

Creation of Public Use Files: Lessons Learned from the Comparative Effectiveness Research (CER) Public Use Data Pilot Project

Erkan Erdem, PhD, IMPAQ International, LLC, Columbia, MD, USA

Sergio Prada, PhD, IMPAQ International, LLC, Columbia, MD, USA

American Evaluation Association, Anaheim, California

November 5, 2011



CER PUF Project

- ❑ Produce and disseminate public use files (PUFs) from Medicare claims data sets for comparative effectiveness research (CER)
 - Transparency
 - Open Government
 - Increase access to Medicare data through de-identified PUFs

Basic Stand Alone (BSA) PUFs

- ❑ Create a set of eight BSA PUFs
 - One PUF for each type of service (e.g., inpatient)
 - Comply with HIPAA Safe Harbor
 - De-identify to protect beneficiary confidentiality
 - Also, protect individual providers in some PUFs
 - Balance between privacy protection and analytic utility

BSA PUFs

Eight BSA PUFs using Fee-for-Service claims

- Inpatient
- Outpatient
- Skilled Nursing Facilities (SNF)
- Home Health Agencies (HHA)
- Hospice
- Carrier (Physician/Supplier)
- Durable Medical Equipment (DME)
- Prescription Drug Events (PDE)

Available at www.cms.gov/BSAPUFS

Environmental Scan

- ❑ Stakeholders interviews
 - Experts on de-identification
 - Health information privacy experts
 - Governmental representatives
- ❑ Six case studies of de-identified individual-level PUFs
- ❑ Literature review and legal analysis
 - Methods for data access (PUF, Data Enclave, Remote Data Center)
 - Identify statutory, regulatory and technical barriers related to the creation and dissemination of the data

Environmental Scan: Findings

Literature review

- Statistical Disclosure Limitation (SDL) methods
- Disclosure risk measures
- Analytic utility measures

Case studies

- No unique recipe for success
- Statistical Policy Working Paper #22 Checklist by Federal Committee on Statistical Methodology widely used

Environmental Scan: Findings

- ❑ Recommendations for new initiatives
 - Limited scope at the beginning
 - Focus on information rather than data
- ❑ Access
 - Provide tiered data access
- ❑ Outstanding challenges
 - Increasing availability of personally identifiable data
 - Need for expertise in risk modeling

Needs Assessment

- ❑ Review of what researchers need in Medicare datasets
 - Public sector
 - University/Academia
 - Private/Non profit
- ❑ Needs assessment allowed us to
 - Assess the usefulness of BSA PUFs
 - Develop an understanding of acceptable de-identification methods
 - Formulate future PUFs

Needs Assessment: Findings

- ❑ Existing data access procedure
 - Long and costly
 - Hinders small projects and research on rare diseases
- ❑ Use of BSA PUFs
 - Descriptive stats, research design, hypothesis testing
- ❑ Future PUFs useful for CER provided that
 - Sufficient set of variables is included
 - Minimal perturbation employed

Needs Assessment: Findings

- ❑ Recommendations for PUFs
 - Set appropriate expectations for the files
 - Avoid perturbation as researchers demand precise and unbiased estimates
 - Require a simple & limited DUA
 - Ask users to provide some information and agree with terms
 - Provide multiple levels of data (tiered access) with differing restrictions and type of data including actual data

Steps in BSA PUFs

□ Source data

- Disjoint 5% sample of Medicare beneficiaries for each PUF from 2008 Beneficiary Summary File

□ Steps to create the BSA PUFs

- Choice of fields to be included in each PUF with priority given to clinical information such as diagnosis and procedures
- Data cleaning and preparation
- De-identification
- Creation of PUF
- Re-identification testing and statistical certification
- Preparation of documentation, codebook, etc.
- Dissemination

De-identification of BSA PUFs

❑ De-identification step 1:

- Coarsening, categorization, aggregation, or rounding
 - Ex: Age categories, 3-digit ICD-9 diagnosis codes, rounded payment amounts

❑ De-identification step 2:

