Readme for Medicare project

Angela Wang [amwang@stanford.edu](mailto:amwang@stanford.edu)

Last updated: 7/27/2011

-Research question at hand: what effect does hospital concentration have on the bargaining power of MA plans?

-Starting in 2008, hospitals were required to report the utilization data of their MA patients to the Medicare. Thus our analysis begins with 2008 data.

-The current project deals only with 2008 data, but the code can be modified so that subsequent years can be added. The macro %year. is used in all code.

-Change working directories to correspond to your own.

##Piece-by-Piece: building our dataset

All dataset processing for this project is completed in SAS. All regression analysis is completed in STATA. The philosophy behind this is that SAS can handle complex data sets while STATA is much more efficient for analysis.

Stat-Transfer (available on all of our servers) can be used to convert files from .sas7bdat to .dta and vise-versa. Elsewise, xpt files can be used to export/import from SAS to STATA and vise-versa.

Statamp servers: age2, 4, 5, 6, 8

SAS servers: age3, 5

Please refer to the “Building hospital regression datasets” at end of this file for dataflow, dataset dependencies and code dependencies.

Levels: Analysis is conducted on both the discharge as well as the beneficiary-MA status level.

Outline

Beneficiary/Utilization: Denominator and MedPAR files

-Overview

-Patient characteristics

-demographics

-geocodes

-ICD9 diagnosis codes/HCC

-Stay characteristics

-MA indicator/weights

-construction

-assignment

-charges, cost, revenue, revenue/cost

Hospital-level data:

-hospital characteristics

-original set

-hospital system

-cost reports

-assignment to stay-level data

-National: CMS

-construction of npr, ccr

-exclusion of hospitals

- CA-only: OSHPD

-construction of two different npr variables

-ongoing issues

IVs: benchmarks and ffs spending

-exclusions

Hospital market structure variables:

-generate hospital choice dataset

-split hospitals and beneficiaries into mutually exclusive and exhaustive regions

-calculate differential distances

-conditional logits

-calculation of hosp market structure variables

Hospital competition structure variables:

-merge all necessary data pieces together

-zero-stage regressions

-probits

-generate IV variables

-two-part models: and CA-only samples

-National

-beneficiary-level

-discharge-level

-Califoria

-beneficiary-level

-discharge-level

Quick Tips

-How to batch run a job

-SAS

-SQL

-Terminal/Bash

Sample output for construct\_regions.sas

Code files and description

Data files and description

-analysis\_stata/: dta files

-workingdata/: sas7bdat files

Building hospital competition datasets

###Medicare beneficiary data: Denominator and MedPAR (unique identifier: hicbic)

One should begin by understanding the contents of the Medicare utilization and enrollment data from CMS (Center for Medicare & Medicaid Services). The [RESDAC (Research Data Assistance Center) website](http://www.resdac.org/ddvh/Index.asp) provides detailed codebook (aka data documentation, data dictionaries, record layouts, and data layouts) info regarding all the datasets that are available to researchers. The pertinent files for this project are the "Medicare Denominator File" (commonly referred to as "denom"), which is a enrollment/summary file, and the "Medicare MedPAR file" (referred to as "MedPAR"), which is the utilization file. These files are \*\*big\*\*. New programmers should take a look at the RESDAC documentation to get acquainted with the data and determine what variables are available and where they are located.

-The denominator file contains all enrollees (aka beneficiary, eligible, or member) in Medicare for the calendar year and their demographic information including: date of birth, date of death, zip code of residence, sex, ethnicity, monthly indicators for hmo enrollment, monthly indicators eligibility for medicare, etc. For 2008, this file contains 46+ million observations.

-The MedPAR File contains inpatient hospital and skilled nursing facility (SNF) final action stay records. Each MedPAR record represents a stay in an inpatient hospital or SNF. An inpatient "stay" record summarizes all services rendered to a beneficiary from the time of admission to a facility through discharge. Each MedPAR record may represent one claim or multiple claims, depending on the length of a beneficiary's stay and the amount of inpatient services used throughout the stay. For 2008, this file contains 13+ million observations.

Chris mentions the "Inpatient SAFs" (ip) in his original email. We choose to use the MedPAR files instead of the ip. The unit of analysis for the Inpatient SAFs is a claim whereas for the MedPAR it is a stay (an inpatient stay may have several claims). With (some experienced) manipulation you can transform the ip file to MedPAR form.

Our data is stored on the NBER servers. Jean Roth (jroth@nber.org), one of the two data/server managers at the NBER, processes the yearly data that comes from the CMS every year. For 2008, the denom and MedPAR files are split into 100 files that need to be appended back together to form your personal working dataset. There are several sample sizes that you can use ranging from 1% (maybe smaller available as well) to the full 100% file. Working with a smaller file will allow you to cut down on processing time during a debugging phase. Ultimately though, all analysis will be run on the 100% files.

The directory locations for these two sets of files are here:

/disk/agedisk1/medicare/data.NOBACKUP/u/c/100pct/denom - denominator

/disk/agedisk1/medicare/data.NOBACKUP/u/c/100pct/medpar - medpar

There's a subdirectory for each year of data under each of these subdirectories. The NBER has data through 2005 currently, and we can use data as far back as 1996. In the locations above you'll find data for 2002-2008.

All of these files have had their HICs encrypted, using a common encryption algorithm across files and two sets of years 2002-05 and 2006-08. In the 2002-05 files, you will see a variable called ehic in each file. In the 06-08 files, you will see a variable called bene\\_id. \*We rename the bene\\_id variable to hicbic for consistency with other sets of data.\*

See "construct\\_denom.sas" and "construct\\_medpar.sas" for code to build these files. Liberal comments are used in the denom file. These personal CMS working files are used heavily in the construction of other intermediary datasets further on in the project. After you have these constructed to your liking, make sure you have a backup stored somewhere as this is a processing intensive step.

Below I detail the variables that we use from these two datasets.

####Patient characteristics

#####Demographics (variables: a6569, a7074, a7579, a8089, a9099, female, black)

Age, female, black, and derivative interacted indicators can be constructed directly from the denominator file.

For our analysis, age is always taken to be their age at January 1, 2008. There may be some consideration to recode ages for the discharge-level regressions, so that age reflects the beneficiaries age at time of discharge.

Black is the only reliable indicator that researchers use from the race variable, so we follow suit.

Code to implement this appears at various points throughout the project when needed (eg. "analysis\\_denom.sas" lines 41-62)

#####Patient location/geocode (variables: pzip, SSA)

5-digit zip-codes are also constructed directly from 9-digit bene\_zip in the denominator file. SAS has a handy zipcode-to-geocode crosswalk(sashelp.zipcode.sas7bdat) which will geocode any valid zip-code to the centroid of the zip-code and they also have some [handy documentation](http://support.sas.com/resources/papers/proceedings10/219-2010.pdf) on this process. We considered using 9-digit or 3-digit zips, but ultimately we decided that analysis on those levels would be respectively too granular or too coarse.

