**Title:** SAF Construction Documentation

**Author:** Brendan Rabideau

**Date Created:** 12/07/2016

**Updated on:** 6/15/2017

**Objective:** To give an overview of the process used to create the standard analytic files for each condition of interest for the Hospital Readmissions Reduction Program analysis (Nuckols, 50367), Vertical Integration project (Sood), and Hospital and PAC Productivity analysis (Romley). This document chronicles the core programs that go into generating the SAFs beginning with the raw Medicare data. Included are descriptions of the main purposes of each program, as well as descriptive information of the important raw, intermediate, and final datasets as of the above ‘Updated On’ date.

**Notes:**

* Raw data found here:
  + /schaeffer-b/sch-protected/VERTICAL-INTEGRATION/Data
* Intermediate and final data found here:
  + /schaeffer-b/sch-protected/from-projects/VERTICAL-INTEGRATION/rabideau/Data
  + /schaeffer-b/sch-protected/from-projects/VERTICAL-INTEGRATION/rabideau/Data/MDS
* Program found here:
  + /schaeffer-b/sch-protected/from-projects/VERTICAL-INTEGRATION/rabideau/Programs/MainAnalysis
  + /schaeffer-b/sch-protected/from-projects/50367\_Nuckols/rabideau/Programs/HRRP

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Program:** collapse\_mds\_all.do  
**Inputs:** Raw MDS2.0 and MDS3.0 (2006-2013; n=130,041,719)  
**Outputs:** mds\_stays\_2006\_2013\_1.dta (n= 31,169,306)  
**Purpose:** This code takes the assessment level MDS2.0 and MDS3.0 and groups and collapses observations into a stay level file. The key variables to retain from this file are admission date, discharge date, provider ID, and beneficiary ID.   
**Notes:** stat transfer to .sas7bdat. See MDS memo

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Program**: 00\_Assign\_Macro\_Variables\_and\_Libraries.sas  
**Inputs:** None  
**Outputs:** None  
**Purpose:** This program sets all of the libname statements and key macrovariables that will be invoked in each subsequent program. This makes updating these programs much more manageable because these values are not hardcoded in dozens of programs throughout the algorithm, but instead are all in 1 location. Some key macrovariables that are set here are the condition-cohort selection criteria (&diagselect\_dx), the condition-cohort that is being looked at on this run (&dx), start and end years (&YY and &YYE), and raw/intermediate/final libnames for the data.   
**Notes:** Most times you’ll just need to update &DX to make the algorithm run with a new condition cohort. It is one of the few things that need to be manually modified to make the algorithm work. Note that &DX must match the name of the DX\_readmission.sas program invoked in the shell script ‘algorithm\_inputs’ (e.g. Let DX = HF in the 00\_Assign program; hf\_readmission\_V2014\_short.sas in the shell script)

