



Differences in How the Medicare 5% Files are Generated

The Research Data Assistance Center (ResDAC) is a CMS contractor that provides free assistance to academic and non-profit researchers interested in using Medicare and/or Medicaid data for their research. ResDAC is staffed by a consortium of epidemiologists, public health specialists, health services researchers, biostatisticians, and health informatics specialists from the University of Minnesota.

*TN-011
March 2007
Authors: Kelly Merriman, Faith M. Asper*

In 2005, CMS hired two contractors to process and distribute CMS data. The Research Data Distribution Center (RDDC) began processing and distributing CMS Research Identifiable Files, Limited Data Sets, and Public Use Files. The Iowa Foundation for Medical Care (IFMC) creates and distributes Medicare data for the study of Chronic Conditions, called the Chronic Condition Data Warehouse (CCW).

The process used by the RDDC and IFMC to create the 5% sample files varies slightly from the method CMS has used to create the 5% files in the past. Although CMS, the RDDC, and IFMC use the same terminal digits of the HIC to create the 5% beneficiary sample files, namely 05, 20, 45, 70, or 95 in position 8 and 9 of the Health Insurance Claim (HIC) number, the extraction method varies. HIC changes that result in different 8th and 9th digits occur very infrequently; empirical testing has estimated changes to be less than 0.5% of HICs per year.

RDDC 5% Files

The RDDC identifies the 5% sample from the 100% CMS Denominator file first. In other words, the RDDC identifies those beneficiaries in the 100% Denominator file that have 05, 20, 45, 70, or 95 in positions 8 and 9 of the HIC number. Then, after identifying the cohort from the Denominator file, they crosswalk the HIC to the RDDC-Bene ID associated with that beneficiary. The RDDC-Bene ID is a unique identifier for a beneficiary created for processing and privacy purposes by the RDDC. However, the assignment of BIDs is unique for each data shipment, so a BID for a given beneficiary will vary by research project. Finally, using the RDDC-Bene ID, the RDDC extracts the claims data. This extract methodology extracts all claims for all beneficiaries with a HIC in the 5% Denominator file, even if the HIC on the claim is not the same.

This is slightly different from the way CMS has generated the 5% sample files in the past. The method CMS has used is to extract claims and Denominator records directly based on selecting records with 05, 20, 45, 70, or 95 in position 8 and 9 of the HIC. No finder file was created.

Because the Denominator file is created and frozen in March of the following year, there may be claims for people who don't have a Denominator file record. This can happen if the beneficiary changes HICs sometime during the reference year with the

new HIC not being in the 5% sample. Then claims records for the first part of the year, when the HIC with 05, 20, 45, 70, or 95 in the 8th and 9th positions was used, would be included in the 5% claims files. But for the second part of the year, when the beneficiary's newly changed HIC no longer satisfies the criteria for the 5% sample, no claims for this beneficiary would be included in the 5% claims files and, since the beneficiary's HIC does not satisfy the 5% criteria in March of the following year, he/she would not be included in the 5% Denominator file for that year. Note that HIC changes that result in different 8th and 9th digits occur very infrequently, estimated to be less than 0.5% of HICs per year.

For example, if a beneficiary uses the HIC, 123456789B, from January through April and then uses the HIC, 234567895A, from May to December, only those records from May through December were included in the CMS 5% files (in this example, the HIC has "95" in positions 8 and 9 from May to December). Comparatively, the RDDC 5% files, will contain records for this beneficiary for the entire year.

IFMC's CCW 5% Enhanced Files

The Iowa Foundation for Medical Care (IFMC) generates the CCW data based on the 5% sample of beneficiaries. IFMC identifies the 5% sample from the Beneficiary Summary File, which is the CCW version of the Denominator file. In the CCW Beneficiary Summary file, the last two digits of the Claim Account Number (CAN) are in the set {05, 20, 45, 70, 95}, and each beneficiary has Part A or B coverage for at least one month and is alive for at least one day during the year. Once a beneficiary enters the sample, the beneficiary will stay in the sample from that time forward. The result of this methodology is that all claims for all beneficiaries with a HIC in the Beneficiary Summary File are extracted, even if the HIC on the claims differs from the HIC in the CCW Beneficiary Summary file.

This method is slightly different from that used by CMS to generate 5% sample files in the past, but is the same as how the RDDC creates the RDDC 5% files. So, for a description of how this sample differs from the CMS 5% sample, please refer to the above section.