We also create a complete 5-digit SSA (Social Security Administration) state-county code by concatenating the SSA state-code and SSA county-codes from the denominator so that we can later merge in our IV: the Medicare Advantage benchmark payment rate.

Note that the SSA state-county code is different from the more commonly used FIPS (federal info processing standards) state-county codes. There are plenty of SSA-FIPS crosswalks available if any merging needs to be done. This has been unnecessary for our analysis so far.

#####Diagnosis codes (DGNSCD1-DGNSCD10)

We eventually recode ICD-9 (int'l classification of disease codes/diagnosis codes) from MedPAR to HCC (hazard characteristic code) using the CMS-HCC model that was in place in 2008, the [2007 HCC model software](https://www.cms.gov/MedicareAdvtgSpecRateStats/06a\_Risk\_adjustment\_prior.asp). Note that the CMS-HCC scheme we use is not exhaustive, so not all of our diagnosis codes will have a matching HCC. CMS uses a modified version of the model that has 70 indicators in 2007, while the true HCC model has 180 or so. The percentage of ICD-9 codes <5% that do not have a match are not worrisome.

The HCC risk model is used to adjust Medicare capitation payments to private health care plans for the health expenditure risk of their enrollees. We are using the HCC in a different capacity: as binary controls for disease conditions. Essentially, we are binning the 6000+ different ICD-9 codes into 70 disease categories for control purposes.

"analysis\_medpar\_HCC\_byhicbic.sas" is used to recode the 10 diagnosis codes by stay in MedPAR to 70 HCC indicators by hicbic.

The HCC indicators for each hicbic represents a summary of all HCCs that the beneficiary was diagnosed with over the calendar year. We map the 10 diagnosis codes to their respective HCCs for each stay and then reshape to obtain 70 indicator dummies for each stay. We then take the maximum value of each HCC indicator by hicbic to obtain the hicbic level file.

####Stay (discharge) characteristics

#####MA (Medicare Advantage) indicator and weight (variables: MA, weight)

HMO/GHO/MCO/MA what are they and how do they differ?

An HMO (health maintenance organization) or a GHO (group health organization) is a type of MCO (managed care organization) that provide some form of health care coverage. The terms "HMO" and "GHO" are used interchangeably to indicate a managed care plan. Medicare Advantage (MA) is the managed care plan for Medicare.

The MedPAR file contains a variable called the MedPAR GHO Paid Code, but we choose not to use this. [ResDAC provides a nice write up on GHO/HMO encoding in MedPAR files](http://www.resdac.org/tools/TBs/TN-009\_MedicareManagedCareEnrolleesandUtilFiles\_508.pdf). This code indicates whether or not a GHO has paid the provider for the claim (in ip)/stay (in medpar). However, an empirical analysis conducted by ResDAC showed that the indictor was correct only 95% of the time, so they recommend that researchers use the monthly HMO indicators from the denominator data. To assign an HMO status to each stay, we first construct monthly MA indicators for each hicbic. Then we match each MedPAR stay by discharge month to its respective HMO status. This process actually turns out to be a lot more complicated than it sounds. The easiest way to understand this is to look through the sas files.

######Constructing our MA indicator and weights "hmo\\_status\\_byhicbic.sas"

We use two sets hmo1-hmo12 and buy1-buy12 monthly indicator variables and the death date (if applicable) to construct 12 monthly MA indicators for each hicbic. A beneficiary's enrollment may not be consistent across the year, so we use the rules listed below to determine a beneficiaries eligibility to remain in our dataset and construct some weights for those enrollees who have both MA and TM enrollment in 2008. The buy-in indicator is a 0/1 dummy for whether or not the bene was eligible for

Establish monthly eligibility for Medicare:

1) Eligibility ends the first full month after month of death. Benes received benefits up to the date of death that implies that they are eligible for Medicare during their month of death.

2) Buyin indicator must be 1 and you must be living/died during that month.

Keep valid beneficiaries with valid eligibility data:

1) Benes must have either zero or only 1 switch between MA and TM plans

2) Benes must have continuous eligibility for the entire year (or until death if applicable)

3) Benes must not have "aged" into Medicare: eligibility switch from noneligible-->eligible

TM and MA weights:

Since some proportion of benes switch plans during the year, we construct weights so that we can utilize both the TM and MA part of their data.

1) weightMA = months MA/12 and weightTM = months TM/12; weightTM and weightMA should sum to 1 since we only keep benes who have continuous eligibility for the entire year.

2) If bene dies in 2008, weightMA = months MA/months eligible and weightTM = months TM/ months eligible; we do this so as to prevent the down-weighting of benes who die.

######Assigning our MA indicator

The MA indicator is assigned based on the month of discharge of a stay. "analysis\\_rcc\\_byhicbic.sas" and also “analysis\\_rcc\\_bydischarge.sas”

For each stay, we match the admission month to the hicbic’s respective hmo status for that month.

#####Charges, Costs, and Revenue (rcc)

Two-levels of financial data exist. Here we discuss the values that are unique for each stay. We can aggregate the first three of these variables to a beneficiary-ma status level by summing. Price is recalculated at the beneficiary level after all the primary variables have been summed.

On an aggregate level, hospitals also have these values which are reported in CMS cost reports. We use the hospital-level ratios to derive/interpolate financial data for both our TM and MA stays. Those financials are further discussed in the following hospital-level data section

######Charges (totchrg)

Definition: dollar amounts “charged” for a service by a health care provider. This is often different from the actual payments made to providers

Other names: gross revenue

Origin: This value is taken directly from each MedPAR stay record. No processing needed! Hurray!

######Cost (cost)

Definition: Cost to the hospital for services provided

Other names: expenditures

Origin: Derived using cost reports for all stays using hospital-level cost-to-charge ratio

totchrg\*ccr = cost

######Revenue (revenue)

Definition: amount that the provider (hospital) actually makes. Theoretically the identity "revenue = cost-charges" should hold true.

Other names: payment, medpar\_payment, net revenue

Origin: For TM folks, this value is the sum of several payment variables from MedPAR. (medpar\_payment= BLDDEDAM+COIN\_AMT+PMT\_AMT+PRPAYAMT+DED\_AMT)

For MA folks, this value is derived using a net payment ratio (revenue/costs) revenue totchrg\*npr

######Revenue/cost (price)

Definition: revenue/cost. This 4th DV is not dependent on volume (charges)

Other names: new DV

Origin: revenue/cost as calculated above for all beneficiaries. One caveat is that MA patients at the same hospital all have the same “price”. This is because both the revenue and cost components for MA patients are derived payments.

