

***** Data Cleaning *****

*** combine minus3 to 0 to get hospitalizations, ED visits, and outpatient visits in the past year

```
use "Q:\PGMS\mgeorge\GC project\data files\Updated data 9_2019\DX_cnt_5y.dta"
rename PATID, lower
keep patid index_dt in_cnt_minus3 in_cnt_minus2 in_cnt_minus1 in_cnt_0 ed_cnt_minus3
ed_cnt_minus2 ed_cnt_minus1 ed_cnt_0 out_cnt_minus3 out_cnt_minus2 out_cnt_minus1 out_cnt_0
out_rheum_cnt_minus3 out_rheum_cnt_minus2 out_rheum_cnt_minus1 out_rheum_cnt_0
gen hospitalizations = in_cnt_minus3 + in_cnt_minus2 + in_cnt_minus1 + in_cnt_0
gen ed_visits = ed_cnt_minus3+ ed_cnt_minus2+ ed_cnt_minus1+ ed_cnt_0
gen outpatient_visits = out_cnt_minus3+ out_cnt_minus2+ out_cnt_minus1+ out_cnt_0
gen outpatient_rheum = out_rheum_cnt_minus3+ out_rheum_cnt_minus2+ out_rheum_cnt_minus1+
out_rheum_cnt_0
label variable hospitalizations "# hosp past year"
label variable ed_visits "# ED visits past year"
label variable outpatient_visits "# outpatiet (av) past year"
label variable outpatient_rheum "# outpatiet rheum (av66) past year"
```

```
gen hosp_cat = 1 if hospitalizations == 0
replace hosp_cat = 2 if hospitalizations == 1
replace hosp_cat = 3 if hospitalizations == 2
replace hosp_cat = 4 if hospitalizations >=3 & hospitalizations != .
label define hosp_catname 1 "none" 2 "1" 3 "2" 4 ">=3"
label values hosp_cat hosp_catname
```

```
gen ed_visit_cat = 1 if ed_visits == 0
replace ed_visit_cat = 2 if ed_visits == 1
replace ed_visit_cat = 3 if ed_visits == 2
replace ed_visit_cat = 4 if ed_visits >=3
label values ed_visit_cat hosp_catname
```

```
keep patid index_dt hospitalizations ed_visits outpatient_visits outpatient_rheum
save "Q:\PGMS\mgeorge\GC project\data files\updated data\raw\DX_cnt_5y_derived_baseline_data.dta"
```

```
**** use medication data from the last 3 months of the baseline period
use "Q:\PGMS\mgeorge\GC project\data files\updated data\raw\Exp_Rx_5y.dta"
rename PATID-Antibiotics_20, lower
keep patid index_dt hcq_0 lef_0 mtx_0 ssa_0 opioid_0 nsaid_0 antibiotics_0
save "Q:\PGMS\mgeorge\GC project\data files\updated data\raw\Exp_Rx_5y_derived_baseline_data.dta"
```

```
*** create steroid censoring dates:
use "Q:\PGMS\mgeorge\GC project\data files\updated data\GC_avg_dose_5y.dta"
drop gc_dose_minus3- gc_dose_minus1
reshape long gc_dose_, i(patid index_dt) j(gc_time)
gen gc_dose_cat = 1 if gc_dose_ == .
replace gc_dose_cat = 2 if gc_dose_ >0 & gc_dose_ <=5
replace gc_dose_cat = 3 if gc_dose_ >5 & gc_dose_ <=10
replace gc_dose_cat = 4 if gc_dose_ >10 & gc_dose_ != .
gen change = 0
bysort patid index_dt: replace change = 1 if gc_dose_cat[_n] != gc_dose_cat[_n-1] & _n != 1
bysort patid index_dt: replace change = change[_n] + change[_n-1] if _n != 1
drop if change != 0 // now we are left only with unchanged prescriptions
bysort patid index_dt: egen max_time = max(gc_time) // now get maximum follow up time for each
episode
drop if gc_time != 0 // just keep first
entry for each episode
gen gc_stable_follow_up = (max_time + 1)*90 // each time period with a stable dose
(adding one since we start with 0) times 90
save "Q:\PGMS\mgeorge\GC project\data files\updated data\raw\gc_stable_follow_up_time.dta"
```

**** Now merge files together:

```
use "Q:\PGMS\mgeorge\GC project\data files\Updated data 9_2019\raw\cohort.dta"
merge 1:1 patid index_dt using "Q:\PGMS\mgeorge\GC project\data files\Updated data
9_2019\raw\Covariates_sabr.dta"
drop _merge
merge 1:1 patid index_dt using "Q:\PGMS\mgeorge\GC project\data files\Updated data
9_2019\raw\Outcome_Sabr.dta"
drop _merge
```

```

merge 1:1 patid index_dt using "Q:\PGMS\mgeorge\GC project\data files\Updated data
9_2019\raw\Exp_Rx_5y_derived_baseline_data.dta"
drop _merge
merge 1:1 patid index_dt using "Q:\PGMS\mgeorge\GC project\data files\Updated data
9_2019\raw\DX_cnt_5y_derived_baseline_data.dta"
drop _merge
merge 1:1 patid index_dt using "Q:\PGMS\mgeorge\GC project\data files\Updated data
9_2019\raw\gc_stable_follow_up_time.dta"
drop _merge
merge 1:1 patid index_dt using "Q:\PGMS\mgeorge\GC project\data files\Updated data
9_2019\raw\GC_avg_dose_5y.dta"
drop _merge
merge 1:1 patid index_dt using "Q:\PGMS\mgeorge\GC project\data files\Updated data
9_2019\raw\coh_bs_provid.dta"
drop _merge

*****

gen year = year(index_dt)

** censor at ICD-10 change
gen right_censor_string = "8.31.2015"
gen right_censor_date = date(right_censor_string, "MDY")
format right_censor_date %d
drop right_censor_string

** drop 2006 (no time for 1 year baseline) and if no follow up possible because index date is >=
the right-censor (allowing 30 days of follow-up)
drop if index_dt >= right_censor_date
drop if year == 2006
drop if death_date < index_dt // a few people die before follow up begins

gen female = 1 if sex == "F"
replace female = 0 if sex == "M"

gen year2011_2015 = 0 if year < 2011
replace year2011_2015 = 1 if year >= 2011 & year != .

gen racecat = 1 if race == "1"
replace racecat = 2 if race == "2"
replace racecat = 3 if race == "3"
replace racecat = 4 if race == "4"
replace racecat = 5 if race == "5"
replace racecat = 6 if race == "6"
replace racecat = 7 if race == "0" | race == "X" | race == ""

label define racecatname 1 "White" 2 "Black" 3 "Other" 4 "Asian" 5 "Hispanic" 6 "Native American"
7 "Unknown"
label values racecat racecatname

gen racecat2 = 1 if race == "1"
replace racecat2 = 2 if race == "2"
replace racecat2 = 3 if race == "5"
replace racecat2 = 4 if race == "3" | race == "4" | race == "6" | race == "0" | race == "X" |
race == ""
label define racecatname2 1 "White" 2 "Black" 3 "Hispanic" 4 "Other/Unknown"
label values racecat2 racecatname2

gen agecat = 1 if age <40
replace agecat =2 if age >=40 & age < 55
replace agecat = 3 if age >=55 & age < 65
replace agecat = 4 if age >=65 & age < 70
replace agecat = 5 if age >=70 & age < 80
replace agecat = 6 if age >=80 & age != .
label define agecatname 1 "<40" 2 "40-55" 3 "55-65" 4 "65-70" 5 "70-80" 6 ">=80"
label values agecat agecatname

gen agecat2 = 1 if age <40
replace agecat2 =2 if age >=40 & age < 55
replace agecat2 = 3 if age >=55 & age < 65
replace agecat2 = 4 if age >=65 & age < 70

```

```

replace agecat2 = 5 if age >=70 & age < 75
replace agecat2 = 6 if age >=75 & age < 80
replace agecat2 = 7 if age >=80 & age != .
label define agecatname2 1 "<40" 2"40-55" 3 "55-65" 4 "65-70" 5 "70-75" 6 "75-80" 7 ">=80"
label values agecat2 agecatname2

gen agecat3 = 1 if age <40
replace agecat3 = 2 if age >=40 & age <45
replace agecat3 = 3 if age >=45 & age <50
replace agecat3 = 4 if age >=50 & age <55
replace agecat3 = 5 if age >=55 & age <60
replace agecat3 = 6 if age >=60 & age <65
replace agecat3 = 7 if age >=65 & age <70
replace agecat3 = 8 if age >=70 & age <75
replace agecat3 = 9 if age >=75 & age <80
replace agecat3 = 10 if age >=80 & age <85
replace agecat3 = 11 if age >=85 & age != .
label define agecatname3 1 "<40" 2"40-45" 3 "45-50" 4 "50-55" 5 "55-60" 6 "60-65" 7 "65-70" 8
"70-75" 9 "75-80" 10 "80-85" 11 ">=85"
label values agecat3 agecatname3

gen disability = 0 if bene_entlmt_rsn_orig == "0"
replace disability = 1 if (bene_entlmt_rsn_orig == "1" | bene_entlmt_rsn_orig == "2" |
bene_entlmt_rsn_orig == "3")

gen dual_elig = 0 if medicaid_elig == 0
replace dual_elig = 1 if medicaid_elig == 2
replace dual_elig = 2 if medicaid_elig == 1
label define dual_eligname 0 "not dual elig" 1 "partial dual" 2 "full dual"
label values dual_elig dual_eligname

gen hosp_inf_baseline_cat = 1 if hosp_inf_baseline_cnt == 0
replace hosp_inf_baseline_cat = 2 if hosp_inf_baseline_cnt == 1
replace hosp_inf_baseline_cat = 3 if hosp_inf_baseline_cnt == 2
replace hosp_inf_baseline_cat = 4 if hosp_inf_baseline_cnt >=3 & hosp_inf_baseline_cnt != .
label values hosp_inf_baseline_cat hosp_catname

gen drug2 = 1 if dtype == 1
replace drug2 = 2 if dtype == 11
replace drug2 = 3 if dtype == 10
replace drug2 = 4 if dtype == 9
replace drug2 = 5 if dtype == 8
replace drug2 = 6 if dtype == 7
replace drug2 = 7 if dtype == 6
replace drug2 = 8 if dtype == 2
replace drug2 = 9 if dtype == 5
replace drug2 = 10 if dtype == 4

label define drug2name2 1 "mtx" 2 "abatacept" 3 "adalimumab" 4 "certolizumab" 5 "etanercept" 6
"golimumab" 7 "infliximab" 8 "rituximab" 9 "tocilizumab" 10 "tofacitinib"
label values drug2 drug2name2

gen prior_biologics = biologics_pri_n - 1 if drug2 != 1 // remove the current biologic unless
it is a mtx treated patient
replace prior_biologics = biologics_pri_n if drug2 == 1
replace prior_biologics = 0 if prior_biologics == .

gen prior_biologics_cat = 1 if prior_biologics == 0
replace prior_biologics_cat = 2 if prior_biologics == 1
replace prior_biologics_cat = 3 if prior_biologics == 2
replace prior_biologics_cat = 4 if prior_biologics >=3 & prior_biologics != .
label define priorbiologicscatname 1 "none" 2 "1" 3 "2" 4 ">=3"
label values prior_biologics_cat priorbiologicscatname

replace cancer_scr_breast_prior = 0 if cancer_scr_breast_prior == .
replace cancer_scr_colon_prior = 0 if cancer_scr_colon_prior == .
replace cancer_scr_prostate_prior = 0 if cancer_scr_prostate_prior == .
replace cancer_scr_cervical_prior = 0 if cancer_scr_cervical_prior == .
replace cancer_scr_lung_prior = 0 if cancer_scr_lung_prior == .

replace cancer_scr_breast_ly = 0 if cancer_scr_breast_ly == .