- Suppression of records based on count of beneficiaries (and providers) for unique combinations (or cells) of all variables in the PUF
- Safe if there are at least 11 beneficiaries (and 11 providers) in a given cell in the full (100%) Medicare population.
- Goal of less than 10% of suppression rate

Lessons Learned from BSA PUFs

- ❑ Each type of care requires a unique treatment
 - Variation in # of claims per beneficiary
 - Variation in dimensionality of clinical data (e.g., drug code)
 - Variation in payment pattern (e.g., payment link to DRG)
 - Variation in number of providers
- ❑ Rule of 11 minimum beneficiaries and providers simplified choice of risk criteria
 - Avoid perturbation as researchers demand precise and unbiased estimates
- ❑ Access to actual data necessary

Checklist for PUFs

- Review underlying laws and regulations
- Contact stakeholders for viability
- Identify appropriate SDL techniques for the content
- Use samples for added protection
- Define a risk measure, risk assessment method, and utility loss metric
- Do not make SLD techniques public
- Prepare detailed documentation and data dictionary
- Invest in dissemination methods such as briefs, webinars, dashboards, and challenges

New PUFs – Chronic Conditions

- Profile-level file “table” with information from the full Medicare population
 - Profiles defined by age, gender, chronic conditions, and dual-eligibility
 - Averages of Medicare payments and utilization for each profile

Dissemination

- ❑ Download BSA Medicare Claims PUFs
<http://www.cms.gov/BSAPUFS>
- ❑ Medicare Claims PUFs Challenge
<http://www.health2challenge.org/2011/06/01/medicare-claims-data/>
- ❑ Academy Health Medicare Claims PUF Webinar Series
<http://www.academyhealth.org/Training/ResourceDetail.cfm?itemnumber=7097>
- ❑ IMPAQ Chronic Conditions PUF Dashboard
<http://www.impaqint.com/project-showcase/public-use-data-project-chronic/default.html>

Appendix

Creation of Public Use Files: Lessons Learned from the Comparative Effectiveness Research (CER) Public Use Data Pilot Project

Erkan Erdem, PhD, IMPAQ International, LLC, Columbia, MD, USA

Sergio Prada, PhD, IMPAQ International, LLC, Columbia, MD, USA

American Evaluation Association, Anaheim, California

November 5, 2011



BSA PUFs

Inpatient	Part D Event	Hospice	DME
Claim-level	Event-level	Beneficiary-level	Line item-level
Age	Age	Age	Age
Gender	Gender	Gender	Gender
Base DRG code	Drug name	Terminal diagnosis	ICD-9-CM diagnosis code
ICD-9-CM primary procedure code	Drug strength	Cancer indicator	HCPCS procedure code
Medicare quintile average payment amount	Unit of strength of the drug	Covered days	Count of supplies or services
Length of stay	Dose form	Payment by Medicare	Payment by Medicare
	Drug class		
	Quantity dispensed		
	Days supply		
	Total drug cost		
	Payment by beneficiary		
	Drug type		

BSA PUFs (cont.)

Outpatient	SNF	HHA	Carrier
Procedure-level	Beneficiary-level	Beneficiary-level	Line item-level
Age	Age	Age	Age
Gender	Gender	Gender	Gender
ICD-9-CM diagnosis code	Number of admissions	Number of admissions	ICD-9-CM diagnosis code
HCPCS procedure code	Number of days of Rehabilitation Services	Number of therapy visits	HCPCS procedure code
Count of service	Number of days of Rehabilitation Plus Extensive Services	Number of skilled nursing care visits	Count of service
Payment by Medicare	Payment by Medicare	Number of home health aide visits	Payment by Medicare
		Payment by Medicare	Place of Service
			Type of Service
			Provider Type

Creation of Public Use Files: Lessons Learned from the Comparative Effectiveness Research (CER) Public Use Data Pilot Project

Erkan Erdem, PhD, IMPAQ International, LLC, Columbia, MD, USA

Sergio Prada, PhD, IMPAQ International, LLC, Columbia, MD, USA

American Evaluation Association, Anaheim, California

November 5, 2011