Other notes: we may drop observations that have outlier price amounts. This is still to be determined as of 7/26/2011

###Medicare hospitals (unique identifier: mprovno)

####Hospital Location/geocode (variable: hzip)

The raw file (/disk/homes2b/nber/cafendul/hosp\\_prices/Hospital\\_Geocoding\\_Result2.dbf) for this part of the project was obtained from Chris. We sent a file with hospitals and hospital addresses to Scott, our map library contact, who helped geocode them into latitude and longitude coordinates. This file is imported into SAS and saved as "hosp\\_geocodes.sas7bdat". The cleaned version, "hosp\\_geocode\\_clean.sas7bdat" has a total of 3560 unique hospitals.

"clean\\_hosp\\_geocode.sas" is used to clean up the raw data and process it to an analysis ready file.

Hospital data was checked against their listing on the [Medicare Data website](http://data.medicare.gov/dataset/Hospital-General-Information/v287-28n3).

Some hospitals have miscoded zip codes. These zipcodes are corrected using the data step.

Exact duplicates are dropped, but many hospitals still contained multiple entries with conflicting geocodes.

Hospital duplicates come in several flavors:

- Miscoded zip codes: one entry has the wrong address/zipcode, the duplicate entry which was coded using the wrong zip was dropped

- Conflicting address:

1) Hospitals that have multiple campuses but only one hospital code. We take the main campus to be the location for our geocode.

2) Hospitals with two addresses, we take the average of these geocodes to be our final geocode. Usually these geocodes are similar and have a difference of less than 0.01deg.

To further validate our hospital data we compare it to the geocodes provided by the AHA (American Hospital Association) hospital files. We don't use these files straight up because they only cover a portion of the hospitals we need geocodes for. That said, only 147 hospitals appear in the cost reports but not the AHA. If lat and long differ by more than 0.5deg, the lat and long were updated to what was reported in the AHA.

Note about AHA files: You'll need to sign a consent form through NBER/Jean to work with these files.

####Hospital characteristics (variables: hchar1-hchar7, beds, hchar8)

Chris provides the hospital characteristic file. (/disk/homes2b/nber/cafendul/hosp\_prices/hosp\_chars.sas7bdat.gz)

The hospital characteristics include size (small, med, large) based on the number of beds, ownership (non-profit, for-profit, and teaching) and whether or not it is a teaching hospital. This dataset can be used as is. These variables are renamed to have a common prefix to aid processing. This renaming occurs in (construct\_regions.sas line 176).

hchar group 1, 2, and 3 and group 4, 5, and 6 form linear combinations. We drop 3 and 6 from our analysis to ease the Stata processing load. If left in, Stata would drop one variable out of each these groups anyways.

We later added the hchar8 as a dummy variable to account for membership of a hospital in a hospital system. Jean extracted the crosswalk from mprovno to hospital-system-id from the AHA data. (aha\_extract2008.dta) I process it to become aha\_sysid.dta. Since we introduced this variable into our analysis later on, we did not merge it in and construct it until it was time to construct the hospital market structure variables in “analysis.do”. Not all hospitals that appear in MedPAR are exhaustively included in AHA. There are around 100 hospital that we don’t account for and we code these as not belonging to a hospital system.

Mapping of hospital characteristics:

own\_fp=hchar1 ownership for profit

own\_np=hchar2 ownership nonprofit

own\_gv=hchar3 ownership government

small\_beds=hchar4 small number of beds

med\_beds=hchar5 medium number of beds

large\_beds=hchar6 large number of beds

teaching=hchar7 teaching hospital

hosp\_sys=hchar8 hospital belongs to a hospital system

####Financials/Cost reports

We originally conducted a national level analysis using only the data from CMS. However, CA also collects its own hospital financials data (reported by OSHPD Office of Statewide Health Planning and Development), so we have begun to create a dataset to complete a CA-only analysis.

The calculated npr and ccr from the cost reports are matched to stay-level data by mprovno and discharge date. Since fiscal years do not necessarily correspond with calendar year begin and end dates, we combine cost report data from FY2008 and FY2009 so that the entire calendar year is accounted for. A cost report is matched if the fy end date is greater than or equal to the discharge date and the fy begin date is less than or equal to the admission date.

#####National: CMS

Chris provides us with CMS cost reports (/disk/homes2b/nber/cafendul/hosp\_prices/hosp\_chars\_new.sas7bdat.gz).

We use this data to calculate a net payment ratio (npr) and cost-to-charge ration(ccr) for each cost report. "npr\_ccr\_bymprovno.sas" Using these ratios, we can then impute the cost for each stay and the revenue for MA stays. The cleaned dataset is “hosp\_costs.sas7bdat”.

We can calculate net revenue to gross revenue ratio (npr) using data directly from the hospital cost reports. Note that these represent net and gross revenue (total charges) for all patients, for all different parts of the hospital complex. To the extent that we can isolate those net revenues and gross revenues that do not include Medicare payments of various kinds, we can improve upon it. One thing we can definitely do is to remove Medicare inpatient net revenue from the numerator, and Medicare inpatient total charges from the denominator. Unfortunately, we can’t go further than this. While we can identify other pots of net revenue that are paid by Medicare (i.e., SNF, HHA), we can't find the corresponding charge figures for each of those sources.

net payment ratio = [net patient revenue - Medicare net patient revenue]/[total charges - Medicare charges]

Cost to charge ratios (ccr) are calculated directly using available variables in the cost reports. A detailed codebook for CMS cost report variables is not available at this time.

ccr = [total cost1 + total cost2]/ total gross revenue

######Exclusion of hospitals

We exclude hospitals that have weird npr or ccr ratios. Npr and ccr ratios must fit the following rational criteria.

1. 0<npr<1, revenue and charges must be positive and charges must exceed revenues
2. ccr<1, revenue(charges) must exceed costs

#####CA OSHPD Cost reports

Since the OSHPD data is more granular, we can construct two different net payment ratios.

1) npr that is most analogous to the national level npr:

rc\_ratio\_CA\_rev = (Net patient revenue by all other payers, including MA, except for TM)/(gross inpatient revenue total - gross inpatient revenue by TM + gross outpatient revenue total).

2) npr that uses the hospital reported managed care values for both net patient revenue and gross inpatient and outpatient revenue

rc\_ratio\_CA = (Net Patient Revenue Medicare: Managed Care only)/(Gross Inpatient Revenue Medicare: Managed Care only + Gross Outpatient Revenue Medicare: Managed Care only)

######Current ongoing issues:

-Not all MedPAR CA hospitals appear the OSHPD dataset. We are still in the process of determining how many hospitals we cannot account for.