```

```

replace cancer_scr_colon_ly = 0 if cancer_scr_colon_ly == .
replace cancer_scr_prostate_ly = 0 if cancer_scr_prostate_ly == .
replace cancer_scr_cervical_ly = 0 if cancer_scr_cervical_ly == .
replace cancer_scr_lung_ly = 0 if cancer_scr_lung_ly == .

encode division, gen(region)

** to clean up steroid data and make steroid categories:
foreach d in 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 minus1 minus2 minus3 {
replace gc_dose_`d' = 0 if gc_dose_`d' == .
}

foreach d in 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 minus1 minus2 minus3 {
gen gc_dose_`d'_cat = 1 if gc_dose_`d' == 0
replace gc_dose_`d'_cat = 2 if gc_dose_`d' >0 & gc_dose_`d' <=5
replace gc_dose_`d'_cat = 3 if gc_dose_`d' >5 & gc_dose_`d' <=10
replace gc_dose_`d'_cat = 4 if gc_dose_`d' >10 & gc_dose_`d' != .
}

label define gccatname 1 "none" 2 "<=5" 3 "5-10" 4 ">10"
label values gc_dose_*_cat gccatname

replace prov_66_av_ra = 0 if providerid_baseline == ""
egen provider_year = concat(providerid_baseline year) if providerid_baseline != "" // makes a
unique identifier for each provider in each year

** hospitalized infections in the 180 days prior to index date - will do sensitivity analysis
removing these patients
gen hosp_inf_baseline_180 = 1 if hosp_inf_baseline_180_cnt != .
replace hosp_inf_baseline_180 = 0 if hosp_inf_baseline_180_cnt == .

xtile income_cat = estinc, nq(5)
label variable income_cat "income category based on quintiles of median household income"
replace income_cat = 6 if income_cat == .
label variable income_cat "category based on quintiles of median household income, 6th cat
missing"
label define income_catname 1 "1st quintile" 2 "2nd quintile" 3 "3rd quintile" 4 "4th quintile" 5
"5th quintile" 6 "missing"
label values income_cat income_catname

*** follow-up time, ignoring steroid dose changes:
gen follow_up_time = enr_end_date - index_dt
replace follow_up_time = right_censor_date + 30 - index_dt if (right_censor_date + 30 - index_dt
<= follow_up_time)
replace follow_up_time = prsc_end_dt - index_dt if prsc_end_dt - index_dt <= follow_up_time
replace follow_up_time = death_date - index_dt if death_date != . & (death_date - index_dt <
follow_up_time)
label variable follow_up_time "follow up time ignoring steroid dose change"

** on stable dose for 1 year?
gen stable_1yr = 0
replace stable_1yr = 1 if gc_dose_0_cat == gc_dose_1_cat & gc_dose_0_cat == gc_dose_2_cat &
gc_dose_0_cat == gc_dose_3_cat
tab stable_1yr
tab gc_dose_0_cat stable_1yr, row
tab gc_dose_0_cat stable_1yr if follow_up_time >=360, row // at least 12 months of follow-up,
no change from the baseline dose

***** time to infection, censoring if steroid dose changes *****
gen time1 = enr_end_date - index_dt + 1 // add one day since the index date
is day 1 of follow-up in terms of our outcomes - need to see if baseline assessment included this
or not
replace time1 = right_censor_date + 31 - index_dt if (right_censor_date + 31 - index_dt <= time1)
// add 30 days for the month after R censor date. 1 more day because index date is day 1
replace time1 = prsc_end_dt - index_dt + 1 if prsc_end_dt - index_dt + 1 <= time1
// censor at prescription end date. add 1 because index date is day 1
replace time1 = gc_stable_follow_up if gc_stable_follow_up <=time1
// this has the +1 built in already
replace time1 = death_date - index_dt + 1 if death_date - index_dt + 1 <= time1
// here censoring at death although I can look at it as a competing risk
also

```

```

replace time1 = fst_hosp_inf_dt - index_dt + 1 if fst_hosp_inf_dt - index_dt + 1 <= time1
gen fail1 = 0
replace fail1 = 1 if fst_hosp_inf_dt - index_dt + 1 == time1
replace fail1 = 2 if death_date - index_dt + 1 == time1
label variable time1 "time to infection censoring w steroid dose change"

gen censor_reason = 1 if time1 == enr_end_date - index_dt + 1
replace censor_reason = 2 if time1 == right_censor_date + 31 - index_dt
replace censor_reason = 3 if time1 == prsc_end_dt - index_dt + 1
replace censor_reason = 4 if time1 == gc_stable_follow_up
replace censor_reason = 5 if time1 == death_date - index_dt + 1
replace censor_reason = 6 if time1 == fst_hosp_inf_dt - index_dt + 1
label define censorreason 1 "end enrollment" 2 "end dataset" 3 "DMARD course" 4 "GC dose change"
5 "death" 6 "outcome"
label values censor_reason censorreason

gen time1_primary = enr_end_date - index_dt + 1 // add one day since the
index date is day 1 of follow-up in terms of our outcomes - need to see if baseline assessment
included this or not
replace time1_primary = right_censor_date + 31 - index_dt if (right_censor_date + 31 - index_dt
<= time1_primary) // add 30 days for the month after R censor date. 1 more day because index
date is day 1
replace time1_primary = prsc_end_dt - index_dt + 1 if prsc_end_dt - index_dt + 1 <= time1_primary
// censor at prescription end date. add 1 because index date is day 1
replace time1_primary = gc_stable_follow_up if gc_stable_follow_up <= time1_primary
replace time1_primary = death_date - index_dt + 1 if death_date - index_dt + 1 <= time1_primary
replace time1_primary = fst_hosp_inf_prim_dt - index_dt + 1 if fst_hosp_inf_prim_dt - index_dt +
1 <= time1_primary
gen fail1_primary = 0
replace fail1_primary = 1 if fst_hosp_inf_prim_dt - index_dt + 1 == time1_primary
replace fail1_primary = 2 if death_date - index_dt + 1 == time1_primary
label variable time1_primary "time to infection primary dx censoring w steroid dose change"

***** time to infection, not censoring if steroid dose changes *****
gen time2 = enr_end_date - index_dt + 1 // add one day since the index date
is day 1 of follow-up in terms of our outcomes - need to see if baseline assessment included this
or not
replace time2 = right_censor_date + 31 - index_dt if (right_censor_date + 31 - index_dt <= time2)
// add 30 days for the month after R censor date. 1 more day because index date is day 1
replace time2 = prsc_end_dt - index_dt + 1 if prsc_end_dt - index_dt + 1 <= time2 // censor at
prescription end date. add 1 because index date is day 1
replace time2 = death_date - index_dt + 1 if death_date - index_dt + 1 <= time2
replace time2 = fst_hosp_inf_dt - index_dt + 1 if fst_hosp_inf_dt - index_dt + 1 <= time2
gen fail2 = 0
replace fail2 = 1 if fst_hosp_inf_dt - index_dt + 1 == time2
replace fail2 = 2 if death_date - index_dt + 1 == time2
label variable time2 "time to infection no censoring w steroid dose change"

gen time2_primary = enr_end_date - index_dt + 1 // add one day since the
index date is day 1 of follow-up in terms of our outcomes - need to see if baseline assessment
included this or not
replace time2_primary = right_censor_date + 31 - index_dt if (right_censor_date + 31 - index_dt
<= time2_primary) // add 30 days for the month after R censor date. 1 more day because index
date is day 1
replace time2_primary = prsc_end_dt - index_dt + 1 if prsc_end_dt - index_dt + 1 <= time2_primary
// censor at prescription end date. add 1 because index date is day 1
replace time2_primary = death_date - index_dt + 1 if death_date - index_dt + 1 <= time2_primary
replace time2_primary = fst_hosp_inf_prim_dt - index_dt + 1 if fst_hosp_inf_prim_dt - index_dt +
1 <= time2_primary
gen fail2_primary = 0
replace fail2_primary = 1 if fst_hosp_inf_prim_dt - index_dt + 1 == time2_primary
replace fail2_primary = 2 if death_date - index_dt + 1 == time2_primary
label variable time2_primary "time to infection primary dx no censoring w steroid dose change"

