= V4 Outline MultiLine NoSorting TabWidth=30

H="directory"

// Global Macros use $ symbol to be called.

global nhats "D:\NHATS\Shared\base\_data\NHATS cleaned"

global medi "D:\NHATS\Shared\base\_data\CMS\_claims\Stata"

//Intermediate Data Path

global intpath "D:\NHATS\Projects\Ankuda\_RP1\data\int\_data"

// Final Data Path

global datapath "D:\NHATS\Projects\Ankuda\_RP1\data\final\_data"

//Log files path

global outpath "D:\HRS\Projects\surgery\zc\_surgery\_and\_serious\_illness\output\in\_progress"

PROJECT GOAL:

Examine incident dementia via Medicare claims data vs NHATS annual cognitive assessment

1. nhats 2011-2018

2. incident claims-based dementia cohort:

a. ffs Medicare for 2011 & 2012

b. no-claims-based diagnosis of dementia from 2011 to 2012

c. 1+ claims-based diagnosis of dementia from 2013 to 2018

3. NHATS variable: prob\_dem

H="get study sample"

KNOWLEDGE POINT 1: CROSSWALK NHATS TO CLAIMS DATA

use "${nhats}\sp\_round\_1\_10.dta",clear

//exclude sample replenished in wave 5

drop if yearsample==2015

//create an index year from nhats interview date

gen index\_year=ivw\_year

gen index\_date=ivw\_date

save "${intpath}\nhats\_index.dta",replace

//xwalk nhats and claims data

merge m:1 spid using "${medi}\xwalk\_2018.dta"

keep if \_merge==3

drop \_merge

save "${intpath}\nhats\_index\_xwalk.dta",replace

//create ffs

use "${medi}\mbsf\_06\_18.dta",clear

gen ffs=bene\_hmo\_cvrage\_tot\_mons==0

save "${intpath}\mbsf\_ffs.dta",replace

//merge nhats and mbsf

use "${intpath}\nhats\_index\_xwalk.dta",clear

joinby bid\_nhats\_1 using "${intpath}\mbsf\_ffs.dta"

duplicates report bid\_nhats\_1 wave

return list

duplicates drop bid\_nhats\_1 wave,force

save "${intpath}\nhats\_ffs.dta",replace

H="merge dx codes from claims to nhats"

KNOWLEDGE POINT 2: GET ICD-BASED DIAGNOSIS FROM THE CLAIMS

ip: ip\_06\_18

other claims data: x\_09\_18

//get icd-based dementia from the claims

foreach x in ip sn hh hs op pb dm {

di "`x'"

if "`x'"=="ip" {

use bid\_nhats\_1 admit\_date icd\_dgns\_cd\* using "D:/NHATS/Shared/base\_data/CMS\_claims/Stata/`x'\_06\_18.dta", clear

}

else {

use bid\_nhats\_1 admit\_date icd\_dgns\_cd\* using "D:/NHATS/Shared/base\_data/CMS\_claims/Stata/`x'\_09\_18.dta", clear

}

gen file\_origin="`x'"

rename icd\_dgns\_cd\* dgnscd\*

rename dgnscd\* diag\*

gen dementia\_claims=0

save "${intpath}\claims\_`x'.dta",replace

}

diagnosis 1-12: pb; dm

dignosis 1-25: all other claims data

foreach y in ip sn hh hs op pb dm {

if "`y'"=="pb" | "`y'"=="dm" {

use "${intpath}\claims\_`y'.dta",clear

foreach i of varlist diag1-diag12 {

replace dementia\_claims=1 if inlist(`i',"F0150","F0151","F0280","F0281","F0390","G3101","G3109") | inlist(`i',"R4181","G309","G300","G301","G308","G311","G312","G914")

}

}

else {

use "${intpath}\claims\_`y'.dta",clear

foreach k of varlist diag1-diag25 {

replace dementia\_claims=1 if inlist(`k',"F0150","F0151","F0280","F0281","F0390","G3101","G3109") | inlist(`k',"R4181","G309","G300","G301","G308","G311","G312","G914")