-Hospitals that report no MA spending in OSHPD have MA discharges according to the MedPAR data.

####IV construction: MA benchmark payment rate and ffs (fee for service) spending

The MA benchmark payment rate and ffs spending are both available by county. We create two IVs for each county. (ffs\_spend.do which outputs benchmark\_new.dta)

We first calculate ffs\_spend using data from the Medicare Advantage Special Rate Stats

[http://www.cms.gov/MedicareAdvtgSpecRateStats/05\_FFS\_Data.asp](http://www.cms.gov/MedicareAdvtgSpecRateStats/05_FFS_Data.asp" \t "_blank).

ffs\_spend=partapercapitawoimedshgme+partadsh/(partaenrollment\*12)+partbpercapita

We can then merge on the SSA code and then calculate the IVs:

IV1: b\_divide\_ffs = benchmark/ffs\_spend

IV2: b\_minus\_ffs = benchmark-ffs\_spend

#####Exclusion of counties

There is one ffs outlier county that we choose to remove: Loving, TX. SSA: 45762 and zip: 79754. It has an ffs-spending value that is way above the norm. Although Miami-Dade, FL also has a high ffs-spending, but it is large enough and important enough in the debate over MA that is needs to stay.

We also choose to drop all beneficiaries in Alaska since it doesn’t really have any kind of Medicare managed care market to speak of. Including it could cause problems in estimation.

###Hospital market structure variables

Theory and methodology for constructing these variables mirrors that as describe in Dan’s “Is Hospital Competition Socially Wasteful” paper, section III. 1. and 2. However, we deviate from the paper when constructing our analysis regions. See sample output for construct\_regions.sas at the end of this document for a better idea of what’s going on.

####Generating hospital choice dataset

The big picture idea is that we’re creating an artificial dataset of all possible hospital choices that a patient in any particular zip code with unique set of patient characteristic will have.

1. Determine the unique zip codes and hospitals that exist in each region and geocode each.
2. Cross each valid zip code within a region with all hospitals in that region within a 100-mile great-circle radius of the zip code.
3. If the hospital choice actually exists in MedPAR, then the hospital-choice variable is 1. Elsewise it is 0.

Other distances measures such as travel time may eventually be used in the restriction of 2.

We use HRRs that exhaustively cover all zip codes to divide the US up into regions. ([http://www.dartmouthatlas.org/tools/downloads.aspx](http://www.dartmouthatlas.org/tools/downloads.aspx" \t "_blank))

Our regions are as follows:

\*1: CT, ME, NH, RI, and VT

\*2: MA

\*3: Bronx, East Long Island, Manhattan HRRs

\*4: all NY, NJ or PA HRRs (except NYC);

\*5: DE, DC, FL, GA, MD, NC, SC, VA, and WV

\*6: IN and OH

\*7: MI and WI

\*8: IL

\*9: AL, KY, MS, and TN;

\*10: IA, KS, MN, MO, NE, ND, and SD

\*11: AR, LA, OK, and TX

\*12: AZ, CO, ID, MT, NV, NM, UT, and WY;

\*13: AK, HI, OR, WA, and all of CA except the LA HRR

\*14: LA HRR;

#####Calculate differential distances

We then calculate two sets of differential distance measures: using same-type relative distances and different-type relative distances. I won’t delve into more detail here as the theory behind this is contained in Dan’s paper. The code will also be illuminating if you want to see exactly how this is all done.

####Conditional logits

clogit analysis (aptly named analysis.do) is run on each regional file and predicted values for each hospital choice are generated. We use these predicted values to generate the market structure variables.

Right hand side controls include all demographic and hospital level interactions that are generated in a series of loops in analysis.do.

####Calculation of market structure variables

“hosp\_mrkt\_strct.do” constructs all of the market structure variables (hosp\_mkrt\_zip.dta) that are needed for our analysis. The method follows that of the paper exactly.

A total of 10 variables are constructed at the 5-digit zipcode level: HHI\_pat\_k\_star, CAP\_pat\_k\_star, hchar1\_pat\_k\_star- hchar8\_pat\_k\_star

###Hospital competition regressions

“IV0.do” merges all necessary files together to run zero-stage regressions and form our final analysis datasets. “IV\_dis.do”, “IV\_hic.do”, (also combined as “IV.do”) and “IV\_CA.do” run all of our regression models.

We follow the methodology as outlined by Wooldridge here ([http://www.stata.com/statalist/archive/2011-03/msg00188.html](http://www.stata.com/statalist/archive/2011-03/msg00188.html" \t "_blank))

####zero-stage regressions

We run this on the national beneficiary-level data.  It is not necessary to run this on the discharge level data, because the purpose of this model is to assign each beneficiary a predicted probability of MA enrollment; since this is a beneficiary-level phenomenon, it doesn't make sense to estimate it on the subset of benes with an inpatient hospital discharge.

1. Reduced form probit

We estimate a reduced form probit for the dependent variable using all exogenous variables. Get the fitted probabilities, ma\_hat.

For our model, we run a probit regression for the patient's traditional medicare (TM) or Medicare Advantage (MA) status, using all of the RHS variables that we will include in the second stage model (patient demographics, market-level measures constructed from the hospital choice models), plus our instrument (the Medicare Advantage payment rate in the patient's county of residence minus the ffs-spending or the benchmark divided by the ffs\_spending). We would expect our instrument to be positively associated with enrollment in MA (since the payment rate increases benefits to prospective enrollees, making them more likely to enroll). Once this is done, calculate ma\_hat for all individuals.

2. Construct instruments (not regressors) phat\*X1, phat\*X1\*X2

Create interaction terms between the MA phat variable and all of the variables we want interacted with the MA indicator. HHI-pat-star, CAP-pat-star and hosp\_char1-pat-star – hosp\_char8-pat-star

####Two-part models

We want to estimate the equation:

Y = b0+ b1\*Y1 +b2\* Y1\*X1 + b3\*Y1\*X1\*X2

Using IVs phat, phat\*X1, phat\*X1\*X2.

We don't have any three-way interactions, so our set up is simpler. We estimate a model that looks like this:

Y = b0 + b1\*Y1 + b2\*X2 + b3\*Y1\*X1 + b4\*Y1\*X1

So to run this, we will do an IV model where the first stage uses phat, phat\*X1 and phat\*X2 to instrument for Y1, Y1\*X1 and Y1\*X2, respectively. Then the first stage will include these latter three variables in instrumented form.

We complete the actual analysis using a two-part model with four different samples: national beneficiary level, national discharge level, and CA-only beneficiary and discharge level. The above estimation technique only comes into play in our IV regressions.