```

***** Creation of Provider Preference *****

```
*** merge the biologic and methotrexate yearly files
use "Q:\PGMS\mgeorge\GC project\data files\yearly prescriber data\Biologics_2007.dta"
gen methotrexate = 0
rename PATID- mtx_pri3m, lower
sort patid index_date
bysort patid: gen patient_number = _n
drop if patient_number != 1
drop patient_number
save "Q:\PGMS\mgeorge\GC project\data files\yearly prescriber data\Biologics_2007.dta", replace

use "Q:\PGMS\mgeorge\GC project\data files\yearly prescriber data\Biologics_2008.dta"
gen methotrexate = 0
rename PATID- mtx_pri3m, lower
sort patid index_date
bysort patid: gen patient_number = _n
drop if patient_number != 1
drop patient_number
save "Q:\PGMS\mgeorge\GC project\data files\yearly prescriber data\Biologics_2008.dta", replace

use "Q:\PGMS\mgeorge\GC project\data files\yearly prescriber data\Biologics_2009.dta"
gen methotrexate = 0
rename PATID- mtx_pri3m, lower
sort patid index_date
bysort patid: gen patient_number = _n
drop if patient_number != 1
drop patient_number
save "Q:\PGMS\mgeorge\GC project\data files\yearly prescriber data\Biologics_2009.dta", replace

use "Q:\PGMS\mgeorge\GC project\data files\yearly prescriber data\Biologics_2010.dta"
gen methotrexate = 0
rename PATID- mtx_pri3m, lower
sort patid index_date
bysort patid: gen patient_number = _n
drop if patient_number != 1
drop patient_number
save "Q:\PGMS\mgeorge\GC project\data files\yearly prescriber data\Biologics_2010.dta", replace

use "Q:\PGMS\mgeorge\GC project\data files\yearly prescriber data\Biologics_2011.dta"
gen methotrexate = 0
rename PATID- mtx_pri3m, lower
sort patid index_date
bysort patid: gen patient_number = _n
drop if patient_number != 1
drop patient_number
save "Q:\PGMS\mgeorge\GC project\data files\yearly prescriber data\Biologics_2011.dta", replace

use "Q:\PGMS\mgeorge\GC project\data files\yearly prescriber data\Biologics_2012.dta"
gen methotrexate = 0
rename PATID- mtx_pri3m, lower
sort patid index_date
bysort patid: gen patient_number = _n
drop if patient_number != 1
drop patient_number
save "Q:\PGMS\mgeorge\GC project\data files\yearly prescriber data\Biologics_2012.dta", replace

use "Q:\PGMS\mgeorge\GC project\data files\yearly prescriber data\Biologics_2013.dta"
gen methotrexate = 0
rename PATID- mtx_pri3m, lower
sort patid index_date
bysort patid: gen patient_number = _n
drop if patient_number != 1
drop patient_number
save "Q:\PGMS\mgeorge\GC project\data files\yearly prescriber data\Biologics_2013.dta", replace

use "Q:\PGMS\mgeorge\GC project\data files\yearly prescriber data\Biologics_2014.dta"
gen methotrexate = 0
rename PATID- mtx_pri3m, lower
sort patid index_date
bysort patid: gen patient_number = _n
```

```

drop if patient_number != 1
drop patient_number
save "Q:\PGMS\mgeorge\GC project\data files\yearly prescriber data\Biologics_2014.dta", replace

use "Q:\PGMS\mgeorge\GC project\data files\yearly prescriber data\Biologics_2015.dta"
gen methotrexate = 0
rename PATID- mtx_pri3m, lower
sort patid index_date
bysort patid: gen patient_number = _n
drop if patient_number != 1
drop patient_number
save "Q:\PGMS\mgeorge\GC project\data files\yearly prescriber data\Biologics_2015.dta", replace
**

use "Q:\PGMS\mgeorge\GC project\data files\yearly prescriber data\MTX_2007.dta"
gen methotrexate = 1
rename PATID- mtx_pri3m, lower
save "Q:\PGMS\mgeorge\GC project\data files\yearly prescriber data\MTX_2007.dta", replace

use "Q:\PGMS\mgeorge\GC project\data files\yearly prescriber data\MTX_2008.dta"
gen methotrexate = 1
rename PATID- mtx_pri3m, lower
save "Q:\PGMS\mgeorge\GC project\data files\yearly prescriber data\MTX_2008.dta", replace

use "Q:\PGMS\mgeorge\GC project\data files\yearly prescriber data\MTX_2009.dta"
gen methotrexate = 1
rename PATID- mtx_pri3m, lower
save "Q:\PGMS\mgeorge\GC project\data files\yearly prescriber data\MTX_2009.dta", replace

use "Q:\PGMS\mgeorge\GC project\data files\yearly prescriber data\MTX_2010.dta"
gen methotrexate = 1
rename PATID- mtx_pri3m, lower
save "Q:\PGMS\mgeorge\GC project\data files\yearly prescriber data\MTX_2010.dta", replace

use "Q:\PGMS\mgeorge\GC project\data files\yearly prescriber data\MTX_2011.dta"
gen methotrexate = 1
rename PATID- mtx_pri3m, lower
save "Q:\PGMS\mgeorge\GC project\data files\yearly prescriber data\MTX_2011.dta", replace

use "Q:\PGMS\mgeorge\GC project\data files\yearly prescriber data\MTX_2012.dta"
gen methotrexate = 1
rename PATID- mtx_pri3m, lower
save "Q:\PGMS\mgeorge\GC project\data files\yearly prescriber data\MTX_2012.dta", replace

use "Q:\PGMS\mgeorge\GC project\data files\yearly prescriber data\MTX_2013.dta"
gen methotrexate = 1
rename PATID- mtx_pri3m, lower
save "Q:\PGMS\mgeorge\GC project\data files\yearly prescriber data\MTX_2013.dta", replace

use "Q:\PGMS\mgeorge\GC project\data files\yearly prescriber data\MTX_2014.dta"
gen methotrexate = 1
rename PATID- mtx_pri3m, lower
save "Q:\PGMS\mgeorge\GC project\data files\yearly prescriber data\MTX_2014.dta", replace

use "Q:\PGMS\mgeorge\GC project\data files\yearly prescriber data\MTX_2015.dta"
gen methotrexate = 1
rename PATID- mtx_pri3m, lower
save "Q:\PGMS\mgeorge\GC project\data files\yearly prescriber data\MTX_2015.dta", replace

*** merge biologic and MTX files together: ***
use "Q:\PGMS\mgeorge\GC project\data files\yearly prescriber data\MTX_2007.dta"
merge 1:1 patid methotrexate using "Q:\PGMS\mgeorge\GC project\data files\yearly prescriber
data\Biologics_2007.dta"
. sort patid methotrexate
. bysort patid: gen patient_number = _n
. drop if patient_number == 2 // drops methotrexate patients who also receive biologics during
the year
. drop patient_number
. drop _merge
save "Q:\PGMS\mgeorge\GC project\data files\yearly prescriber data\2007_merged.dta"

use "Q:\PGMS\mgeorge\GC project\data files\yearly prescriber data\MTX_2008.dta"

```

```

merge 1:1 patid methotrexate using "Q:\PGMS\mgeorge\GC project\data files\yearly prescriber
data\Biologics_2008.dta"
. sort patid methotrexate
. bysort patid: gen patient_number = _n
. drop if patient_number == 2 // drops methotrexate patients who also receive biologics during
the year
. drop patient_number
. drop _merge
save "Q:\PGMS\mgeorge\GC project\data files\yearly prescriber data\2008_merged.dta"

use "Q:\PGMS\mgeorge\GC project\data files\yearly prescriber data\MTX_2009.dta"
merge 1:1 patid methotrexate using "Q:\PGMS\mgeorge\GC project\data files\yearly prescriber
data\Biologics_2009.dta"
. sort patid methotrexate
. bysort patid: gen patient_number = _n
. drop if patient_number == 2 // drops methotrexate patients who also receive biologics during
the year
. drop patient_number
. drop _merge
save "Q:\PGMS\mgeorge\GC project\data files\yearly prescriber data\2009_merged.dta"

use "Q:\PGMS\mgeorge\GC project\data files\yearly prescriber data\MTX_2010.dta"
merge 1:1 patid methotrexate using "Q:\PGMS\mgeorge\GC project\data files\yearly prescriber
data\Biologics_2010.dta"
. sort patid methotrexate
. bysort patid: gen patient_number = _n
. drop if patient_number == 2 // drops methotrexate patients who also receive biologics during
the year
. drop patient_number
. drop _merge
save "Q:\PGMS\mgeorge\GC project\data files\yearly prescriber data\2010_merged.dta"

use "Q:\PGMS\mgeorge\GC project\data files\yearly prescriber data\MTX_2011.dta"
merge 1:1 patid methotrexate using "Q:\PGMS\mgeorge\GC project\data files\yearly prescriber
data\Biologics_2011.dta"
. sort patid methotrexate
. bysort patid: gen patient_number = _n
. drop if patient_number == 2 // drops methotrexate patients who also receive biologics during
the year
. drop patient_number
. drop _merge
save "Q:\PGMS\mgeorge\GC project\data files\yearly prescriber data\2011_merged.dta"

use "Q:\PGMS\mgeorge\GC project\data files\yearly prescriber data\MTX_2012.dta"
merge 1:1 patid methotrexate using "Q:\PGMS\mgeorge\GC project\data files\yearly prescriber
data\Biologics_2012.dta"
. sort patid methotrexate
. bysort patid: gen patient_number = _n
. drop if patient_number == 2 // drops methotrexate patients who also receive biologics during
the year
. drop patient_number
. drop _merge
save "Q:\PGMS\mgeorge\GC project\data files\yearly prescriber data\2012_merged.dta"

use "Q:\PGMS\mgeorge\GC project\data files\yearly prescriber data\MTX_2013.dta"
merge 1:1 patid methotrexate using "Q:\PGMS\mgeorge\GC project\data files\yearly prescriber
data\Biologics_2013.dta"
. sort patid methotrexate
. bysort patid: gen patient_number = _n
. drop if patient_number == 2 // drops methotrexate patients who also receive biologics during
the year
. drop patient_number
. drop _merge
save "Q:\PGMS\mgeorge\GC project\data files\yearly prescriber data\2013_merged.dta"

use "Q:\PGMS\mgeorge\GC project\data files\yearly prescriber data\MTX_2014.dta"
merge 1:1 patid methotrexate using "Q:\PGMS\mgeorge\GC project\data files\yearly prescriber
data\Biologics_2014.dta"
. sort patid methotrexate
. bysort patid: gen patient_number = _n

