CMS 2008 Basic Stand Alone (BSA) Prescription Drug Events (PDE) Public Use File (PUF)

Frequently Asked Questions (FAQ)

1. What is the *CMS 2008 BSA PDE PUF*?

The *CMS 2008 BSA PDE PUF* is a free downloadable file containing a subset of the information contained on prescription drug events provided to a 5% sample of 2008 Medicare beneficiaries. Each of the 50,325,988 records in the file pertains to one prescription drug event. Each record includes:

1. Gender of the beneficiary;
2. Age of the beneficiary at the end of 2008, reported as (1) under 65 years of age; (2) 65 to 69; (3) 70 to 74; (4) 75 to 79; (5) 80 to 84; (6) 85 and older;
3. Drug name;
4. Drug strength and units of drug strength;
5. Drug dose form;
6. Drug class;
7. Quantity dispensed;
8. Days supply;
9. Total drug cost;
10. Indicator for how much the patient paid (categorized);
11. Drug type.

2. How was this PUF created?

The *CMS 2008 BSA PDE PUF* originates from a disjoint 5% random sample of beneficiaries from the 100% Beneficiary Summary File for 2008. To exclude any overlap with the beneficiaries in the existing 5% CMS research sample,[1] the beneficiaries in that other sample were excluded, and a 5-in-95 random draw was made of the remaining 95% of beneficiaries. All prescription drug events for the selected 5% of beneficiaries were then included in the sample from which the *CMS 2008 BSA PDE PUF* was developed.

The selected events were subjected to a thorough de-identification process. The methods used to protect the identity of beneficiaries are described in the answer to the next question.

[1] http://www.resdac.org/tools/TBs/TN-011_How5percentMedicarefilescreated_508.pdf

3. What has been done to protect the privacy of Medicare beneficiaries?

Of paramount importance in the release of the PUF is the protection of beneficiary confidentiality. To that end, all directly identifiable information has been removed in accordance with the HIPAA Privacy Rules.

Other important steps were taken:

- Only a small subset of possible variables was selected for inclusion in the file. This reduced the possible information that could be used to identify the beneficiaries included in the new 5% sample.
- For the variables selected for inclusion, categorization was used to protect identities. For example, in place of date of birth or current age in years, the file was created with age categorized into seven intervals: (1) under 65 years of age; (2) 65 to 69; (3) 70 to 74; (4) 75 to 79; (5) 80 to 84; (6) 85 and older. This categorization allows researchers to differentiate patterns in other data (e.g., in the frequency of a particular diagnosis at hospital admission) between younger and older beneficiaries but not to use age or date of birth as a highly identifying variable.
- The final protection was provided by excluding about 5% of records from the final PUF, those for which the combination of values for all eleven variables in the file were extremely uncommon in the Medicare population. No combination that occurred for fewer than 11 beneficiaries in the full Medicare population was allowed into the final PUF. This criterion tended to exclude events with uncommon drugs.

4. How was provider confidentiality protected?

There is no risk of provider identification in the *CMS 2008 BSA PDE PUF* as the PUF does not contain any information about individual providers.

5. What data cleaning steps were performed to obtain the initial 5% PDE sample?

The 5% random PDE sample was cleaned by removing events with data anomalies (e.g., zero drug cost or days supply and patient payment greater than total drug cost). De-identification procedures were then performed on the resulting initial 5% PDE sample (see FAQ #3 above).

6. What is a prescription drug "event"?

An event is similar to a claim in other types of care under Medicare. Each prescription including refills generates a prescription drug event in the PDE database. Each record in the CMS 2008 BSA PDE PUF is one such occurrence.

7. Can I know which events belong to the same Medicare beneficiary?

The *CMS 2008 BSA PDE PUF* does not allow users to link multiple events on the file for those beneficiaries with more than one event in 2008. The record identification field on the PUF contains a new series of random numbers generated just for the *CMS 2008 BSA PDE PUF* and used to sort events in a random order. Users wishing to work with a subsample of the line items on the file can use the record identifier to draw a random subset of records.

8. Why are some drug information not available in the *CMS 2008 BSA PDE PUF*?

The initial 5% sample of prescription events is at the detailed National Drug Code (NDC) level. This file is merged with the RxNorm database of the National Institutes of Health to find out the information on the drug (name, strength, dose form, etc.). It is also merged with the Veterans Affairs National Drug File (VA-NDF) to find out the class of the drug. If an NDC is not found in one of these databases, then the relevant information is not available (i.e., missing) in the *CMS 2008 BSA PDE PUF.* Please refer to General Documentation for additional information.

9. Most variables are represented as codes in the file. How can I use them?

All necessary information is provided together with the *CMS 2008 BSA PDE PUF.* A data user must download all of the components. Data User's Guide is a SAS read-in file which would create a SAS data file with the descriptions of the coded variables. If you use another software (e.g., STATA), please download all the lookup files for drug name, strength, etc. and merge them with the *CMS 2008 BSA PDE PUF* individually to bring in the descriptions of these variables.

10. How is the *CMS 2008 BSA PDE PUF* different from the 5% CMS standard research sample?

There is no overlap in terms of beneficiaries between the 5% CMS standard research sample and the *CMS 2008 BSA PDE PUF*. These two 5% samples are disjoint.

11. What are the limitations of the *CMS 2008 BSA PDE PUF*?

The *CMS 2008 BSA PDE PUF* is intended to give researchers a convenient initial look at data drawn from CMS Prescription drug events. The file contains measures of demographic characteristics of beneficiaries, information on the drug (drug name, strength, dose form, class, etc.) and total drug cost. In order to preserve confidentiality, suppression criteria have been applied to variables and beneficiaries on the initial file. Some variables are rounded or categorized. Researchers should read the General Documentation and the Data Dictionary and Codebook to determine the appropriateness of the PUF for addressing specific research questions.

12. How may I request additional data?

See the Files for Order section of the CMS Web site http://www.cms.gov/home/rsds.asp. This site lists available CMS data files, data file properties, information about data-use agreements, as well as ordering and payment information.

13. What is the plan for future data releases?

The *CMS 2008 BSA Inpatient Claims PUF* was released in February 2011. The *CMS 2008 BSA PDE PUF* is released together with the *CMS 2008 BSA Hospice Beneficiary PUF* and *CMS 2008 BSA DME Line Items PUF.* The current plan is to release additional 2008 Basic Stand Alone (BSA) PUFs in 2011. These PUFs will be based on some or all of the following files: Skilled Nursing Facility (SNF) claims, Outpatient claims, Physician/Supplier claims, and Home Health Agency claims.

14. How may I provide feedback on the CMS 2008 BSA PDE PUF?

Questions and comments can be submitted to Research Data Assistance Center (http://www.resdac.org/) via resdac@umn.edu or 1-888-9RESDAC.