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Program:** Gen\_pta\_base\_datasets.sas  
**Inputs:** Raw MedPAR (2006-2013; n=133,550,584), HHA (2006-2013; n=113,780,196), mds\_stays\_2006\_2013\_1 (n=31,169,306) **Outputs:** pta\_in\_base\_dataset\_pac.sas7bdat (n=113,780,196) pta\_in\_base\_dataset.sas7bdat (n=101,174,145) **Purpose:** To read in the raw data, standardize names across years, identify different facility types (STA, IRF, SNF, MDS, LTC, HHA), and separate out observations that will form the base of our index admissions dataset and our post-index dataset. Removes claims from US territories and non-STACHs from the dataset that will become our index admission sample. **Notes:**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Program:** Gen\_bene\_dataset.sas  
**Inputs:** Raw denominator and MBSF (2006-2013; n=399,259,818) **Outputs:** bene\_dataset.sas7bdat (n=69,550,495) **Purpose:** To read in raw data, standardize names across years, and output a file with demographic information that is unique at the beneficiary level. **Notes:**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Program:** Gen\_cov\_dataset.sas  
**Inputs:** Raw denominator and MBSF (2006-2013; n=399,259,818) **Outputs:** coverage13\_12.sas7bdat (n=69,550,494) **Purpose:** To read in raw data, standardize names across years, and output a file with coverage information that is unique at the beneficiary level. In the event of multiple years per beneficiary, this program flattens out the files and creates 12 monthly coverage flags per year. For example, Y12BUY1, Y12BUY2… Y13BUY12 are variables that indicate coverage from 01/01/12 to 12/31/13.  **Notes:**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Program:** 00\_build\_stay\_dataset.sas  
**Inputs:** pta\_in\_base\_datasaet.sas7bdat (n=101,174,145) **Outputs:** stay\_dataset.sas7bdat (n=101,144,868) **Purpose:** To create a short-term acute care hospital, stay-level file from raw claims that will form the basis of our index stay dataset. This program collapses multiple claims from the same stay (if we start with inpatient data, though since we start with MedPAR data this is largely irrelevant), removes duplicate stays and stays longer than 365 days. **Notes:**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Program:** 00B\_build\_stay\_dataset\_all\_simple.sas  
**Inputs:** stay\_dataset (n=101,144,868) pta\_in\_base\_dataset\_pac.sas7bdat (n=113,780,196) **Outputs:** stay\_dataset\_all.sas7bdat (n=214,925,064) **Purpose:** Begins to build the stay-level dataset that will become our post-index file. This program essentially appends together our STACH stays that might be readmissions, as well as all of our PAC stays into a single, long-form file where the observation is at the bene-stay-facility level. **Notes:**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Program:** 0A\_create\_flag\_transfers\_dataset.sas  
**Inputs:** stay\_dataset.sas7bdat (n=101,144,868) **Outputs:** flag\_transfers.sas7bdat (n=4,092,758) **Purpose:** To flag stays with a discharge and subsequent admission to a different STACH hospital within 1 day of discharge from the original hospital. This program flags an entire transfer chain in the event of multiple transfers, not just the hospitalization immediately before the final hospitalization. **Notes:**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Program:** 1\_Create\_initial\_index\_with\_transfers.sas (FROM HERE ON WE WILL USE THE PN SAMPLE FOR OUR N’s)  
**Inputs:** stay\_dataset.sas7bdat (n=101,144,868) flag\_transfers.sas7bdat (n=4,092,758) mds2\_2006\_2010.sas7bdat (n=18,777,323), mds3\_2010\_2013.sas7bdat (n=11,333,799) **Outputs:** index01.sas7bdat (n=11,548,556 (PN)) **Purpose:** To filter our dataset to include only index stays that are relevant to our condition of interest, within our reporting period, and who meet the appropriate transfer criteria. Appropriate transfer criteria means that the final hospitalization in the transfer chain must be for our condition of interest. This program will also flag the non-final hospitalizations in a transfer chain so we do not drop out our index stay because a prior hospitalization for a transfer occurred within 30 days of it (see DX\_readmission.sas for this criteria).   
**Notes:**When parsing down our dataset to only our conditions of interest, we use a wider range of diagnosis codes than what goes into the final definition of our condition. I’m not entirely sure why this is, but it came included with the CMS code. I assume this is so that we can have more flexibility in defining our cohort without having to rebuild all of the data from scratch. So while we drop many observations here for not being a condition of interest, we will drop more observations later for not being a narrower definition of our condition of interest it (see DX\_readmission.sas for this criteria).

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Program:** 2\_add\_Bene\_data.sas  
**Inputs:** index01.sas7bdat (n=11,548,556) bene\_dataset.sas7bdat (n=69,550,495) **Outputs:** index02.sas7bdat (n=11,548,453) **Purpose:** Merge on demographic data, including date of death. Keep only if in both datasets **Notes:**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Program:** 3\_add\_coverage.sas  
**Inputs:** index02.sas7bdat (n=11,548,453) coverage13\_12.sas7bdat (n=69,550,494) **Outputs:** index03.sas7bdat (n=11,548,453) **Purpose:** Merge on coverage information and create flags for continuously covered for the appropriate number of months pre and post-index. Keep only if in both datasets **Notes:**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Program:** 4\_bundled\_index\_cases.sas  
**Inputs:** index03.sas7bdat (n=11,548,453) **Outputs:** indexMatch.sas7bdat (n=7,482,850) **Purpose:** Creates an index file that is unique at the beneficiary level, while retaining information that allows us to identify beneficiaries with multiple stays. This is necessary because the post-index file is not unique at the beneficiary level, so if we merge this before transforming the data we end up with a many-to-many merge. This allows us to do a 1-to-many merges, and then disaggregate the files to the bene-stay level afterwards.  **Notes:**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Program:** 5\_Post\_Index.sas  
**Inputs:** stay\_dataset\_all.sas7bdat (n=212,395,896); indexMatch.sas7bdat (n=7,482,850) **Outputs:** postindex\_DX\_0613.sas7bdat (n=25,231,855) **Purpose:** Formats and finalizes the post-index file. Only keeps post-index stays from beneficiaries that are in the index file, and only if they happen after the first occurrence of that beneficiary’s index stay and within 180 days of that beneficiary’s final index discharge. **Notes:** There is a line in this program (IF DIS(I)>0 AND (0<=admdiff<=/\*365\*/180) THEN DO;) where the upper-bound of the variable admdiff is important. Admdiff should be equal to the number of days in your post-discharge episode of care. If this number is too low, you will not capture all of the post-acute stays in your episode. If this number is too high, you will face a nuanced technical problem in future programs where certain beneficiaries are dropped for not having post-acute care within the X-day episode, but they do have post-acute care in the X+1 and beyond period (see the variable ‘interval’ in the DX\_readmission.sas programs) To make sure this does not happen, set this number equal to the length of your episode of care