```



```

. drop if patient_number == 2 // drops methotrexate patients who also receive biologics during
the year
. drop patient_number
. drop _merge
save "Q:\PGMS\mgeorge\GC project\data files\yearly prescriber data\2014_merged.dta"

use "Q:\PGMS\mgeorge\GC project\data files\yearly prescriber data\MTX_2015.dta"
merge 1:1 patid methotrexate using "Q:\PGMS\mgeorge\GC project\data files\yearly prescriber
data\Biologics_2015.dta"
. sort patid methotrexate
. bysort patid: gen patient_number = _n
. drop if patient_number == 2 // drops methotrexate patients who also receive biologics during
the year
. drop patient_number
. drop _merge
save "Q:\PGMS\mgeorge\GC project\data files\yearly prescriber data\2015_merged.dta"

*****
*** append the datasets together
use "Q:\PGMS\mgeorge\GC project\data files\yearly prescriber data\2007_merged.dta"
append using "Q:\PGMS\mgeorge\GC project\data files\yearly prescriber data\2008_merged.dta"
append using "Q:\PGMS\mgeorge\GC project\data files\yearly prescriber data\2009_merged.dta"
append using "Q:\PGMS\mgeorge\GC project\data files\yearly prescriber data\2010_merged.dta"
append using "Q:\PGMS\mgeorge\GC project\data files\yearly prescriber data\2011_merged.dta"
append using "Q:\PGMS\mgeorge\GC project\data files\yearly prescriber data\2012_merged.dta"
append using "Q:\PGMS\mgeorge\GC project\data files\yearly prescriber data\2013_merged.dta"
append using "Q:\PGMS\mgeorge\GC project\data files\yearly prescriber data\2014_merged.dta"
append using "Q:\PGMS\mgeorge\GC project\data files\yearly prescriber data\2015_merged.dta"

gen start_date = mtx_date if methotrexate == 1
replace start_date = biologics_date if methotrexate == 0

gen year = year(start_date)
egen provider_year = concat(providerid year) if providerid != ""

drop if death_date - start_date < 180 // need 180 days to assess outcome
drop if start_date >20273 // need 180 days in 2015 to assess
outcome
drop if providerid == ""

replace gc_90_30d = 0 if gc_90_30d != 1
replace gc_180_30d = 0 if gc_180_30d != 1

gen female = 1 if sex == "F"
replace female = 0 if sex == "M"

gen racecat2 = 1 if race == "1"
replace racecat2 = 2 if race == "2"
replace racecat2 = 3 if race == "5"
replace racecat2 = 4 if race == "4" | race == "3" | race == "6" | race == "0" | race == "X" |
race == ""
label define racecatname2 1"White" 2"Black" 3"Hispanic" 4 "Other/Unknown"
label values racecat2 racecatname2

gen age = (index_date - birth_date)/365.25
gen agecat2 = 1 if age <40
replace agecat2 =2 if age >=40 & age < 55
replace agecat2 = 3 if age >=55 & age < 65
replace agecat2 = 4 if age >=65 & age < 70
replace agecat2 = 5 if age >=70 & age < 75
replace agecat2 = 6 if age >=75 & age < 80
replace agecat2 = 7 if age >=80 & age != .
label define agecatname2 1 "<40" 2"40-55" 3 "55-65" 4 "65-70" 5 "70-75" 6 "75-80" 7 ">=80"
label values agecat2 agecatname2

gen drug2 = 1 if methotrexate == 1
replace drug2 = 2 if drug == "abatacept_iv" | drug == "abatacept_sq"
replace drug2 = 3 if drug == "adalimumab_sq"
replace drug2 = 4 if drug == "anakinra"
replace drug2 = 5 if drug == "certolizumab_iv" | drug == "certolizumab_sq"
replace drug2 = 6 if drug == "etanercept_sq"

```

```

replace drug2 = 7 if drug == "golimumab_iv" | drug == "golimumab_sq"
replace drug2 = 8 if drug == "infliximab"
replace drug2 = 9 if drug == "rituximab"
replace drug2 = 10 if drug == "tocilizumab_iv" | drug == "tocilizumab_sq"
replace drug2 = 11 if drug == "tofacitinib"

label define drug2name 1 "mtx" 2 "abatacept" 3 "adalimumab" 4 "anakinra" 5 "certolizumab" 6
"etanercept" 7 "golimumab" 8 "infliximab" 9 "rituximab" 10 "tocilizumab" 11 "tofacitinib"
label values drug2 drug2name

gen prior_biologics_cat = 1 if biologics_prior_cnt == .
replace prior_biologics_cat = 2 if biologics_prior_cnt == 1
replace prior_biologics_cat = 3 if biologics_prior_cnt == 2
replace prior_biologics_cat = 4 if biologics_prior_cnt >=3 & biologics_prior_cnt != .
label define priorbiologicscatname 1 "none" 2 "1" 3 "2" 4 ">=3"
label values prior_biologics_cat priorbiologicscatname

bysort provider_year: gen patient_num = _n
                        // number patients for each provider
bysort provider_year: egen number_patients = max(patient_num)
                        // how many patients in the dataset did each provider see in the past year

tab number_patients if patient_num == 1 // see
how many patients each provider sees
tab number_patients if patient_num == 1 & number_patients >=10 // number of total providers
seeing at least 10 patients

gen include_provider = 0
replace include_provider = 1 if number_patients >=10 & number_patients != . // most of the
patients cut out are from missing provider. 25-30% of providers eliminated, but represents a much
smaller number of patients

tab include_provider // proportion of patients included
tab include_provider if patient_num == 1 // proportion of physicians included

sum number_patients if patient_num == 1 & include_provider == 1, detail

bysort provider_year: egen gc_90_prop = mean(gc_90) if include_provider == 1 //
proportion of patients receiving gc within 90 days
bysort provider_year: egen gc_180_prop = mean(gc_180) if include_provider == 1 //
proportion of patients receiving gc within 90 days
bysort provider_year: egen gc_90_30d_prop = mean(gc_90_30d) if include_provider == 1
// proportion of patients receiving 30 days of gc within 90 days
bysort provider_year: egen gc_180_30d_prop = mean(gc_180_30d) if include_provider == 1
// proportion of patients receiving 30 days of gc within 90 days

** summarizing information by provider
sum gc_90_prop if patient_num == 1, detail
sum gc_180_prop if patient_num == 1, detail
sum gc_90_30d_prop if patient_num == 1, detail
sum gc_180_30d_prop if patient_num == 1, detail

** now make a model with all patients included for predicted gc
logistic gc_90 i.agecat2 i.racecat2 female i.drug2 biologics_pri3m mtx_pri3m
i.methotrexate#i.biologics_pri3m i.methotrexate#i.mtx_pri3m hcq_lef_ssa_pri3m
i.prior_biologics_cat charlson_score mi chf pvd cvd cpd dem ulc mld diab diab_sev renal tumor
liver meta i.year // include interaction term between MTX and previous MTX and biologic since
these mean different things in MTX vs biologic patients
predict predict_gc90, pr
sum predict_gc90, detail
bysort provider_year: egen gc_90_observed = sum(gc_90) if include_provider == 1
// number of patients who received gc_90 in this time
bysort provider_year: egen gc_90_expected = sum(predict_gc90) if include_provider == 1
// sum of the expected values
gen obsexp_gc90 = gc_90_observed/gc_90_expected if include_provider == 1
sum obsexp_gc90 if patient_num == 1, detail

logistic gc_180 i.agecat2 i.racecat2 female i.drug2 biologics_pri3m mtx_pri3m
i.methotrexate#i.biologics_pri3m i.methotrexate#i.mtx_pri3m hcq_lef_ssa_pri3m
i.prior_biologics_cat charlson_score mi chf pvd cvd cpd dem ulc mld diab diab_sev renal tumor
liver meta i.year

```

```

predict predict_gc180, pr
sum predict_gc180, detail
bysort provider_year: egen gc_180_observed = sum(gc_180) if include_provider == 1
bysort provider_year: egen gc_180_expected = sum(predict_gc180) if include_provider == 1
gen obsexp_gc180 = gc_180_observed/gc_180_expected if include_provider == 1
sum obsexp_gc180 if patient_num == 1, detail

logistic gc_90_30d i.agecat2 i.racecat2 female i.drug2 biologics_pri3m mtx_pri3m
i.methotrexate#i.biologics_pri3m i.methotrexate#i.mtx_pri3m hcq_lef_ssa_pri3m
i.prior_biologics_cat charlson_score i.mi i.chf i.pvd i.cvd i.cpd i.dem i.ulc i.mld i.diab
i.diab_sev i.renal i.tumor i.liver i.meta i.year
predict predict_gc90_30d, pr
sum predict_gc90_30d, detail
bysort provider_year: egen gc_90_30d_observed = sum(gc_90_30d) if include_provider == 1
// number of patients who received gc_90_30d in this time
bysort provider_year: egen gc_90_30d_expected = sum(predict_gc90_30d) if include_provider == 1
// sum of the expected values
gen obsexp_gc90_30d = gc_90_30d_observed/gc_90_30d_expected if include_provider == 1
sum obsexp_gc90_30d if patient_num == 1, detail
tway scatter obsexp_gc90_30d number_patients2 if patient_num == 1 ///
**** other key figure

logistic gc_180_30d i.agecat2 i.racecat2 female i.drug2 biologics_pri3m mtx_pri3m
i.methotrexate#i.biologics_pri3m i.methotrexate#i.mtx_pri3m hcq_lef_ssa_pri3m
i.prior_biologics_cat charlson_score mi chf pvd cvd cpd dem ulc mld diab diab_sev renal tumor
liver meta i.year
predict predict_gc180_30d, pr
sum predict_gc180_30d, detail
bysort provider_year: egen gc_180_30d_observed = sum(gc_180_30d) if include_provider == 1
bysort provider_year: egen gc_180_30d_expected = sum(predict_gc180_30d) if include_provider == 1

gen obsexp_gc180_30d = gc_180_30d_observed/gc_180_30d_expected if include_provider == 1
sum obsexp_gc180_30d if patient_num == 1, detail

*** we can look at whether distribution obsexp is different among providers with high expected
vs. providers with low expected: it is not
gen expected_over_n = gc_90_30d_expected/number_patients
sum expected_over_n if patient_num == 1 & include_provider == 1
histogram obsexp_gc90_30d if expected_over_n < 0.25 & patient_num == 1 & include_provider == 1
histogram obsexp_gc90_30d if expected_over_n > 0.25 & patient_num == 1 & include_provider == 1

*****
save "Q:\PGMS\mgeorge\GC project\data files\yearly prescriber data\2007to2015
provider_preference_11_5_19.dta"

** first use this file that includes measures for each provider_year:
save "Q:\PGMS\mgeorge\GC project\data files\yearly prescriber data\2007to2015
provider_preference_nonedited_11_5_19.dta"
drop if patient_num !=1
drop if include_provider != 1
keep provider_year providerid year number_patients gc_90_prop gc_180_prop gc_90_30d_prop
gc_180_30d_prop obsexp_gc90 obsexp_gc180 obsexp_gc90_30d obsexp_gc180_30d
save "Q:\PGMS\mgeorge\GC project\data files\yearly prescriber data\2007to2015
provider_preference_nonedited_11_5_19.dta", replace

** now create another file that removes an individual patient's contribution, this can be merged
in based on both provider_year and patid
use "Q:\PGMS\mgeorge\GC project\data files\yearly prescriber data\2007to2015
provider_preference_11_5_19.dta"
save "Q:\PGMS\mgeorge\GC project\data files\yearly prescriber data\2007to2015
provider_preference_patient_specific_11_5_19.dta"
gen gc_90_prop2 = ((gc_90_prop * number_patients) - gc_90)/(number_patients - 1)
// subtract out data from each individual patient
gen gc_180_prop2 = ((gc_180_prop * number_patients) - gc_180)/(number_patients - 1)
// subtract out data from each individual patient
gen gc_90_30d_prop2 = ((gc_90_30d_prop * number_patients) - gc_90_30d)/(number_patients - 1)
// subtract out data from each individual patient
gen gc_180_30d_prop2 = ((gc_180_30d_prop * number_patients) - gc_180_30d)/(number_patients - 1)
// subtract out data from each individual patient

gen gc_90_observed2 = gc_90_observed - gc_90