1. National beneficiary level regressions.  We run these regressions on five different dependent variables: the "first" part of the two-part model (a 0/1 variable indicating whether the beneficiary had any inpatient hospital spending or not), charges, cost, net revenues, and price.  We run five different specifications:  
     
   - OLS with interactions  
   - OLS without interactions  
   - OLS with no MA dummy  
   - IV with interactions  
   - IV without interactions

We run the OLS regressions once and the IV regressions twice, once for each of the instruments we use. This means running 35 regressions: 3 OLS regressions with 5 DVs = 15 OLS regressions, and 2 IV regressions with 2 instruments and 5 DVs = 20 IV regressions.

1. National discharge level regressions are similar. We run these regressions on five different dependent variables: first part of the two-part model (the positive charge indicator), charges, cost, net revenue, and price.  We run five different versions again:  
     
   - OLS with interactions  
   - OLS without interactions  
   - OLS with no MA dummy  
   - IV with interactions  
   - IV without interactions  
     
   Run the IV regressions twice, once for each of the instruments we are contemplating using.  This means running 35 regressions: 3 OLS regressions with 5 DVs = 15 OLS regressions, and 2 IV regressions with 2 instruments and 5 DVs = 20 IV regressions.
2. CA only beneficiary level regressions.  We run the "IV with interactions" model for the five DVs described (2) above.  We run it once for each instrument and run these three different ways:
   1. Restrict attention to beneficiaries with CA as their state of residence.  We also exclude *any beneficiaries who had one or more discharges at a non-CA hospital* (or a CA hospital that does not appear in our OSHPD hospital data).  Continue to use the Medicare cost-report-based measure of net revenue for this version.  Five regressions.
   2. Restrict the sample in the same way as the previous step.  Re-run the net patient revenue regression only, but create a new version of this variable using net-revenue-to-charge information from the OSHPD hospital data.  One regression.
   3. Restrict the sample in the same way as the previous step.  Re-run the net patient revenue regression only, but create a new version of this variable using net-revenue-to-charge information from the OSHPD hospital data that is specific to MA patients only.  One regression.

This is a total of 5+1+1=7 regressions.  Run these once for each instrument, for a total of 14 regressions.

1. CA only discharge level regressions.  Run the "IV with interactions" model for the five DVs described (2) above.  Run it once for each instrument.  Run these three different ways:
   1. Restrict attention to discharges for beneficiaries with CA as their state of residence.  Also exclude *any discharges from non-CA hospitals* (or CA hospitals that do not appear in our OSHPD hospital data).  Continue to use the Medicare cost-report-based measure of net revenue for this version.  Five regressions.
   2. Restrict the sample in the same way as the previous step.  Re-run the net patient revenue and revenue-to-cost regressions only, but create a new version of this variable using net-revenue-to-charge information from the OSHPD hospital data.  Two regressions.
   3. Restrict the sample in the same way as the previous step.  Re-run the net patient revenue and revenue-to-cost regressions only, but create a new version of this variable using net-revenue-to-charge information from the OSHPD hospital data that is specific to MA patients only. Two regressions.

This is a total of 5+2+2=9 regressions. Run these once for each instrument, for a total of 18 regressions.

##Working in the Unix environment

How to batch run and monitor a program:

>cd myworkingdir/

to start a job in the background:

>sasage3 sasfilename.sas & or >sasage5 sasfilename.sas &

>statamp –b dofilename.do & or >nohup statamp –b dofilename.do &

to monitor output:

>tail –f sasfilename.lst or >tail –f sasfilename.log

>tail –f dofilename.log

to figure out how the job is doing:

>ps (if job is on current login session)

>top (if job was on a different login session)

kill a program at any point:

>kill -9 PID

Stata also comes in the interactive mode on the servers with statamp installed: 2, 4, 5, 6, or 8. A job will end upon logoff from the server.

To invoke:

>statamp

###If your SAS program doesn't run:

Learning to love SAS will take time. Breath.

-Semicolon delimiter. Is it at the end of every proc/data step?

-Make sure that you have a \*\*libname\*\* specified. Unlike STATA, SAS wants you to tell it exactly where it can find the datasets it will work on.

-Single vs. double quotes matter. Check to see if you're using the right ones.

-Commas make a big difference, check to see if you need or don't need them

-Shell commands: cd, rm, bunzip (things that your bash terminal will understand) can be evoked with an “x” followed by your command in double quotes

-Are the macro variables referenced correctly: “&macro\_name.”? Make sure macros are correctly defined before you reference them.

###The power of SQL

Learning to run queries (aka chop up your dataset(s), combine, reshuffle your data) in SQL will save you tons of time and headaches in the long run. Any DATA step can pretty much be replaced with a more efficient and cleaner SQL step. The [handbook](http://support.sas.com/documentation/onlinedoc/91pdf/sasdoc\_91/base\_sqlproc\_6992.pdf) provides lots of examples is a great go-to for any sql coding questions. The code for this project also has plenty examples. This handbook is actually pretty useful and I’ve gone through it several times to pick up good ideas to integrate into coding practices.

We also have a MySQL book floating around in the room. The syntax between proc sql and MySQL is a bit different, but the book is still very useful.

###Most used Terminal/Bash commands

If all else fails, just close the window and start up a new session. :-)

> kill -9 PID: bring the process to the highest priority (-9) and then kill it. PID is the process number which you can look up by using 'top' or 'ps' (only if you are still in the same bash session). You can kill any process that belongs to you

> ls -lha: show all(-a) files in long(-l), human(-h)-readable format

> exit: exit the program or session

> rm -rf: remove a file without prompt(-f) and all child directories (-r)

> top: list the top process of all users that are currently running on a machine. You can also kill a process while in top by hitting 'k' and then typing the PID.