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Program:** 6\_Final\_Index.sas  
**Inputs:** index03.sas7bdat (n=11,548,453) **Outputs:** readmissions\_index\_DX\_0613.sas7bdat (n=11,342,915) **Purpose:** Formats and finalizes the index file **Notes:**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Program:** DX\_readmission.sas  
**Inputs:** readmissions\_index\_DX\_0613.sas7bdat (n=11,342,915); postindex\_DX\_0613.sas7bdat (n=25,231,855) **Outputs:** collapsed\_DX.sas7bdat (2,186,722) **Purpose:** Apply inclusion and exclusion criteria to our sample (including more narrow condition of interest criteria), merges index and post-index files (creating a long-file), creates outcomes variables (readmission and mortality), applies the unplanned readmission algorithm, and flattens the file out to the bene-stay level while maintaining the post-index information that is important in creating the post-discharge episodes of care. This flattened file contains PAC outcomes variables such as charges and charges. **Notes:**

* Has Part A (or A+B) coverage during the month of index admission: 9,753,781
* No Aspiration: 8,861,922
* Meets Narrower PN Definition: 3,757,480 (big drop here, but see attached workbook for CMS’ definition of PN. The broader definition included in program 1\_Create\_initial\_index\_with\_transfers.sas had at least double the number of allowable primary DX codes. I anticipate this will raise flags so let me know if you want to discuss this further)
* Not an Intermediary Hospitalization in a Transfer: 8,664,638
* Age>=65: 7,296,738
* Survived Hospitalization: 7,828,175
* 12 Months consecutive Part A (or A+B) Prior Coverage: 7,180,158
* 6 Months Consecutive Part A (or A+B) Post Coverage or Death: 7,180,158
* Not Released Against Medical Advice: 8,807,320
* Final Count after Inclusion/Exclusion Criteria Applied: 2,251,705
* Not Within 30 days of prior eligible DX index stay: 2,188,156

This program has several nested macros that get called in, so the program Readmissions\_Macros\_2014.sas must be in the same directory to successfully run. This macro was provided directly from CMS, and the most important function is to apply the planned readmission algorithm so that we can distinguish between planned and unplanned readmissions. Additionally, this program relies on applying formats to diagnosis and procedure codes. The format catalogs are found in directories called ‘CCMap2008\_2012’, ‘CCS\_2013’, and ‘CC\_and\_CCS\_FORMATS\_TRN’, which are currently found in the same folder as the core programs.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Program:** Day\_array.sas  
**Inputs:** collapse\_DX.sas7bdat (n=2,186,722) **Outputs:** freq\_array\_DX.sas7bdat (n=2,186,722) **Purpose:** To sort and make sense of the post-acute information when generating the post-discharge episode of care. This is accomplished by creating a day array where variables day1-dayN are created that represent where in the healthcare system they are n days after discharge from their index stay. This program then uses day array to create PAC outcomes such as any PAC use, community release (defined as the absence of any PAC services), total days of PAC use (by facility type – STA, IRF, SNF, MDS, LTC, HHA). In addition, the program merges on Elixhauser and MBSF chronic condition flags, as well as generates index-stay chronic conditions flags. Finally, this program creates some control variables to capture things such as quarter and season. **Notes:** The day array program also has nested macros. These macros are called comoanaly2012\_2015.sas and comformat2012\_2015.sas. These macros are provided by HCUP and are used to generate the Elixhauser comorbidities. The macros from HCUP are currently found here: <https://www.hcup-us.ahrq.gov/toolssoftware/comorbidity/comorbidity.jsp> and must be in the same directory as the core programs to execute properly.

**Program:** prior\_nh\_hosp.sas **Inputs:** freq\_array\_dx.sas7bdat, Raw MedPAR (2006-2013; n=133,550,584), Raw MDS2.0 and MDS3.0 (2006-2013; n=130,041,719) **Outputs:** freq\_array\_dx.sas7bdat **Purpose:** This program adds flags that identify if a beneficiary has had prior nursing home stays or hospitalizations according to the raw MDS or MedPAR data. For the prior nursing home stays, this program looks at the most recent assessment date in the raw MDS data instead of relying on the discharge date on the processed MDS stay-level file because the discharge dates on that file are oftentimes questionable. Flags are created for having had an assessment within the last 90 or 180 days (prior\_nh\_90, prior\_nh\_180). The prior\_hosp365 flag is a continuous variable that counts the number of STACH or CAH hospitalizations a beneficiary was discharged from within 365 days of the index hospital admission.  **Notes:**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Program:** backfill\_mds.sas **Inputs:** freq\_array\_dx.sas7bdat, Raw MDS2.0 and MDS3.0 (2006-2013; n=130,041,719) **Outputs:** freq\_array\_dx.sas7bdat **Purpose:** . It attempts to replicate the Residential History File (RHF) methodology by setting all days that are designated as “MDS” to missing, and then filling them in based on raw assessment dates and types. See here: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3015013/>  
**Notes:** Optional to run this program