```

```

gen gc_90_expected2 = gc_90_expected - predict_gc90
gen obsexp_gc902 = gc_90_observed2/gc_90_expected2

gen gc_180_observed2 = gc_180_observed - gc_180
gen gc_180_expected2 = gc_180_expected - predict_gc180
gen obsexp_gc1802 = gc_180_observed2/gc_180_expected2

gen gc_90_30d_observed2 = gc_90_30d_observed - gc_90_30d
gen gc_90_30d_expected2 = gc_90_30d_expected - predict_gc90_30d
gen obsexp_gc90_30d2 = gc_90_30d_observed2/gc_90_30d_expected2

gen gc_180_30d_observed2 = gc_180_30d_observed - gc_180_30d
gen gc_180_30d_expected2 = gc_180_30d_expected - predict_gc180_30d
gen obsexp_gc180_30d2 = gc_180_30d_observed2/gc_180_30d_expected2

drop if include_provider != 1
keep provider_year patid gc_90_prop2 gc_180_prop2 gc_90_30d_prop2 gc_180_30d_prop2 obsexp_gc902
obsexp_gc1802 obsexp_gc90_30d2 obsexp_gc180_30d2
save "Q:\PGMS\mgeorge\GC project\data files\yearly prescriber data\2007to2015
provider_preference_patient_specific_11_5_19.dta", replace

*****
*** now merge these back into the cohort of interest:
merge m:1 provider_year using "Q:\PGMS\mgeorge\GC project\data files\yearly prescriber
data\2007to2015 provider_preference_nonedited_11_5_19.dta" // this merges by provider_year
drop if _merge == 2
drop _merge
drop providerid // this is now a duplicate variable - the one to keep is
providerid_baseline

merge m:1 patid provider_year using "Q:\PGMS\mgeorge\GC project\data files\yearly prescriber
data\2007to2015 provider_preference_patient_specific_11_5_19.dta"
drop if _merge == 2
drop _merge
codebook gc_90_prop2 if provider_year != ""

** replace values with the patient-specific values when those are present so that patients aren't
contributing to their measurement (as expected this results in lower predictive ability)
replace gc_90_prop = gc_90_prop2 if gc_90_prop2 != .
replace gc_180_prop = gc_180_prop2 if gc_180_prop2 != .
replace gc_90_30d_prop = gc_90_30d_prop2 if gc_90_30d_prop2 != .
replace gc_180_30d_prop = gc_180_30d_prop2 if gc_180_30d_prop2 != .

replace obsexp_gc90 = obsexp_gc902 if obsexp_gc902 != .
replace obsexp_gc180 = obsexp_gc1802 if obsexp_gc1802 != .
replace obsexp_gc90_30d = obsexp_gc90_30d2 if obsexp_gc90_30d2 != .
replace obsexp_gc180_30d = obsexp_gc180_30d2 if obsexp_gc180_30d2 != .
drop gc_90_prop2 gc_180_prop2 gc_90_30d_prop2 gc_180_30d_prop2 obsexp_gc902 obsexp_gc1802
obsexp_gc90_30d2 obsexp_gc180_30d2 // drop the version 2 to avoid confusion, these now are
replaced by patient-specific values if patients were in the derivation dataset in the year of
interest

```

```

***** Apply Exclusions *****
***** follow_up time for rituximab *****
drop if drug2 == 8 & (prsc_end_dt - index_dt) < 30 // do not include rituximab
patients that just received 2 initial infusions (typically 14 days apart but allow as much as 30
days apart)
***** exclusions *****
gen elig_lyr = 0
replace elig_lyr = 1 if index_dt - enr_start_date >=360 & index_dt - enr_start_date != . // this
was part of the dataset creation

*** so start with patients with >= 6months of a medication course with at least 6 months data
prior to this in the dataset (for 1 year baseline)
*** index date 2007-August31 2015

gen exclude_other_dz = 0 if prior_psa_exc ==0 & prior_as_exc ==0 & prior_ibd_exc ==0 &
prior_sle_exc ==0
replace exclude_other_dz =1 if prior_psa_exc ==1 | prior_as_exc ==1 | prior_ibd_exc ==1 |
prior_sle_exc ==1

gen missing_data = 0
replace missing_data = 1 if region == . | residence_urban == .

drop if exclude_other_dz == 1

drop if prior_cancer_exc == 1

drop if prior_hiv_exc == 1

drop if missing_data == 1

drop if providerid_baseline == . // drop if no identifiable provider

drop if obsexp_gc90_30d == . // drop if didn't have provider seeing at least 10
other RA patients

gen steroids_baseline = 1 if gc_dose_0_cat > 1
replace steroids_baseline = 0 if gc_dose_0_cat == 1

drop if gc_dose_0_cat > 2 // limit to no glucocorticoids or <=5mg/day

```

***** IV Analysis *****

** alternative drug categories:

```
gen drug_cat = 1 if drug2 == 1
replace drug_cat = 2 if drug2 == 3 | drug2 == 4 | drug2 == 5 | drug2 == 6 | drug2 == 7
replace drug_cat = 3 if drug2 == 2
replace drug_cat = 4 if drug2 == 8
replace drug_cat = 5 if drug2 == 9
replace drug_cat = 6 if drug2 == 10
tab drug_cat
label define drug_catname 1"MTX" 2"TNF" 3"Abatacept" 4"Rituximab" 5 "Tocilizumab" 6"Tofacitinib"
label values drug_cat drug_catname
```

** alternative prior biologics

```
gen prior_biologics_cat2 = 1 if prior_biologics_cat == 1
replace prior_biologics_cat2 = 2 if prior_biologics_cat == 2
replace prior_biologics_cat2 = 3 if prior_biologics_cat >2
label define priorbiologicscat2name 1"None" 2"1" 3">1"
label values prior_biologics_cat2 priorbiologicscat2name
```

```
gen age2 = age*age
```

** create dummy variables

```
tab agecat3, gen(age_)
tab racecat2, gen(race)
tab income_cat, gen(income)
tab year, gen(year_)
tab drug2, gen(drug2_)
tab prior_biologics_cat, gen(prior_biologics)
tab hosp_inf_baseline_cat, gen(infection_baseline)
tab dual_elig, gen(dual)
tab region, gen(region)
tab ed_visit_cat, gen(ed_visit_cat)
tab hosp_cat, gen(hosp_cat)
tab prior_biologics_cat2, gen(prior_biologics_cat2_)
tab drug_cat, gen(drug_cat)
```

*** dichotomize the IV definitions in the dataset of interest:

```
xtile gc_90_30d_bi = gc_90_30d_prop, nq(2)
bysort gc_90_30d_bi: sum gc_90_30d_prop
tab gc_90_30d_bi steroids_baseline, row
```

```
gen gc_90_30d_bi_high = 1 if gc_90_30d_bi ==2
replace gc_90_30d_bi_high = 0 if gc_90_30d_bi == 1
```

```
xtile obsexp_gc90_30d_bi = obsexp_gc90_30d, nq(2)
bysort obsexp_gc90_30d_bi: sum obsexp_gc90_30d
tab obsexp_gc90_30d_bi steroids_baseline, row
```

```
gen obsexp_gc90_30d_bi_high = 1 if obsexp_gc90_30d_bi ==2
replace obsexp_gc90_30d_bi_high = 0 if obsexp_gc90_30d_bi == 1
```

** Evaluate complier rate

```
tab gc_90_30d_bi_high steroids_baseline
tab obsexp_gc90_30d_bi_high steroids_baseline
```

** Complier rate with adjustment

** all covariates

```
logistic steroids_baseline i.gc_90_30d_bi_high i.agecat2 i.racecat2 female i.year i.drug2
DME_Walker_1y DME_Oxygen_1y DME_Lift_1y DME_Wheelchair_1y DME_HospitalBed_1y DME_1y
SNFstay_12mos_prior charlson_score cov_dm cov_htn cov_ckd cov_copd cov_asthma cov_cvd cov_obese
cov_chf cov_cad cov_pud cov_severe_ra cov_anemia cov_esrd cov_mi cov_depression cov_chronic_pain
i.disability i.dual_elig i.income_cat i.hosp_inf_baseline_cat i.prior_biologics_cat ssa_0 lef_0
hcq_0 mtx_0 nsaid_0 opioid_0 ppi_prior_90 antibiotics_0 outpatient_visits outpatient_rheum
i.ed_visit_cat i.hosp_cat i.region residence_urban cancer_scr_breast_prior cancer_scr_colon_prior
cancer_scr_prostate_prior cancer_scr_cervical_prior cancer_scr_lung_prior cov_flu_vaccine
cov_pneumo_vaccine cov_zoster_vaccine
margins gc_90_30d_bi_high, atmeans
```

```
logistic steroids_baseline i.obsexp_gc90_30d_bi_high i.agecat2 i.racecat2 female i.year i.drug2
DME_Walker_1y DME_Oxygen_1y DME_Lift_1y DME_Wheelchair_1y DME_HospitalBed_1y DME_1y
SNFstay_12mos_prior charlson_score cov_dm cov_htn cov_ckd cov_copd cov_asthma cov_cvd cov_obese
cov_chf cov_cad cov_pud cov_severe_ra cov_anemia cov_esrd cov_mi cov_depression cov_chronic_pain
i.disability i.dual_elig i.income_cat i.hosp_inf_baseline_cat i.prior_biologics_cat ssa_0 lef_0
hcq_0 mtx_0 nsaid_0 opioid_0 ppi_prior_90 antibiotics_0 outpatient_visits outpatient_rheum
i.ed_visit_cat i.hosp_cat i.region residence_urban cancer_scr_breast_prior cancer_scr_colon_prior
cancer_scr_prostate_prior cancer_scr_cervical_prior cancer_scr_lung_prior cov_flu_vaccine
cov_pneumo_vaccine cov_zoster_vaccine
margins obsexp_gc90_30d_bi_high, atmeans
```