> ps: list all of the running process for the current session

> tail -f: monitor the tail end of an output .log. I use this \_all\_ the time to check how regressions are running

> cat: open and read a document in the screen

> pico or vim: to edit documents on the fly

> cd: change directories

“cd ..”: go up a level

“cd”: go to home dir

“cd –”: go to last dir

> bzip, bunzip and gzip, gunzip: zip and unzip files using these two utilities

- Most used shortcut keys:

ctrl+c: escape from current line/start a new line

ctrl+c, ctrl+x: exit out of current program

ctrl+u: delete everything ahead of the cursor on this line

up/down arrows: recall and cycle through previous commands

ctrl+r: last stata command

SAMPLE LOG FROM REGION 20: LA PMSA

(note: region 20 was eventually remapped region 14: LA HRR)

Patient zips, latitudes, and longitudes are repressed

hchar1 ownership for-profit

hchar2 ownership non-profit

hchar3 ownership government

hchar4 small

hchar5 medium

hchar6 large

hchar7 teaching

mprovno hospital id

hlat hospital latitude

hlon hospital longitude

zip9 9 digit patient zipcode

dist distance between hospital and zip9

1. raw analysis dataset with hospital characteristics for all possible mprovno-zip9 pairs (ref: analysis from .sas file)

Obs mprovno zip9 hlat hlon dist hchar1 hchar2 hchar3 hchar4 hchar5 hchar6 hchar7

1 050018 zip1 34.0646 -118.239 5.4022 1 0 0 0 1 0 0

2 050040 zip1 34.3242 -118.448 25.829 0 0 1 0 1 0 1

3 050056 zip1 34.6894 -118.160 48.775 0 1 0 0 1 0 0

4 050058 zip1 34.1284 -118.258 9.7613 0 1 0 0 1 0 0

5 050063 zip1 34.0956 -118.292 7.8262 1 0 0 0 1 0 1

6 050078 zip1 33.7379 -118.306 17.478 0 1 0 0 1 0 0

7 050091 zip1 33.9892 -118.225 1.5756 1 0 0 0 1 0 0

8 050096 zip1 34.0704 -117.945 18.490 1 0 0 1 0 0 1

9 050103 zip1 34.0511 -118.217 4.8468 0 1 0 0 1 0 1

10 050104 zip1 33.9311 -118.204 4.7611 0 1 0 0 1 0 0

2a. product data from proc means to identify closest two hospitals for each characteristic (ref: product)

mprovno\_ mprovno\_

Obs zip9 hchar1 hchar2 hchar3 hchar4 hchar5 hchar6 hchar7 \_TYPE\_ \_FREQ\_ dist\_1 dist\_2 1 2

1 zip1 . . . . . . 0 129 50 1.5756 4.3449 050091 050663

2 zip1 . . . . . . 1 129 38 3.5416 4.6257 050149 050471

3 zip1 . . . . . 0 . 130 85 1.5756 3.5416 050091 050149

4 zip1 . . . . . 1 . 130 3 5.5329 9.6000 050373 050625

5 zip1 . . . . 0 . . 132 13 5.5329 5.8266 050373 050660

6 zip1 . . . . 1 . . 132 75 1.5756 3.5416 050091 050149

7 zip1 . . . 0 . . . 136 78 1.5756 3.5416 050091 050149

8 zip1 . . . 1 . . . 136 10 5.8266 7.9353 050660 050751

9 zip1 . . 0 . . . . 144 81 1.5756 3.5416 050091 050149

10 zip1 . . 1 . . . . 144 7 5.5329 6.9041 050373 050717

11 zip1 . 0 . . . . . 160 48 1.5756 4.3449 050091 050663

12 zip1 . 1 . . . . . 160 40 3.5416 4.6257 050149 050471

13 zip1 0 . . . . . . 192 47 3.5416 4.6257 050149 050471

14 zip1 1 . . . . . . 192 41 1.5756 4.3449 050091 050663

15 zip2 . . . . . . 0 129 50 1.5733 4.2850 050091 050104

16 zip2 . . . . . . 1 129 38 4.0546 5.1370 050149 050471

17 zip2 . . . . . 0 . 130 85 1.5733 4.0546 050091 050149

18 zip2 . . . . . 1 . 130 3 5.9311 10.026 050373 050625

19 zip2 . . . . 0 . . 132 13 5.9311 6.2133 050373 050660

20 zip2 . . . . 1 . . 132 75 1.5733 4.0546 050091 050149

21 zip2 . . . 0 . . . 136 78 1.5733 4.0546 050091 050149

22 zip2 . . . 1 . . . 136 10 6.2133 8.3489 050660 050751

23 zip2 . . 0 . . . . 144 81 1.5733 4.0546 050091 050149

24 zip2 . . 1 . . . . 144 7 5.9311 6.5053 050373 050717

25 zip2 . 0 . . . . . 160 48 1.5733 4.5090 050091 050663

26 zip2 . 1 . . . . . 160 40 4.0546 4.2850 050149 050104

27 zip2 0 . . . . . . 192 47 4.0546 4.2850 050149 050104

28 zip2 1 . . . . . . 192 41 1.5733 4.5090 050091 050663

2b. product\_alt data to identify closest two hospitals for each opposite characteristic (ref: product\_alt)

mprovno\_ mprovno\_

Obs zip9 ophchar1 ophchar2 ophchar3 ophchar4 ophchar5 ophchar6 ophchar7 \_TYPE\_ \_FREQ\_ dist\_1 dist\_2 1 2

1 zip1 . . . . . . 0 129 50 1.5756 4.3449 050091 050663

2 zip1 . . . . . . 1 129 38 3.5416 4.6257 050149 050471

3 zip1 . . . . . 0 . 130 85 1.5756 3.5416 050091 050149

4 zip1 . . . . . 1 . 130 3 5.5329 9.6000 050373 050625

5 zip1 . . . . 0 . . 132 13 5.5329 5.8266 050373 050660

6 zip1 . . . . 1 . . 132 75 1.5756 3.5416 050091 050149

7 zip1 . . . 0 . . . 136 78 1.5756 3.5416 050091 050149

8 zip1 . . . 1 . . . 136 10 5.8266 7.9353 050660 050751

9 zip1 . . 0 . . . . 144 81 1.5756 3.5416 050091 050149

10 zip1 . . 1 . . . . 144 7 5.5329 6.9041 050373 050717

11 zip1 . 0 . . . . . 160 48 1.5756 4.3449 050091 050663

12 zip1 . 1 . . . . . 160 40 3.5416 4.6257 050149 050471

13 zip1 0 . . . . . . 192 47 3.5416 4.6257 050149 050471

14 zip1 1 . . . . . . 192 41 1.5756 4.3449 050091 050663

3. merge product and product\_alt on analysis to identify dsame and dopp for each mprovno-zip9 pair (ref: analysis\_full)

Obs mprovno zip9 hlat hlon dist hchar1 hchar2 hchar3 hchar4 hchar5 hchar6 hchar7 dsame1