For a comparison of the CCW 5% Enhanced files and the CMS standard 5% files, please see the following grid:

Table 1 : Differences between the 5% Sample Files

	CMS Standard 5% Sample Files	CCW 5% Files	Enhanced CCW 5% Files
Distributor	CMS & Research Data Distribution Center (RDDC)	Same as Enhanced CCW 5% files	Iowa Foundation for Medical Care
Years Available	1991-1996 (CMS); 1997-2005 (RDDC)	Same as Enhanced CCW 5% files	1999-2005
Time periods available	Available as Calendar Year files	Same as Enhanced CCW 5% files	Available as a Calendar Year file or available by Quarter.
File Cost	Approximately \$18,000/year for all SAFs and Denominator	Same as Enhanced CCW 5% files	Approximately \$7,000/year for all SAFs, Denominator, & Assessments
File Format	1991-1996: EBCDIC; 1997-current: SAS® or csv text	Same as Enhanced CCW 5% files	1999 - current: SAS® or ASCII text
Media Available	1991-1996: 3490e compressed IBM Standard Label Tape Cartridges; 1997-current: CD, DVD, USB Hard Drive	Same as Enhanced CCW 5% files	1999-current: CD, DVD, USB Hard Drive
Unique Identifier	1991-1996: DESY Sort Key or HIC; 1997-current: RDDC Bene Id - This identifier is unique to beneficiary and to the DUA.	Same as Enhanced CCW 5% files	CCW Bene Id - This identifier is unique to the beneficiary and to the DUA.
Sample Selection	Denominator: Any beneficiary where the last two digits of the CAN are in the set {05, 20, 45, 70, 95} SAF: Any claim in reference year where CAN is in the set {05, 20, 45, 70, 95}	Denominator File: Benes with CAN in the set {05, 20, 45, 70, 95} as of March of (reference year + 1), with Part A or B coverage for at least one month and alive at least one day during reference year.	Claim Account Number (CAN) with last two digits in the set {05, 20, 45, 70, 95} prior to September 2005, with Part A or B coverage for at least one month and alive at least one day during reference year; once a beneficiary enters the sample, the beneficia
HIC changes	If a beneficiary has experienced a change in his/her CAN and the last two digits of the CAN are NOT 05, 20, 45, 70, 95, then the beneficiary will be removed from the sample.	Same as CMS Standard 5% files	All claims and eligibility information from 1999 forward for this 5% sample are included in the CCW. This means that if a beneficiary had a change in hid/her HIC and the change caused this beneficiary to fall out of the standard CMS 5% sample, the CCW wil

Table 1 : Differences between the 5% Sample Files continued

	CMS Standard 5% Sample Files	CCW 5% Files	Enhanced CCW 5% Files
Distributor	CMS & Research Data Distribution Center (RDDC)	Same as Enhanced CCW 5% files	Iowa Foundation for Medical Care
Files available: Claims, Enrollment, Assessment	Denominator, Inpatient, Skilled Nursing Facility, Outpatient, HHA, Hospice, Carrier, and DME SAFs, MedPAR short-stay/long-stay/snf stay	Same as Enhanced CCW 5% files	Denominator, Beneficiary Summary File, Chronic Conditions Summary File, Inpatient, Skilled Nursing Facility, Outpatient, HHA, Hospice, Carrier, DME, Minimum Data Set (MDS), Outcome Assessment and Information Set (OASIS), Inpatient Rehabilitation Facility
Claim File Variables	All	Same as Enhanced CCW 5% files	The CCW data files are a subset of the CMS Standard 5% sample files. Only variables that were useful for research were kept in the files. (For example, CMS internal processing variables were excluded or variables that were null.) Some of the Beneficiary
Linking Process to Other CMS Files	Files link to other CMS files using the Enterprise Cross Reference process to match. The CMS ECR process uses beneficiary identifiers such as name, address, date of birth, and gender to match a person across files.	Same as Enhanced CCW 5% files	Enhanced CCW 5% files link to other CMS files using the Enterprise Cross Reference process to match. The CMS ECR process uses beneficiary identifiers such as name, address, date of birth, and gender to match a person across files.
Linked Files	Claims and Denominator files after 1997 can be linked to the Minimum Data Set (MDS), OASIS, IRF-PAI and Swing Beds.	Same as Enhanced CCW 5% files	Claims and Enrollment files from 1999-forward can be linked to MDS, OASIS, IRF-PAI, and Swing Beds.
Beneficiary's Chronic Conditions	Annual Person Summary File will include 21 chronic condition flags	Same as Enhanced CCW 5% files.	Chronic Conditions Summary File contains a flag to indicate which of the 21 pre-defined chronic conditions the beneficiary has according to the algorithm. The algorithms do have certain criteria as to whom will be included in the search. (i.e. No beneficiaries with Medicare managed care, only those with Part A&B coverage.) See User Manual for details.

If you have any questions or comments, ResDAC staff can be contacted at 1-888-ResDAC or resdac@umn.edu

Internet Citation:

Differences in How the Medicare 5% Files are Generated. Technical Brief, ResDAC Publication Number TN-011, March 2007. Research Data Assistance Center, University of Minnesota, Minneapolis, MN. <http://www.resdac.umn.edu>.

Technical Document authored by Kelly Merriman and Faith M. Asper, ResDAC staff.

Copyright © 2008 Regents of the University of Minnesota. All Rights Reserved.

The analyses upon which this publication is based were performed under Contract Number 500-01-0043, entitled, "Research Data Assistance Center (ResDAC)", sponsored by CMS, Department of Health & Human Services.