** covariates used in our final IV analyses

```
logistic steroids_baseline i.obsexp_gc90_30d_bi_high age age2 i.racecat2 female i.drug_cat
i.prior_biologics_cat2 ssa_0 charlson_score cov_chronic_pain outpatient_visits outpatient_rheum
i.ed_visit_cat i.hosp_inf_baseline_cat i.region i.income_cat residence_urban cov_asthma mtx_0
i.dual_elig
margins obsexp_gc90_30d_bi_high, atmeans
```

*** partial F adjusting for all other covariates

```
anova steroids_baseline i.agecat2 i.racecat2 female i.year i.drug2 DME_Walker_1y DME_Oxygen_1y
DME_Lift_1y DME_Wheelchair_1y DME_HospitalBed_1y DME_1y SNFstay_12mos_prior charlson_score cov_dm
cov_htn cov_ckd cov_copd cov_asthma cov_cvd cov_obese cov_chf cov_cad cov_pud cov_severe_ra
cov_anemia cov_esrd cov_mi cov_depression cov_chronic_pain i.disability i.dual_elig i.income_cat
i.hosp_inf_baseline_cat i.prior_biologics_cat ssa_0 lef_0 hcq_0 mtx_0 nsaid_0 opioid_0
ppi_prior_90 antibiotics_0 outpatient_visits outpatient_rheum i.ed_visit_cat i.hosp_cat i.region
residence_urban cancer_scr_breast_prior cancer_scr_colon_prior cancer_scr_prostate_prior
cancer_scr_cervical_prior cancer_scr_lung_prior cov_flu_vaccine cov_pneumo_vaccine
cov_zoster_vaccine i.gc_90_30d_bi_high
```

```
anova steroids_baseline i.agecat2 i.racecat2 female i.year i.drug2 DME_Walker_1y DME_Oxygen_1y
DME_Lift_1y DME_Wheelchair_1y DME_HospitalBed_1y DME_1y SNFstay_12mos_prior charlson_score cov_dm
cov_htn cov_ckd cov_copd cov_asthma cov_cvd cov_obese cov_chf cov_cad cov_pud cov_severe_ra
cov_anemia cov_esrd cov_mi cov_depression cov_chronic_pain i.disability i.dual_elig i.income_cat
i.hosp_inf_baseline_cat i.prior_biologics_cat ssa_0 lef_0 hcq_0 mtx_0 nsaid_0 opioid_0
ppi_prior_90 antibiotics_0 outpatient_visits outpatient_rheum i.ed_visit_cat i.hosp_cat i.region
residence_urban cancer_scr_breast_prior cancer_scr_colon_prior cancer_scr_prostate_prior
cancer_scr_cervical_prior cancer_scr_lung_prior cov_flu_vaccine cov_pneumo_vaccine
cov_zoster_vaccine i.obsexp_gc90_30d_bi_high
```

*** logistic regression models with all covariates:

```
logistic steroids_baseline i.agecat2 i.racecat2 female i.year i.drug2 DME_Walker_1y DME_Oxygen_1y
DME_Lift_1y DME_Wheelchair_1y DME_HospitalBed_1y DME_1y SNFstay_12mos_prior charlson_score cov_dm
cov_htn cov_ckd cov_copd cov_asthma cov_cvd cov_obese cov_chf cov_cad cov_pud cov_severe_ra
cov_anemia cov_esrd cov_mi cov_depression cov_chronic_pain i.disability i.dual_elig i.income_cat
i.hosp_inf_baseline_cat i.prior_biologics_cat ssa_0 lef_0 hcq_0 mtx_0 nsaid_0 opioid_0
ppi_prior_90 antibiotics_0 outpatient_visits outpatient_rheum i.ed_visit_cat i.hosp_cat i.region
residence_urban cancer_scr_breast_prior cancer_scr_colon_prior cancer_scr_prostate_prior
cancer_scr_cervical_prior cancer_scr_lung_prior cov_flu_vaccine cov_pneumo_vaccine
cov_zoster_vaccine
estat ic
```

```
logistic steroids_baseline i.agecat2 i.racecat2 female i.year i.drug2 DME_Walker_1y DME_Oxygen_1y
DME_Lift_1y DME_Wheelchair_1y DME_HospitalBed_1y DME_1y SNFstay_12mos_prior charlson_score cov_dm
cov_htn cov_ckd cov_copd cov_asthma cov_cvd cov_obese cov_chf cov_cad cov_pud cov_severe_ra
cov_anemia cov_esrd cov_mi cov_depression cov_chronic_pain i.disability i.dual_elig i.income_cat
i.hosp_inf_baseline_cat i.prior_biologics_cat ssa_0 lef_0 hcq_0 mtx_0 nsaid_0 opioid_0
ppi_prior_90 antibiotics_0 outpatient_visits outpatient_rheum i.ed_visit_cat i.hosp_cat i.region
residence_urban cancer_scr_breast_prior cancer_scr_colon_prior cancer_scr_prostate_prior
cancer_scr_cervical_prior cancer_scr_lung_prior cov_flu_vaccine cov_pneumo_vaccine
cov_zoster_vaccine i.gc_90_30d_bi_high
estat ic
```

```
logistic steroids_baseline i.agecat2 i.racecat2 female i.year i.drug2 DME_Walker_1y DME_Oxygen_1y
DME_Lift_1y DME_Wheelchair_1y DME_HospitalBed_1y DME_1y SNFstay_12mos_prior charlson_score cov_dm
cov_htn cov_ckd cov_copd cov_asthma cov_cvd cov_obese cov_chf cov_cad cov_pud cov_severe_ra
cov_anemia cov_esrd cov_mi cov_depression cov_chronic_pain i.disability i.dual_elig i.income_cat
i.hosp_inf_baseline_cat i.prior_biologics_cat ssa_0 lef_0 hcq_0 mtx_0 nsaid_0 opioid_0
ppi_prior_90 antibiotics_0 outpatient_visits outpatient_rheum i.ed_visit_cat i.hosp_cat i.region
residence_urban cancer_scr_breast_prior cancer_scr_colon_prior cancer_scr_prostate_prior
cancer_scr_cervical_prior cancer_scr_lung_prior cov_flu_vaccine cov_pneumo_vaccine
cov_zoster_vaccine
estat ic
```

```
logistic steroids_baseline i.agecat2 i.racecat2 female i.year i.drug2 DME_Walker_ly DME_Oxygen_ly
DME_Lift_ly DME_Wheelchair_ly DME_HospitalBed_ly DME_ly SNFstay_12mos_prior charlson_score cov_dm
cov_htn cov_ckd cov_copd cov_asthma cov_cvd cov_obese cov_chf cov_cad cov_pud cov_severe_ra
cov_anemia cov_esrd cov_mi cov_depression cov_chronic_pain i.disability i.dual_elig i.income_cat
i.hosp_inf_baseline_cat i.prior_biologics_cat ssa_0 lef_0 hcq_0 mtx_0 nsaid_0 opioid_0
ppi_prior_90 antibiotics_0 outpatient_visits outpatient_rheum i.ed_visit_cat i.hosp_cat i.region
residence_urban cancer_scr_breast_prior cancer_scr_colon_prior cancer_scr_prostate_prior
cancer_scr_cervical_prior cancer_scr_lung_prior cov_flu_vaccine cov_pneumo_vaccine
cov_zoster_vaccine i.obsexp_gc90_30d_bi_hi
estat ic
```

**** Check balance in patient characteristics across groups - first across values of the IV, then across steroids vs. no steroids in the same population:

```
pbalchk gc_90_30d_bi_high age age_1 age_2 age_3 age_4 age_5 age_6 age_7 age_8 age_9 age_10 age_11
race1 race2 race3 race4 female disability dual1 dual2 dual3 SNFstay_12mos_prior income1 income2
income3 income4 income5 income6 year_1 year_2 year_3 year_4 year_5 year_6 year_7 year_8 year_9
region1 region2 region3 region4 region5 region6 region7 region8 region9 residence_urban drug2_1
drug2_2 drug2_3 drug2_4 drug2_5 drug2_6 drug2_7 drug2_8 drug2_9 drug2_10 prior_biologics1
prior_biologics2 prior_biologics3 prior_biologics4 ssa_0 lef_0 hcq_0 mtx_0 nsaid_0 opioid_0
ppi_prior_90 antibiotics_0 charlson_score cov_dm cov_htn cov_ckd cov_copd cov_asthma cov_cvd
cov_obese cov_chf cov_cad cov_pud cov_severe_ra cov_anemia cov_esrd cov_mi cov_depression
cov_chronic_pain outpatient_visits outpatient_rheum ed_visit_cat1 ed_visit_cat2 ed_visit_cat3
ed_visit_cat4 hosp_cat1 hosp_cat2 hosp_cat3 hosp_cat4 infection_baselinel infection_baseline2
infection_baseline3 infection_baseline4 DME_Walker_ly DME_Oxygen_ly DME_Lift_ly DME_Wheelchair_ly
DME_HospitalBed_ly DME_ly cancer_scr_breast_prior cancer_scr_colon_prior
cancer_scr_prostate_prior cancer_scr_cervical_prior cancer_scr_lung_prior cov_flu_vaccine
cov_pneumo_vaccine cov_zoster_vaccine if gc_90_30d_bi_high != .
```

```
pbalchk steroids_baseline age age_1 age_2 age_3 age_4 age_5 age_6 age_7 age_8 age_9 age_10 age_11
race1 race2 race3 race4 female disability dual1 dual2 dual3 SNFstay_12mos_prior income1 income2
income3 income4 income5 income6 year_1 year_2 year_3 year_4 year_5 year_6 year_7 year_8 year_9
region1 region2 region3 region4 region5 region6 region7 region8 region9 residence_urban drug2_1
drug2_2 drug2_3 drug2_4 drug2_5 drug2_6 drug2_7 drug2_8 drug2_9 drug2_10 prior_biologics1
prior_biologics2 prior_biologics3 prior_biologics4 ssa_0 lef_0 hcq_0 mtx_0 nsaid_0 opioid_0
ppi_prior_90 antibiotics_0 charlson_score cov_dm cov_htn cov_ckd cov_copd cov_asthma cov_cvd
cov_obese cov_chf cov_cad cov_pud cov_severe_ra cov_anemia cov_esrd cov_mi cov_depression
cov_chronic_pain outpatient_visits outpatient_rheum ed_visit_cat1 ed_visit_cat2 ed_visit_cat3
ed_visit_cat4 hosp_cat1 hosp_cat2 hosp_cat3 hosp_cat4 infection_baselinel infection_baseline2
infection_baseline3 infection_baseline4 DME_Walker_ly DME_Oxygen_ly DME_Lift_ly DME_Wheelchair_ly
DME_HospitalBed_ly DME_ly cancer_scr_breast_prior cancer_scr_colon_prior
cancer_scr_prostate_prior cancer_scr_cervical_prior cancer_scr_lung_prior cov_flu_vaccine
cov_pneumo_vaccine cov_zoster_vaccine if gc_90_30d_bi_high != .
```

```
pbalchk obsexp_gc90_30d_bi_high age age_1 age_2 age_3 age_4 age_5 age_6 age_7 age_8 age_9 age_10
age_11 race1 race2 race3 race4 female disability dual1 dual2 dual3 SNFstay_12mos_prior income1
income2 income3 income4 income5 income6 year_1 year_2 year_3 year_4 year_5 year_6 year_7 year_8
year_9 region1 region2 region3 region4 region5 region6 region7 region8 region9 residence_urban
drug2_1 drug2_2 drug2_3 drug2_4 drug2_5 drug2_6 drug2_7 drug2_8 drug2_9 drug2_10 prior_biologics1
prior_biologics2 prior_biologics3 prior_biologics4 ssa_0 lef_0 hcq_0 mtx_0 nsaid_0 opioid_0
ppi_prior_90 antibiotics_0 charlson_score cov_dm cov_htn cov_ckd cov_copd cov_asthma cov_cvd
cov_obese cov_chf cov_cad cov_pud cov_severe_ra cov_anemia cov_esrd cov_mi cov_depression
cov_chronic_pain outpatient_visits outpatient_rheum ed_visit_cat1 ed_visit_cat2 ed_visit_cat3
ed_visit_cat4 hosp_cat1 hosp_cat2 hosp_cat3 hosp_cat4 infection_baselinel infection_baseline2
infection_baseline3 infection_baseline4 DME_Walker_ly DME_Oxygen_ly DME_Lift_ly DME_Wheelchair_ly
DME_HospitalBed_ly DME_ly cancer_scr_breast_prior cancer_scr_colon_prior
cancer_scr_prostate_prior cancer_scr_cervical_prior cancer_scr_lung_prior cov_flu_vaccine
cov_pneumo_vaccine cov_zoster_vaccine if obsexp_gc90_30d_bi_high != .
```