}

}

keep if dementia\_claims==1

joinby bid using "${intpath}\nhats\_ffs.dta"

save "${intpath}\dx\_nhats`y'.dta",replace

}

H="\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*"

KNOWLEDGE POINT 3: CREATE CONTINUOUS FFS

H="indicator for ffs in 2011/2012"

\*step 1: make a flag for individuals with FFS in 2011 and 2012

use "D:\NHATS\Shared\raw\CMS\NHATS CMS DUA 28016\Merged\STATA\mbsf\_06\_17.dta"

sort bene\_id year

\*make a variable that is 1/0 for 2011 based on FFS full year status

gen ffs\_fullyear\_2011=1 if year==2011

foreach var of varlist bene\_hmo\_ind\_01-bene\_hmo\_ind\_12 {

replace ffs\_fullyear\_2011=0 if `var'!= "0" & `var'!="4" & year==2011

}

tab ffs\_fullyear\_2011

\*same for 2012

gen ffs\_fullyear\_2012=1 if year==2012

foreach var of varlist bene\_hmo\_ind\_01-bene\_hmo\_ind\_12 {

replace ffs\_fullyear\_2012=0 if `var'!= "0" & `var'!="4" & year==2012

}

tab ffs\_fullyear\_2012

\*make a variable that is 0/1 for each person with full ffs in both 2011 and 2012

sort bene\_id year

bysort bene\_id: egen max\_2011=max(ffs\_fullyear\_2011)

bysort bene\_id: egen max\_2012=max(ffs\_fullyear\_2012)

gen ffs\_2011\_2012=0

replace ffs\_2011\_2012=1 if max\_2011==1 & max\_2012==1

replace ffs\_2011\_2012=. if year!=2011

\*the above step means that there is only 1 observation for each person, and only an observation if they had an MBSF in 2011... so effectively no duplicates per person

tab ffs\_2011\_2012

\*now i'm going to only save this indicator

keep bene\_id ffs\_2011\_2012

drop if ffs\_2011\_2012==.

\*now merge to get spid and drop the bene\_id

merge 1:1 bene\_id using "D:\NHATS\Shared\raw\CMS\NHATS CMS DUA 28016\Crosswalks\xwalk\_2016.dta", keepusing(spid)

keep if \_merge==3

drop \_merge

save "D:\NHATS\Projects\Ankuda\_RP1\data\int\_data\cka\_ffs ind.dta", replace

H="indicator for date of first dementia diagnosis"

foreach x in ip sn hh hs op pb dm {

di "`x'"

if "`x'"=="ip" {

use bid\_nhats\_1 admit\_date icd\_dgns\_cd\* using "D:/NHATS/Shared/base\_data/CMS\_claims/Stata/`x'\_06\_18.dta", clear

}

else {

use bid\_nhats\_1 admit\_date icd\_dgns\_cd\* using "D:/NHATS/Shared/base\_data/CMS\_claims/Stata/`x'\_09\_18.dta", clear

}

gen file\_origin="`x'"

rename icd\_dgns\_cd\* dgnscd\*

rename dgnscd\* diag\*

gen dementia\_claims=0

save "${intpath}\claims\_`x'.dta",replace

}

global intpath "D:\NHATS\Projects\Ankuda\_RP1\data\int\_data"

KNOWLEDGE POINT 4: CAPTURE ALL THE DEMENTIA DIAGNOSIS

foreach y in ip sn hh hs op pb dm {

if "`y'"=="pb" | "`y'"=="dm" {

use "${intpath}\claims\_`y'.dta",clear

foreach i of varlist diag1-diag12 {

replace dementia\_claims=1 if inlist(`i',"3310","33111","33119","3312","3317") | inlist(`i',"33182","33189", "2900", "29010", "29011", "29012", "29013") | inlist(`i', "29020" "29021", "2903", "29040", "29041", "29042", "29043", "2908") | inlist(`i', "2940", "29410", "29411", "29420", "29421", "797") | inlist(`i',"F0150", "F0151", "F0280", "F0281", "F0390", "F0391") | inlist(`i', "F04", "G300", "G301", "G308", "G309", "G3101", "G3109") | inlist(`i', "G3183", "G311", "G312", "R4181")

