gov