***** Table 1 *****

*** Cohort description ***

*follow-up time:

bysort obsexp_gc90_30d_bi_high: sum time2, detail

bysort steroids_baseline: sum time2, detail

** across IV:

sum age

bysort obsexp_gc90_30d_bi_high: sum age

tab agecat3 obsexp_gc90_30d_bi_high

tab female obsexp_gc90_30d_bi_high


```

tab racecat2 obsexp_gc90_30d_bi_high
tab year2011_2015 obsexp_gc90_30d_bi_high
tab year obsexp_gc90_30d_bi_high

tab dual_elig obsexp_gc90_30d_bi_high
tab income_cat obsexp_gc90_30d_bi_high

tab region obsexp_gc90_30d_bi_high
tab residence_urban obsexp_gc90_30d_bi_high

tab drug2 obsexp_gc90_30d_bi_high
tab drug_cat obsexp_gc90_30d_bi_high
tab prior_biologics_cat obsexp_gc90_30d_bi_high
tab prior_biologics_cat2 obsexp_gc90_30d_bi_high
tab ssa_0 obsexp_gc90_30d_bi_high
tab lef_0 obsexp_gc90_30d_bi_high
tab hcq_0 obsexp_gc90_30d_bi_high
tab mtx_0 obsexp_gc90_30d_bi_high
tab nsaid_0 obsexp_gc90_30d_bi_high
tab opioid_0 obsexp_gc90_30d_bi_high
tab ppi_prior_90 obsexp_gc90_30d_bi_high

sum charlson_score, detail
bysort obsexp_gc90_30d_bi_high: sum charlson_score, detail
tab cov_dm obsexp_gc90_30d_bi_high
tab cov_htn obsexp_gc90_30d_bi_high
tab cov_ckd obsexp_gc90_30d_bi_high
tab cov_copd obsexp_gc90_30d_bi_high
tab cov_asthma obsexp_gc90_30d_bi_high
tab cov_cvd obsexp_gc90_30d_bi_high
tab cov_obese obsexp_gc90_30d_bi_high
tab cov_chf obsexp_gc90_30d_bi_high
tab cov_cad obsexp_gc90_30d_bi_high
tab cov_pud obsexp_gc90_30d_bi_high
tab cov_severe_ra obsexp_gc90_30d_bi_high
tab cov_anemia obsexp_gc90_30d_bi_high
tab cov_esrd obsexp_gc90_30d_bi_high
tab cov_mi obsexp_gc90_30d_bi_high
tab cov_depression obsexp_gc90_30d_bi_high
tab cov_chronic_pain obsexp_gc90_30d_bi_high

sum outpatient_visits, detail
bysort obsexp_gc90_30d_bi_high: sum outpatient_visits, detail
sum outpatient_rheum, detail
bysort obsexp_gc90_30d_bi_high: sum outpatient_rheum, detail
tab ed_visit_cat obsexp_gc90_30d_bi_high
tab hosp_cat obsexp_gc90_30d_bi_high
tab hosp_inf_baseline_cat obsexp_gc90_30d_bi_high
tab antibiotics_0 obsexp_gc90_30d_bi_high

tab disability obsexp_gc90_30d_bi_high
tab SNFstay_12mos_prior obsexp_gc90_30d_bi_high
tab DME_Walker_1y obsexp_gc90_30d_bi_high
tab DME_Oxygen_1y obsexp_gc90_30d_bi_high
tab DME_Lift_1y obsexp_gc90_30d_bi_high
tab DME_Wheelchair_1y obsexp_gc90_30d_bi_high
tab DME_HospitalBed_1y obsexp_gc90_30d_bi_high
tab DME_1y obsexp_gc90_30d_bi_high

tab cancer_scr_breast_prior obsexp_gc90_30d_bi_high
tab cancer_scr_colon_prior obsexp_gc90_30d_bi_high
tab cancer_scr_prostate_prior obsexp_gc90_30d_bi_high
tab cancer_scr_cervical_prior obsexp_gc90_30d_bi_high
tab cancer_scr_lung_prior obsexp_gc90_30d_bi_high
tab cov_flu_vaccine obsexp_gc90_30d_bi_high
tab cov_pneumo_vaccine obsexp_gc90_30d_bi_high
tab cov_zoster_vaccine obsexp_gc90_30d_bi_high

** cohort characteristics across exposure:
sum age