}

}

else {

use "${intpath}\claims\_`y'.dta",clear

foreach k of varlist diag1-diag25 {

replace dementia\_claims=1 if inlist(`k',"3310","33111","33119","3312","3317") | inlist(`k',"33182","33189", "2900", "29010", "29011", "29012", "29013") | inlist(`k', "29020" "29021", "2903", "29040", "29041", "29042", "29043", "2908") | inlist(`k', "2940", "29410", "29411", "29420", "29421", "797") | inlist(`k',"F0150", "F0151", "F0280", "F0281", "F0390", "F0391") | inlist(`k', "F04", "G300", "G301", "G308", "G309", "G3101", "G3109") | inlist(`k', "G3183", "G311", "G312", "R4181")

}

}

keep if dementia\_claims==1

save "${intpath}\cka\_dx\_`y'.dta",replace

}

\*instead of save "${intpath}\dx\_nhats`y'.dta",replace save "${intpath}\cka\_dx\_`y'.dta",replace

cd "${intpath}"

use cka\_dx\_ip.dta, clear

append using cka\_dx\_sn cka\_dx\_hh cka\_dx\_hs cka\_dx\_op cka\_dx\_pb cka\_dx\_dm

\*now for each person, make a flag for the first dementia case

bysort bid\_nhats\_1: egen first\_dem= min(admit\_date)

\*now only keep each observation if it is the first dementia dx

count

keep if admit\_date==first\_dem

count

\*we don't need all the diagnosis vars, so dropping them

keep bid\_nhats\_1 first\_dem file\_origin

\*we seem to have some duplicates, I'm not sure why

sort bid\_nhats\_1 file\_origin first\_dem

quietly by bid\_nhats\_1 file\_origin first\_dem: gen dup=cond(\_N==1,0,\_n)

keep if dup==0 | dup==1

drop dup

\*there are also duplicates where there is the same data but 2 different files-- like same date for IP and PB, for IP and OP. I'm going to just drop the first one. from approach below I can see that out of 3,626 unique individuals, 312 have 2 clams, 5 have 3

bysort bid\_nhats\_1: gen n=\_n

tab n

keep if n==1

drop n

\*swap out bid for spid

merge 1:1 bid\_nhats\_1 using "D:\NHATS\Shared\raw\CMS\NHATS CMS DUA 28016\Crosswalks\xwalk\_2018.dta"

keep if \_merge==3

drop \_merge

drop bid\_nhats\_1

save "${intpath}\cka\_first dem.dta",replace

H="merging those claims-based indicators with nhats"

\*now put it all together: merge nhats with indicator for ffs in 2011/2012 and indicator for first dementia-- going to use rounds 1-8 bc we only have claims data through 2018

use "D:\NHATS\Shared\base\_data\NHATS cleaned\sp\_round\_1\_8.dta",clear

merge m:1 spid using "D:\NHATS\Projects\Ankuda\_RP1\data\int\_data\cka\_ffs ind.dta"

drop \_merge

merge m:1 spid using "D:\NHATS\Projects\Ankuda\_RP1\data\int\_data\cka\_first dem.dta"

drop \_merge

H="variable cleaning and analysis"

use "D:\NHATS\Projects\Ankuda\_RP1\data\final\_data\cka\_working data.dta",clear

\*now work on making a "keep"indicator for people in the study- those in 2011 cohort who had ffs in 2011 and 2012. will also limit to those in wave 1 so we can see the number of individuals, not number of observations

gen keep=0

replace keep=1 if yearsample==2011 & ffs\_2011\_2012==1

replace keep=0 if wave!=1

tab keep

\*we start with 5,346 individuals

\*now make dementia category indicator: 1) dementia from first 2 years, 2) incident dementia, 3) never dementia. note that there are 19,359 days between 1/1/1960 and 1/1/2013

format first\_dem %td

gen claims\_dementia\_cat=.

