CMS 2008 Basic Stand Alone (BSA) Inpatient Claims Public Use File (PUF) Frequently Asked Questions (FAQ)

1. What is the *CMS 2008 BSA Inpatient Claims PUF*?

The *CMS 2008 BSA Inpatient Claims PUF* is a free downloadable file containing a small but important subset of the information contained on hospital claims for the inpatient services provided to a 5% sample of 2008 Medicare beneficiaries. Each of the 588,415 records in the file pertains to one inpatient stay. 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; and (6) 85 or older;
3. Base Diagnosis Related Group (DRG) for the inpatient stay (defined below);
4. Primary procedure performed during the inpatient stay (also defined below);
5. Number of days the patient was hospitalized during the stay (length of stay), reported as (1) one day; (2) two to four days; (3) five to seven days; and (4) eight days or more;
6. Average Medicare payment for low-, medium-, and high-payment stays within the DRG (defined below).

2. Why was the PUF created?

CMS is committed to increasing access to its Medicare claims data. The PUF was created as part of an effort to provide such access. Other similar files are planned for release in coming months. This PUF is made available as a free download in order to give researchers access to Medicare claims data without the time and cost associated with data files which require more restricted access. Even though this PUF is available to researchers as a free download in CSV format, this format is easily converted into other formats (code for creating SAS datasets is part of the documentation).

3. How was this PUF created?

The *CMS 2008 BSA Inpatient Claims 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 inpatient claims for the selected 5% of beneficiaries were then included in the sample from which the PUF was developed.

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

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

4. 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 six intervals: (1) under 65 years of age; (2) 65 to 69; (3) 70 to 74; (4) 75 to 79; (5) 80 to 84; and (6) 85 or older. This categorization allows researchers to differentiate patterns in other data (e.g., in the frequency of a particular diagnosis at hospital admission) between claims of 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 8% of records from the final PUF, those for which the combination of values for all six variables in the files 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 claims from uncommon DRG categories and uncommon primary procedure codes. However, the criterion also assured that no record in the PUF could be linked to a particular beneficiary no matter how much information a user knew about any beneficiary.

5. How was provider confidentiality protected?

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

6. What is a Base DRG?

Base DRG codes are derived from the Medicare Severity Diagnosis Related Groups (MS-DRGs) that are used by Medicare for setting payment rates for inpatient care. However, Base DRG codes do not differentiate between inpatient stays with major complication or comorbidity (MCC), with a complication or comorbidity that is not major (CC), or with no complication or

comorbidity (w/o CC/MCC). For example, MS-DRG codes 291, 292, and 293 indicate “heart failure and shock w MCC”, “heart failure and shock w CC”, and “Heart failure and shock w/o CC/MCC”. The Base DRG code combines the three MS-DRG categories into one code: “heart failure and shock”.

7. What is a two-digit ICD-9 primary procedure code?

The PUF’s primary procedure codes provide only the first two digits of the ICD-9 primary procedure code on Medicare inpatient claims. The original ICD-9-CM (Volume 3) procedure codes are composed of codes with either 3 or 4 digits. The PUF’s codes, with two digits, provide information on the heading of a category, whereas the third and/or fourth digits (those to the right of the decimal point in the ICD-9 codes) provide greater detail.

8. How is Medicare payment amount represented in the PUF?

Data from 100% of Medicare inpatient claims were used to generate the payment data in the PUF. All inpatient claims in each Base DRG were sorted by payment amount (low to high), and the claims were divided into five approximately equal groups (quintiles; if the 20% line fell in the middle of claims with the same payment amount, the dividing line was moved so the groups above and below were as close as possible to 20%). The payment amounts for each approximate quintile within the DRG were then averaged, and the resulting figures were used as the payment amount reported for all corresponding claims in the DRG in the PUF.

It is uncommon for census data to be used in preparation of a PUF. The use of census data to calculate average Medicare payment by quintile by DRG makes the payment information in the PUF particularly accurate for calculation of average Medicare payment amount by DRG.

9. Can I know which claims belong to the same Medicare beneficiary?

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

10. How is the CMS 2008 BSA Inpatient Claims 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 Inpatient Claims PUF*. These two 5% samples are disjoint.

11. What are the limitations of the CMS 2008 BSA Inpatient Claims PUF?

This PUF is intended to give researchers a convenient initial look at data drawn from CMS inpatient claims records. The file contains measures of demographic characteristics of beneficiaries, diagnosis, primary procedure performed, length of stay, and Medicare payment amount. In order to preserve confidentiality, suppression criteria have been applied to variables and claims 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 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: Part D Event Utilization, Skilled Nursing Facility (SNF) Claims, Hospital Outpatient Claims, Durable Medical Equipment (DME) Claims, Physician/Supplier Claims, Hospice Claims, and Home Health Claims.

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

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