1 050018 zip1 34.0646 -118.239 5.4022 1 0 0 0 1 0 0 1.57557

2 050040 zip1 34.3242 -118.448 25.829 0 0 1 0 1 0 1 3.54160

3 050056 zip1 34.6894 -118.160 48.775 0 1 0 0 1 0 0 3.54160

4 050058 zip1 34.1284 -118.258 9.7613 0 1 0 0 1 0 0 3.54160

5 050063 zip1 34.0956 -118.292 7.8262 1 0 0 0 1 0 1 1.57557

6 050078 zip1 33.7379 -118.306 17.478 0 1 0 0 1 0 0 3.54160

7 050091 zip1 33.9892 -118.225 1.5756 1 0 0 0 1 0 0 4.34488

8 050096 zip1 34.0704 -117.945 18.490 1 0 0 1 0 0 1 1.57557

9 050103 zip1 34.0511 -118.217 4.8468 0 1 0 0 1 0 1 3.54160

10 050104 zip1 33.9311 -118.204 4.7611 0 1 0 0 1 0 0 3.54160

Obs dsame2 dsame3 dsame4 dsame5 dsame6 dsame7 dopp1 dopp2 dopp3 dopp4 dopp5 dopp6 dopp7

1 1.57557 1.57557 1.57557 1.57557 1.57557 1.57557 3.54160 3.54160 5.53286 5.82656 5.53286 5.53286 3.54160

2 1.57557 5.53286 1.57557 1.57557 1.57557 3.54160 1.57557 3.54160 1.57557 5.82656 5.53286 5.53286 1.57557

3 3.54160 1.57557 1.57557 1.57557 1.57557 1.57557 1.57557 1.57557 5.53286 5.82656 5.53286 5.53286 3.54160

4 3.54160 1.57557 1.57557 1.57557 1.57557 1.57557 1.57557 1.57557 5.53286 5.82656 5.53286 5.53286 3.54160

5 1.57557 1.57557 1.57557 1.57557 1.57557 3.54160 3.54160 3.54160 5.53286 5.82656 5.53286 5.53286 1.57557

6 3.54160 1.57557 1.57557 1.57557 1.57557 1.57557 1.57557 1.57557 5.53286 5.82656 5.53286 5.53286 3.54160

7 4.34488 3.54160 3.54160 3.54160 3.54160 4.34488 3.54160 3.54160 5.53286 5.82656 5.53286 5.53286 3.54160

8 1.57557 1.57557 5.82656 5.53286 1.57557 3.54160 3.54160 3.54160 5.53286 1.57557 1.57557 5.53286 1.57557

9 3.54160 1.57557 1.57557 1.57557 1.57557 3.54160 1.57557 1.57557 5.53286 5.82656 5.53286 5.53286 1.57557

10 3.54160 1.57557 1.57557 1.57557 1.57557 1.57557 1.57557 1.57557 5.53286 5.82656 5.53286 5.53286 3.54160

4. final analysis data: differential distance calculations (ref: final\_analysis)

Obs mprovno zip9 hlat hlon dist hchar1 hchar2 hchar3 hchar4 hchar5 hchar6 hchar7 ddsame1

1 050018 zip1 34.0646 -118.239 5.4022 1 0 0 0 1 0 0 3.8267

2 050040 zip1 34.3242 -118.448 25.829 0 0 1 0 1 0 1 22.2871

3 050056 zip1 34.6894 -118.160 48.775 0 1 0 0 1 0 0 45.2333

4 050058 zip1 34.1284 -118.258 9.7613 0 1 0 0 1 0 0 6.2197

5 050063 zip1 34.0956 -118.292 7.8262 1 0 0 0 1 0 1 6.2506

6 050078 zip1 33.7379 -118.306 17.478 0 1 0 0 1 0 0 13.9360

7 050091 zip1 33.9892 -118.225 1.5756 1 0 0 0 1 0 0 -2.7693

8 050096 zip1 34.0704 -117.945 18.490 1 0 0 1 0 0 1 16.9147

9 050103 zip1 34.0511 -118.217 4.8468 0 1 0 0 1 0 1 1.3052

10 050104 zip1 33.9311 -118.204 4.7611 0 1 0 0 1 0 0 1.2195

Obs ddsame2 ddsame3 ddsame4 ddsame5 ddsame6 ddsame7 ddopp1 ddopp2 ddopp3 ddopp4 ddopp5 ddopp6 ddopp7

1 3.8267 3.8267 3.8267 3.8267 3.8267 3.8267 1.8606 1.8606 -0.1306 -0.4243 -0.1306 -0.1306 1.8606

2 24.2531 20.2958 24.2531 24.2531 24.2531 22.2871 24.2531 22.2871 24.2531 20.0021 20.2958 20.2958 24.2531

3 45.2333 47.1993 47.1993 47.1993 47.1993 47.1993 47.1993 47.1993 43.2420 42.9483 43.2420 43.2420 45.2333

4 6.2197 8.1857 8.1857 8.1857 8.1857 8.1857 8.1857 8.1857 4.2284 3.9347 4.2284 4.2284 6.2197

5 6.2506 6.2506 6.2506 6.2506 6.2506 4.2846 4.2846 4.2846 2.2933 1.9996 2.2933 2.2933 6.2506

6 13.9360 15.9020 15.9020 15.9020 15.9020 15.9020 15.9020 15.9020 11.9447 11.6510 11.9447 11.9447 13.9360

7 -2.7693 -1.9660 -1.9660 -1.9660 -1.9660 -2.7693 -1.9660 -1.9660 -3.9573 -4.2510 -3.9573 -3.9573 -1.9660

8 16.9147 16.9147 12.6637 12.9574 16.9147 14.9486 14.9486 14.9486 12.9574 16.9147 16.9147 12.9574 16.9147

9 1.3052 3.2712 3.2712 3.2712 3.2712 1.3052 3.2712 3.2712 -0.6860 -0.9797 -0.6860 -0.6860 3.2712

10 1.2195 3.1856 3.1856 3.1856 3.1856 3.1856 3.1856 3.1856 -0.7717 -1.0654 -0.7717 -0.7717 1.2195

##Code contents:

###SAS files:

-analysis\\_denom.sas: duplicate observation and collapse TM/MA weights to weight variable, add hicbic characteristics

-analysis\\_rcc\\_byhicbic.sas: match cost reports to medpar stays, calculating artificial costs for MA stays and revenue.

collapse to the hicbic, MA level for a regression-ready file

-bene\\_per\\_zip.sas: construct a dataset for beneficiaries per zip code

-clean\\_hosp\\_geocode.sas: clean version of hospital geocodes and zip codes

-construct\\_denom.sas: construct denominator from raw files

-construct\\_medpar.sas: construct medpar from raw files

-construct\\_regions.sas: create regions for hospital market structure regressions

-analysis\\_medpar\\_HCC\\_byhicbic.sas: recode ICD-9 (int'l classification of disease codes) from MedPAR to HCC (hazard characteristic code) dummies

-hmo\\_status\\_byhicbic.sas: assign TM or MA status and weights to all benes based on hmo, buy-in, and death status from denom

-import\\_cty\\_risk.sas: import the cty\_risk benchmarks

-npr\\_ccr\\_bymprovno.sas: construct net payment ratio(npr) and cost to charge ratio(ccr) from cost reports