replace claims\_dementia\_cat=3 if first\_dem==.

replace claims\_dementia\_cat=1 if first\_dem!=. & first\_dem<19359

replace claims\_dementia\_cat=2 if first\_dem!=. & first\_dem>19359

label define dc 1 "prevalent dem" 2 "incident dem" 3 "never dem"

label values claims\_dementia\_cat dc

tab claims\_dementia\_cat if keep==1

\*now make variable for ever any survey-based dementia

sort spid wave

bysort spid: egen any\_dem=max(prob\_dem)

tab claims\_dementia\_cat if keep==1

tab any\_dem if keep==1

\*graph to show percentage of dem in nhats survey in each claims dem category

\*add tracker weight(2011 cohort)

svyset varunit [pweight=anfinwgt], strata(varstrat)

//subpop

\*unweighted

tabulate any\_dem claims\_dementia\_cat if keep==1,col

\*weighted

svy,subpop(keep): tabulate any\_dem claims\_dementia\_cat,col

//overall pop

\*unweighted

tabulate any\_dem claims\_dementia\_cat,col

\*weighted

svy: tabulate any\_dem claims\_dementia\_cat,col

graph hbar (percent)[pweight=tr2011wgt], over(any\_dem, relabel(1 "no survey dem" 2 "has survey dem")) by(claims\_dementia\_cat) blabel(total, format(%9.2f) gap(\*.1)) ytitle("Weighted percentage of NHATS and claims-based dementia (full sample)")

svy,subpop(keep): tabulate any\_dem

svy,subpop(keep): tabulate claims\_dementia\_cat

H="\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*"

H="\*\*DONT RUN\*\*relevant code how prob\_dem was created"

\*\*\*\*\*\*\*\*DONT RUN THE TOP PART\*\*\*\*\*\*\*\*codes are taken from NHATS\_Setup\_1\_9 as a reference

/\*

//Initialize 3 category dementia variable

//probable dementia, possible dementia, no dementia

//probable = diagnosis reported or 2+ ad8 questions (proxy) or <1.5 SD below mean

gen dem\_3\_cat=-1 if ivw\_type!=1 //set to na if not SP interview

replace dem\_3\_cat=1 if sr\_dementia\_ever==1 //if reported directly

replace dem\_3\_cat=1 if dem\_3\_cat==. & dem\_via\_proxy==1 //if proxy ivw indicates dem

replace dem\_3\_cat=3 if dem\_3\_cat==. & dem\_via\_proxy==0 & speaktosp==2 //proxy ind no dem

//now apply to dementia 3 category variable

replace dem\_3\_cat=1 if dem\_3\_cat==. & inlist(speaktosp,1,-1) & inlist(domain65,2,3)

replace dem\_3\_cat=2 if dem\_3\_cat==. & inlist(speaktosp,1,-1) & domain65==1

replace dem\_3\_cat=3 if dem\_3\_cat==. & inlist(speaktosp,1,-1) & domain65==0

la def dem3 1"Probable dementia" 2"Possible dementia" 3"No dementia"

la val dem\_3\_cat dem3

la var dem\_3\_cat "Dementia likelihood, 3 categories"

\*/

\*create a new variable indicate how dememtia status was found out

\*dem\_via\_proxy comes from data from the following path: D:\NHATS\Shared\base\_data\NHATS cleaned\working\round\_1\_8\_cleanv1

gen dem\_report=.

replace dem\_report=0 if sr\_dementia\_ever==1 //if reported directly

replace dem\_report=1 if dem\_3\_cat==. & dem\_via\_proxy==1 //if proxy indicates dem

replace dem\_report=1 if dem\_3\_cat==. & dem\_via\_proxy==0 & speaktosp==2 //proxy ind no dem

label define 0 "self-report dementia" 1 "survey assessment have dementia"

label variable dem\_report "how dementia was found out"

