Liberating Data:
Medicare Claims Public Use Files (PUFs)

Samuel C. “Chris” Haffer, Ph.D.
Data Development & Services Group
Office of Information Products & Data Analytics - CMS
Baltimore, Maryland

Academy Health Annual Research Meeting
June 2013
Liberating Data....Baltimore Style!

CMS Data Access Continuum

Research Identifiable Files

Limited Data Sets

Public Use Files

Medicare Statistical Supplement
Context for PUF Project

- CMS Data are a National Resource
- Access 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!!!!
Statistical Standard
A covered entity may determine that health information is not individually identifiable health information only if: (1) A person with appropriate knowledge of and experience with generally accepted statistical and scientific principles and methods for rendering information not individually identifiable: (i) Applying such principles and methods, determines that the risk is very small that the information could be used, alone or in combination with other reasonably available information, by an anticipated recipient to identify an individual who is a subject of the information; and (ii) Documents the methods and results of the analysis that justify such determination; OR
(2)(i) The following identifiers of the individual or of relatives, employers, or household members of
the individual, are removed: (A) Names; (B) All geographic subdivisions smaller than a State,
including street address, city, county, precinct, zip code, and their equivalent geocodes, except for the
initial three digits of a zip code if, according to the current publicly available data from the Bureau of
the Census: (1) The geographic unit formed by combining all zip codes with the same three initial
digits contains more than 20,000 people; and (2) The initial three digits of a zip code for all such
geographic units containing 20,000 or fewer people is changed to 000. (C) All elements of
dates (except year) for dates directly related to an individual, including
birth date, admission date, discharge date, date of death; and all ages over
89 and all elements of dates (including year) indicative of such age, except that such ages and
elements may be aggregated into a single category of age 90 or older; (D) Telephone numbers; (E)
Fax numbers; (F) Electronic mail addresses; (G) Social security numbers; (H) Medical record
numbers; (I) Health plan beneficiary numbers; (J) Account numbers; (K) Certificate/license numbers;
(L) Vehicle identifiers and serial numbers, including license plate numbers; (M) Device identifiers and
serial numbers; (N) Web Universal Resource Locators (URLs); (O) Internet Protocol (IP) address
numbers; (P) Biometric identifiers, including finger and voice prints; (Q) Full face photographic images
and any comparable images; and (R) Any other unique identifying number, characteristic, or code;
and (ii) The covered entity does not have actual knowledge that the information could be used alone
or in combination with other information to identify an individual who is a subject of the information.
Team

- IMPAQ International
- NORC @ University of Chicago
- Buccaneer Computer Systems & Service, Inc. – A GDIT Company
- George Washington University
- Expert Consultants: Beth Virnig, Marshall McBean
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)
We want you!!

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

Public Use Files

Basic PUFs – www.cms.gov/bsapuufs
SynPUFs - go.cms.gov/SynPUFs
Never Forget the Importance of the Work You Do!

"I heard a definition once:
'Happiness is health and a short memory.'

I wish I invented it because it's very true."

Audrey Hepburn