-tables.sas: [sample] code for creating various summary/frequency tables throughout the project

###STATA files:

-analysis.do: construct and run choice regressions

-HHI\\_mprovno\\_sys.do: construct NEW HHI hospital market stucture variables and intermediaries that takes into account hospital systems

-hosp\\_mrkt\\_strct.do: do file for constructing hospital market structure variables and intermediaries: HHI\\_pat\\_k\\_star, CAP\\_pat\\_k\\_star, hosp\\_char\\_h\\_pat\\_k\\_star

-IV.do: first-pass IVs of MA choice on hospital market structure variables

-IV0.do: merging files together for IV regressions, running 0-stage regression to calculate MA\_hat for IV construction

-zip\\_fips.do: create zip-to-fips xwalk

-zip\\_to\\_region.do: create zip-to-region xwalk

###proc contents:

lists of variables in datasets

###xwalks:

various crosswalks used for regressions

List of files in /analysis\_stata folder:

aha\_extract2008.dta: raw data from Jean with hospital system id to mprovno

aha\_sysid.do: cleans sysid from Jean

aha\_sysid.dta: cleaned hospital system id to mprovno

/analysis: folder that contains analysis1-14 data that comes from analysis.do

analysis.do: hospital choice regressions

analysis.ster: regression results from analysis.do

base.dta: intermediary file for constructing full datasets in IV0.do

benchmark\_new.dta: benchmarks by county (IVs)

bene\_per\_zip.dta: Medicare beneficiaries per zipcode (for descriptive stats purposes)

/ca: folder with CA-only data

base\_ca.dta: base dataset with CA based beneficiaries

IV\_CA.do: IV regressions for CA

npr\_CA.dta: new OSHPD revenue numbers

denom\_clean.dta: intermediary file for constructing full datasets in IV0.do

denom.dta: beneficiary characteristics from CMS denominator

/dis: folder with discharge data

iv\_bydischarge\_b\_div\_ffs.dta: data for regressions using b/ffs

iv\_bydischarge\_b\_minus\_ffs.dta: data for regressions using b-ffs

IV\_dis.do: IV regressions for discharge based data

probit0\_bydischarge.dta: data that feeds into zero-stage regressions

tpm2.ster/tpm2.txt/tpm\_dis.xml: regression output

ffs\_spend.do: construct the new IVs

hcc.dta: clean hicbic level hcc (diagnosis group) dummies

HHI\_mprovno\_sys.dta: hospital market structure variables by zip

/hic: folder with hicbic level data

iv\_byhicbic\_b\_div\_ffs.dta: data for regressions using b/ffs

iv\_byhicbic\_b\_minus\_ffs.dta: data for regressions using b-ffs

IV\_hic.do: IV regressions for hicbic based data

probit0\_byhicbic.dta: data that feeds into zero-stage regressions

tpm2.ster/tpm2.txt/tpm\_hic.xml: regression output

hosp\_chars.dta: hospital characteristics by mprovno

hosp\_mrkt\_struct.do: constructs hospital market structure variables

hosp\_mrkt\_zip.dta: hospital market structure variables by zip code

hosp\_region.dta: hospital geocode info

IV0.do: zero-level regressions

IV\_CA.do: regression file for CA-only regressions

IV.do: full two-part model regressions

/master: folder with master1-master14 files. output from hosp\_mrkt\_strct.do

medpar\_hcc\_byhicbic.dta: raw hicbic level hcc (diagnosis group) dummies

rcc\_bydischarge.dta: revenue charge cost data from medpar by discharge

rcc\_byhicbic.dta: revenue charge cost data from medpar by hicbic

/stata: folder with stata1-14.xpt and .dta files. input and output from analysis.do

variables.txt: list of all variables for IV regression (not up to date)

zip\_hrr\_wcityname\_xwalk.dta: xwalk for zip, hrr, hrrstate, hrrcity, and region

List of files in /workingdata folder:

2007\_HCC.txt: ICD-9 HCC xwalk file

aha\_extract2007.dta: aha hospital geocodes for FY2007

aha\_extract2008.dta: aha hospital geocodes for FY2008

analysis\_HCC\_byhicbic.sas: recoding ICD-9 (int'l classification of disease codes) from MedPAR to HCC (hazard characteristic code) dummies

analysis\_rcc\_bydischarge.sas: match cost reports to medpar stays, calculating artificial costs for MA stays and revenue, collapse to the hicbic, MA level for a regression-ready file

analysis\_rcc\_byhicbic.sas: match cost reports to medpar stays, calculating artificial costs for MA stays and revenue

bene\_per\_zip.sas7bdat:

cty\_risk.sas7bdat/cty\_risk.txt/cty\_risk.xpt: county benchmark payment rate in various file forms

denom100\_2008.sas7bdat: full denominator file

hcc.sas7bdat: ICD-9 HCC xwalk file

hicbic\_medparonly.sas7bdat: intermediate dataset in analysis\_rcc\_byhicbic.sas

hicbic\_medpar.sas7bdat: intermediate dataset in analysis\_rcc\_byhicbic.sas

hicbic.sas7bdat: duplicate entries with different weights for those benes with MA and TM enrollment

hosp\_chars\_new.sas7bdat: cost report financials from chris

hosp\_costs.sas7bdat: ccr and npr

hosp\_geocode\_clean.sas7bdat: cleaned hospital geocodes

hosp\_geocodes.sas7bdat: raw hospital geocodes from scott

hosp\_region.sas7bdat: hospital to region xwalk

hosp.sas7bdat: hospital geocodes only

medpar100.sas7bdat: full medpar file

medpar\_group.sas7bdat: collapsed hospital-choice-demographics by zip

medpar\_hcc\_byhicbic.sas7bdat: hcc byhicbic

medpar\_hmo\_costs.sas7bdat: medpar with merged ccr and npr by discharge date

medpar\_hmo.sas7bdat: intermediate dataset for ma status determination in analysis\_rcc\_bydischarge.sas

medpar\_region.sas7bdat: all valid stays in medpar with regional assignment

rcc\_bydischarge.sas7bdat: revenue charge cost by discharge

rcc\_byhicbic.sas7bdat: revenue charge cost by hicbic

trans\_medparonly.sas7bdat: intermediate dataset for ma status determination in analysis\_rcc\_bydischarge.sas

trans.sas7bdat: assign weights for those benes with MA and TM enrollment

zip\_region.sas7bdat: all valid zips in medpar with regional assignment

zip\_to\_region\_xwalk.txt: five-digit zip to region xwalk

zip\_to\_r.sas7bdat: sas version of zip\_to\_region