H="prob\_dem, time difference & demographics of 3 groups"

use "D:\NHATS\Projects\Ankuda\_RP1\data\final\_data\cka\_working data.dta",clear

\*now work on making a "keep"indicator for people in the study- those in 2011 cohort who had ffs in 2011 and 2012. will also limit to those in wave 1 so we can see the number of individuals, not number of observations

gen keep=0

replace keep=1 if yearsample==2011 & ffs\_2011\_2012==1

replace keep=0 if wave!=1

tab keep

\*we start with 5,346 individuals

\*now make dementia category indicator: 1) dementia from first 2 years, 2) incident dementia, 3) never dementia. note that there are 19,359 days between 1/1/1960 and 1/1/2013

format first\_dem %td

gen claims\_dementia\_cat=.

replace claims\_dementia\_cat=3 if first\_dem==.

replace claims\_dementia\_cat=1 if first\_dem!=. & first\_dem<19359

replace claims\_dementia\_cat=2 if first\_dem!=. & first\_dem>19359

label define dc 1 "prevalent dem" 2 "incident dem" 3 "never dem"

label values claims\_dementia\_cat dc

tab claims\_dementia\_cat if keep==1

\*now make variable for ever any survey-based dementia

sort spid wave

bysort spid: egen any\_dem=max(prob\_dem)

tab claims\_dementia\_cat if keep==1

tab any\_dem if keep==1

KNOWLEDGE POINT 5 track time difference between claims dementia and nhats dementia

tab first\_dem //date of first claims dem diag

tab any\_dem //if ever had nhats dem

tab ivw\_date if any\_dem==1

gen time\_diff=.

replace time\_diff=0 if first\_dem<=ivw\_date

replace time\_diff=1 if first\_dem>ivw\_date

label define time 0 "claims dem first" 1 "survey dem first"

label values time\_diff time

tab time\_diff if keep==1

gen actual\_time\_diff=.

replace actual\_time\_diff=abs(ivw\_date-first\_dem)

tab actual\_time\_diff

gen time\_diff\_cat=.

replace time\_diff\_cat=0 if actual\_time\_diff<=365

replace time\_diff\_cat=1 if actual\_time\_diff >365 & actual\_time\_diff <1095

replace time\_diff\_cat=2 if actual\_time\_diff>=1095

label define time\_cat 0 "less than or equal to one year" 1 "from 1-3 years" 2 "greater than or equal to 3 years"

label values time\_diff\_cat time\_cat

tab time\_diff\_cat if keep==1

tab time\_diff\_cat time\_diff if keep==1

\*run basic demographics for the 3 groups

\*age

tabstat age if keep==1, by(claims\_dementia\_cat)

oneway claims\_dementia\_cat age if keep==1

\*sex

anova female claims\_dementia\_cat if keep==1

tabstat female if keep==1, by(claims\_dementia\_cat)

tab female claims\_dementia\_cat if keep==1,chi2

\*education

anova educ\_hs\_ind claims\_dementia\_cat if keep==1

tabstat educ\_hs\_ind if keep==1, by(claims\_dementia\_cat)

\*income

tabstat imputed\_inc5 if keep==1, by(claims\_dementia\_cat)

tabulate claims\_dementia\_cat if keep==1, summarize(imputed\_inc5)

\*self-reported health

anova srh\_fp claims\_dementia\_cat if keep==1

tabstat srh\_fp if keep==1, by(claims\_dementia\_cat)

//demographics for claims-based dementia

preserve

tab keep claims\_dementia\_cat

keep if keep==1 & claims\_dementia\_cat==2

\*concurrent

sum age

tab female

tab race\_cat

tab educ\_hs\_ind

sum imputed\_inc5

tab srh\_fp

tab medicaid

tab adl\_index

tab nhres

restore

\*one-year after incident claims-based dementia

gen oneyearafter=ivw\_date-first\_dem>0 & ivw\_date-first\_dem<=365

tab oneyearafter if keep==1 & claims\_dementia\_cat==2