



Liberating Claims Data: Medicare Claims Public Use Files (PUFs) & Medicare Data Sharing Program

Samuel C. “Chris” Haffer, Ph.D.

Data Development & Services Group

Office of Information Products & Data Analytics - CMS

Baltimore, Maryland

Context for PUF Project

- CMS Data are a National Resource
- Access has been limited
- Transparency/Open Government
- Innovation:
 - CER/PCOR
 - Data Entrepreneurs
 - Information Intermediaries

Pilot Project Goal

While strictly protecting beneficiary confidentiality....

Increase access to CMS claims data through the creation of de-identified data sets ...

While strictly protecting beneficiary confidentiality!!!!

Phase 1: BASIC PUFs

- Challenge

- Create a set of BASIC or slim/thin files
- Comply with the HIPAA “safe harbor” standard
- Test the utility of the files for research

- Outcome

- 14 Basic PUFs – Inpatient, Outpatient, SNF, DME, Hospice, HHA, PDE, Carrier (Reference Years 2008 and 2010)

Phase 2: Profile PUFs

- Challenge

- Increase utility of the files for research
- Comply with the HIPAA “safe harbor” standard
- Use More Data

- Outcome

- 8 Profile PUFs – Chronic Conditions, Prescription Drug, Institutional Provider, Physician/Supplier (Reference Years 2008 and 2010)

Phase 3: Synthetic PUFs

- Challenge

- Unmet Needs of Phase 1 and Phase 2 PUFs
- Comply with the HIPAA “safe harbor” standard

- Outcome

- DE-SynPUF – (20 Samples x 8 Files each) – Inpatient, Outpatient, Carrier, PDE, Beneficiary Summary (Reference Years 2008, 2009, 2010)

Data Entrepreneurs' Synthetic PUF for Medicare Claims Data (DE-SynPUF)

GOAL:

- Increase access to realistic Medicare claims data files in a timely and less expensive manner to spur the innovation necessary to achieve the goals of better care for beneficiaries and improve the health of the population.

USES:

- train researchers on the use and complexity of conducting analyses with CMS claims data prior to initiating the process to obtain access to actual CMS data;
- support safe data mining innovations that may reveal unanticipated knowledge gains while preserving beneficiary privacy; and,
- allow data entrepreneurs to develop and create software and applications that may eventually be applied to actual CMS claims data;

Data Entrepreneurs' Synthetic PUF for Medicare Claims Data (DE-SynPUF)

- New type of 'synthetic' file useful for data entrepreneurs for software and application development
- Preserve detailed data structure of key variables at beneficiary and claim levels
 - Data is fully 'synthetic' for disclosure safety
 - Limited analytic utility due to lack of preservation of interdependence between variables
- Create file that can be certified to be released as a Public Use File

DE-SynPUF Contents

- 5% sample of enrolled Medicare beneficiaries in 2008
- 3 years of claims (2008, 09, 10)
 - Inpatient
 - Outpatient
 - Carrier
 - PDE – Prescription Drugs

DE-SynPUF Description

- The synthetic beneficiaries have nearly the same number of claims as their seeds, with very similar claim patterns
- The data structure is very similar to the real CMS files, albeit with a smaller number of variables
- Programs and procedure designed using the SynPUF should be fully functional when applied to CMS limited data sets

DE-SynPUF Tables

- Beneficiary Table
 - Rows: 2.4 million
 - Variables: 52
- Inpatient Claims Table
 - Rows: 1.8 million
 - Variables: 81
- Outpatient Claim Table
 - Rows: 21.4 million
 - Variables: 76
- Carrier Claims Table
 - Rows: 120 million
 - Variables: 142
- Prescription Drug Event Claims (PDE) Table
 - Rows: 157 million
 - Variables: 8

We want you!!



- Download and Explore
- Conduct Exploratory Research
- Offer Constructive Feedback
- Encourage Others
- Use as a Teaching Resource

Medicare Data Sharing Program: What it is.

A legal authority, allowing Medicare Parts A and B claims data, and Part D event data to be purchased by organizations certified by HHS as meeting certain requirements for use in producing publicly available provider and supplier report cards using measures of quality, efficiency, effectiveness, and resource use.

Medicare Data Sharing for Performance Measurement

NOW



FUTURE



Medicare Data Sharing Program (ACA §10332)

Section 10332 of the Patient Protection and Affordable Care Act (ACA) added a new subsection to Section 1874 of the Social Security Act, requiring that the Secretary:

1. Establish a process to allow for the use of standardized extracts of Medicare Parts A, B, and D claims data by Qualified Entities (QEs)
2. who will evaluate and report on the performance of providers of services and suppliers
3. using measures of quality, efficiency, effectiveness, and resource use.
4. Defines QEs as public or private entities that are determined by the Secretary to be qualified to use Medicare claims data to make such evaluations of provider/supplier performance
5. agree to meet specific requirements regarding the transparency of their methods and
6. their use and protection of Medicare data.

Medicare Data Sharing Program (ACA §10332)

7. Requires Medicare claims extracts be combined with other claims data.
8. Specifies the only use of such data and the derived performance information about providers and suppliers be in reports in an aggregate form.
9. Reports must be released and made available to the public,
10. after first making such reports available to any identified provider or supplier and
11. affording an opportunity to appeal and correct errors.
12. Instructs the Secretary to take such actions as she deems necessary to protect the identity of individual beneficiaries, and
13. authorizes her to establish additional requirements that she may specify for QEs to meet, such as ensuring the security of data.

Why is the Medicare Data Sharing Program IMPORTANT?

- **INCREASES ACCESS TO MEDICARE CLAIMS DATA FOR PERFORMANCE MEASUREMENT**
 - a major component in our CMS' strategy of improving care for individuals, achieving better health for populations, and lowering costs through improvement.
- **HIGHLIGHTS CMS' COMMITMENT TO DATA STEWARDSHIP**
 - protect beneficiary confidentiality
 - minimize risk of accidental data disclosure



Certified Qualified Entities

Oregon Health Care Quality Corporation

Oregon

August 31, 2012

Health Improvement Collaborative of Greater Cincinnati

Ohio
Indiana
Kentucky

August 31, 2012

Kansas City Quality Improvement Consortium

Kansas
Missouri

September 4, 2012

Maine Health Management Coalition Foundation

Maine

November 28, 2012

HealthInsight

New Mexico

January 18, 2013

California Healthcare Performance Information System

California

February 6, 2013

Pittsburgh Regional Health Initiative

Western PA

March 27, 2013



FOR MORE INFORMATION:

Public Use Files

Basic PUFs – www.cms.gov/bsapufs

SynPUFs - go.cms.gov/SynPUFs

Medicare Data Sharing Program

www.cms.gov/QEMedicareData

www.QEMedicareData.org