```

```
bysort steroids_baseline: sum age
tab agecat3 steroids_baseline
tab female steroids_baseline
tab racecat2 steroids_baseline
tab year2011_2015 steroids_baseline
tab year steroids_baseline

tab dual_elig steroids_baseline
tab income_cat steroids_baseline

tab region steroids_baseline
tab residence_urban steroids_baseline

tab drug2 steroids_baseline
tab drug_cat steroids_baseline
tab prior_biologics_cat steroids_baseline
tab prior_biologics_cat2 steroids_baseline
tab ssa_0 steroids_baseline
tab lef_0 steroids_baseline
tab hcq_0 steroids_baseline
tab mtx_0 steroids_baseline
tab nsaid_0 steroids_baseline
tab opioid_0 steroids_baseline
tab ppi_prior_90 steroids_baseline

sum charlson_score, detail
bysort steroids_baseline: sum charlson_score, detail
tab cov_dm steroids_baseline
tab cov_htn steroids_baseline
tab cov_ckd steroids_baseline
tab cov_copd steroids_baseline
tab cov_asthma steroids_baseline
tab cov_cvd steroids_baseline
tab cov_obese steroids_baseline
tab cov_chf steroids_baseline
tab cov_cad steroids_baseline
tab cov_pud steroids_baseline
tab cov_severe_ra steroids_baseline
tab cov_anemia steroids_baseline
tab cov_esrd steroids_baseline
tab cov_mi steroids_baseline
tab cov_depression steroids_baseline
tab cov_chronic_pain steroids_baseline

sum outpatient_visits, detail
bysort steroids_baseline: sum outpatient_visits, detail
sum outpatient_rheum, detail
bysort steroids_baseline: sum outpatient_rheum, detail
tab ed_visit_cat steroids_baseline
tab hosp_cat steroids_baseline
tab hosp_inf_baseline_cat steroids_baseline
tab antibiotics_0 steroids_baseline

tab disability steroids_baseline
tab SNFstay_12mos_prior steroids_baseline
tab DME_Walker_1y steroids_baseline
tab DME_Oxygen_1y steroids_baseline
tab DME_Lift_1y steroids_baseline
tab DME_Wheelchair_1y steroids_baseline
tab DME_HospitalBed_1y steroids_baseline
tab DME_1y steroids_baseline

tab cancer_scr_breast_prior steroids_baseline
tab cancer_scr_colon_prior steroids_baseline
tab cancer_scr_prostate_prior steroids_baseline
tab cancer_scr_cervical_prior steroids_baseline
tab cancer_scr_lung_prior steroids_baseline
tab cov_flu_vaccine steroids_baseline
tab cov_pneumo_vaccine steroids_baseline
tab cov_zoster_vaccine steroids_baseline
```

```

**** outcome summary
stset time2, fail(fail2 ==1) scale(365)
stsum, by(steroids_baseline)
stsum, by(obsexp_gc90_30d_bi_high)
stsum, by(gc_90_30d_bi_high)
tab steroids_baseline fail2
tab obsexp_gc90_30d_bi_high fail2
tab gc_90_30d_bi_high fail2

stset time2_primary, fail(fail2_primary ==1) scale(365)
stsum, by(steroids_baseline)
stsum, by(obsexp_gc90_30d_bi_high)
stsum, by(gc_90_30d_bi_high)
tab steroids_baseline fail2_primary
tab obsexp_gc90_30d_bi_high fail2_primary
tab gc_90_30d_bi_high fail2_primary

***** Multivariable model with all covariates included to see risk factors *****
stset time2, fail(fail2 ==1) scale(365) // this treats death as censoring,
does NOT censor at steroid dose changes but does censor with DMARD changes
stcox i.gc_dose_0_cat, cluster(patid)

stcox i.gc_dose_0_cat i.agecat3 female i.racecat2 i.disability i.dual_elig i.SNFstay_12mos_prior
i.income_cat i.region residence urban i.year i.drug2 i.prior_biologics_cat ssa_0 lef_0 hcq_0
mtx_0 nsaid_0 opioid_0 ppi_prior_90 antibiotics_0 charlson_score cov_dm cov_htn cov_ckd cov_copd
cov_asthma cov_cvd cov_obese cov_chf cov_cad cov_pud cov_severe_ra cov_anemia cov_esrd cov_mi
cov_depression cov_chronic_pain outpatient_visits outpatient_rheum i.ed_visit_cat i.hosp_cat
i.hosp_inf_baseline_cat DME_Walker_ly DME_Oxygen_ly DME_Lift_ly DME_Wheelchair_ly
DME_HospitalBed_ly DME_ly cancer_scr_breast_prior cancer_scr_colon_prior
cancer_scr_prostate_prior cancer_scr_cervical_prior cancer_scr_lung_prior cov_flu_vaccine
cov_pneumo_vaccine cov_zoster_vaccine, cluster(patid)

stset time2_primary, fail(fail2_primary ==1) scale(365) // this treats death
as censoring, does NOT censor at steroid dose changes but does censor with DMARD changes
stcox i.gc_dose_0_cat, cluster(patid)

stcox i.gc_dose_0_cat i.agecat3 female i.racecat2 i.disability i.dual_elig i.SNFstay_12mos_prior
i.income_cat i.region residence urban i.year i.drug2 i.prior_biologics_cat ssa_0 lef_0 hcq_0
mtx_0 nsaid_0 opioid_0 ppi_prior_90 antibiotics_0 charlson_score cov_dm cov_htn cov_ckd cov_copd
cov_asthma cov_cvd cov_obese cov_chf cov_cad cov_pud cov_severe_ra cov_anemia cov_esrd cov_mi
cov_depression cov_chronic_pain outpatient_visits outpatient_rheum i.ed_visit_cat i.hosp_cat
i.hosp_inf_baseline_cat DME_Walker_ly DME_Oxygen_ly DME_Lift_ly DME_Wheelchair_ly
DME_HospitalBed_ly DME_ly cancer_scr_breast_prior cancer_scr_colon_prior
cancer_scr_prostate_prior cancer_scr_cervical_prior cancer_scr_lung_prior cov_flu_vaccine
cov_pneumo_vaccine cov_zoster_vaccine, cluster(patid)

***** making datasets for R *****

gen fail2_binary = 1 if fail2 == 1
replace fail2_binary = 0 if fail2 == 0 | fail2 == 2

gen fail1_binary = 1 if fail1 == 1
replace fail1_binary = 0 if fail1 == 0 | fail1 == 2 // 2 was death

gen fail2_primary_binary = 1 if fail2_primary == 1
replace fail2_primary_binary = 0 if fail2_primary ==0 | fail2_primary == 2

gen fail1_primary_binary = 1 if fail1_primary == 1
replace fail1_primary_binary = 0 if fail1_primary ==0 | fail1_primary == 2

keep patid index_dt steroids_baseline obsexp_gc90_30d_bi_high obsexp_gc90_30d_20p gc_90_30d_20p
gc_90_30d_bi_high time1 fail1_binary time2 fail2_binary fail1_primary_binary time1_primary
fail2_primary_binary time2_primary age age2 age_1 age_2 age_3 age_4 age_5 age_6 age_7 race1 race2
race3 race4 female dual1 dual2 dual3 drug_cat1 drug_cat2 drug_cat3 drug_cat4 drug_cat5 drug_cat6
prior_biologics_cat2_1 prior_biologics_cat2_2 prior_biologics_cat2_3 charlson_score
infection_baseline1 infection_baseline2 infection_baseline3 infection_baseline4 opioid_0
antibiotics_0 outpatient_visits outpatient_rheum region1 region2 region3 region4 region5 region6
region7 region8 region9 ed_visit_cat1 ed_visit_cat2 ed_visit_cat3 ed_visit_cat4 ssa_0
cov_chronic_pain residence_urban region1 region2 region3 region4 region5 region6 region7 region8
region9 income1 income2 income3 income4 income5 income6 mtx_0 cov_asthma disability

```

```
cancer_scr_breast_prior cancer_scr_colon_prior cancer_scr_lung_prior DME_1y year lef_0 hcq_0
cov_flu_vaccine cov_zoster_vaccine cov_pneumo_vaccine
```

```
save "Q:\PGMS\mgeorge\GC project\data files\Updated data 9_2019\IV analysis and variability
files\Rfile_OEmeasures_8_2020_update.dta", replace
```

```
***** R file *****
```

```
/*
```

```
# Install package -----
```

```
# This section only needs to execute once
```

```
install.packages("ivtools")
```

```
install.packages("survival")
```

```
install.packages("haven")
```

```
# Load libraries -----
```

```
library(ivtools)
```

```
library(survival)
```

```
library(haven)
```

```
# Load Stata dataset -----
```

```
# Replace DATA.dta with the full path to the Stata dataset
```

```
dat <- read_stata(file = "Q:\PGMS\mgeorge\GC project\data files\Updated data 9_2019\IV analysis
and variability files\Rfile_OEmeasures_8_2020_update.dta")
```

```
# Check dimension of the dataset
```

```
dim(dat)
```

```
# Check the first 6 rows
```

```
head(x = dat, n = 6)
```

```
# Make sure the treatment variable is X, IV is Z, survival time is T, and
```

```
# censoring indicator is d
```

```
dat$X <- dat$steroids_baseline
```

```
dat$Z <- dat$obsexp_gc90_30d_bi_high
```

```
dat$T <- dat$time2
```

```
dat$d <- dat$fail2_binary
```

```
# Two-stage estimation -----
```

```
# Make sure the treatment variable is X, IV is Z, survival time is T, and
```

```
# censoring indicator is d
```

```
fitX.LZ <- glm(formula = X ~ Z, data = dat)
```

```
fitT.LX <- coxph(formula = Surv(T, d) ~ X, data = dat)
```

```
fitIV <- ivcoxph(
```

```
  estmethod = "ts",
```

```
  fitX.LZ = fitX.LZ,
```

```
  fitT.LX = fitT.LX,
```

```
  data = dat,
```

```
  ctrl = TRUE
```

```
)
```

```
summary(fitIV)
```

```
# then with covariates:
```

```
fitX.LZ <- glm(formula = X ~ Z + age + age2 + race2 + race3 + race4 + female + drug_cat2 +
drug_cat3 + drug_cat4 + drug_cat5 + drug_cat6 + prior_biologics_cat2_2 + prior_biologics_cat2_3 +
ssa_0 + charlson_score + cov_chronic_pain + outpatient_visits + outpatient_rheum + ed_visit_cat2
+ ed_visit_cat3 + ed_visit_cat4 + infection_baseline2 + infection_baseline3 + infection_baseline4
+ region2 + region3 + region4 + region5 + region6 + region7 + region8 + region9 + income2 +
income3 + income4 + income5 + income6 + residence_urban + cov_asthma + mtx_0 + dual2 + dual3,
data = dat)
```

```
fitT.LX <- coxph(formula = Surv(T, d) ~ X + age + age2 + race2 + race3 + race4 + female +
drug_cat2 + drug_cat3 + drug_cat4 + drug_cat5 + drug_cat6 + prior_biologics_cat2_2 +
prior_biologics_cat2_3 + ssa_0 + charlson_score + cov_chronic_pain + outpatient_visits +
outpatient_rheum + ed_visit_cat2 + ed_visit_cat3 + ed_visit_cat4 + infection_baseline2 +
infection_baseline3 + infection_baseline4 + region2 + region3 + region4 + region5 + region6 +
region7 + region8 + region9 + income2 + income3 + income4 + income5 + income6 + residence_urban +
cov_asthma + mtx_0 + dual2 + dual3, data = dat)
```

```

fitIV <- ivcoxph(
  estmethod = "ts",
  fitX.LZ = fitX.LZ,
  fitT.LX = fitT.LX,
  data = dat,
  ctrl = TRUE
)
summary(fitIV)

** Primary dx of infection **
**** for primary dx

dat$X <- dat$steroids_baseline
dat$Z <- dat$obsexp_gc90_30d_bi_high
dat$T <- dat$time2_primary
dat$d <- dat$fail2_primary_binary

fitX.LZ <- glm(formula = X ~ Z, data = dat)
fitT.LX <- coxph(formula = Surv(T, d) ~ X, data = dat)

fitIV <- ivcoxph(
  estmethod = "ts",
  fitX.LZ = fitX.LZ,
  fitT.LX = fitT.LX,
  data = dat,
  ctrl = TRUE
)
summary(fitIV)

fitX.LZ <- glm(formula = X ~ Z + age + age2 + race2 + race3 + race4 + female + drug_cat2 +
drug_cat3 + drug_cat4 + drug_cat5 + drug_cat6 + prior_biologics_cat2_2 + prior_biologics_cat2_3 +
ssa_0 + charlson_score + cov_chronic_pain + outpatient_visits + outpatient_rheum + ed_visit_cat2
+ ed_visit_cat3 + ed_visit_cat4 + infection_baseline2 + infection_baseline3 + infection_baseline4
+ region2 + region3 + region4 + region5 + region6 + region7 + region8 + region9 + income2 +
income3 + income4 + income5 + income6 + residence_urban + cov_asthma + mtx_0 + dual2 + dual3,
data = dat)
fitT.LX <- coxph(formula = Surv(T, d) ~ X + age + age2 + race2 + race3 + race4 + female +
drug_cat2 + drug_cat3 + drug_cat4 + drug_cat5 + drug_cat6 + prior_biologics_cat2_2 +
prior_biologics_cat2_3 + ssa_0 + charlson_score + cov_chronic_pain + outpatient_visits +
outpatient_rheum + ed_visit_cat2 + ed_visit_cat3 + ed_visit_cat4 + infection_baseline2 +
infection_baseline3 + infection_baseline4 + region2 + region3 + region4 + region5 + region6 +
region7 + region8 + region9 + income2 + income3 + income4 + income5 + income6 + residence_urban +
cov_asthma + mtx_0 + dual2 + dual3, data = dat)

fitIV <- ivcoxph(
  estmethod = "ts",
  fitX.LZ = fitX.LZ,
  fitT.LX = fitT.LX,
  data = dat,
  ctrl = TRUE
)
summary(fitIV)

*****

*** non-obsexp IV any dx:
dat$X <- dat$steroids_baseline
dat$Z <- dat$gc_90_30d_bi_high
dat$T <- dat$time2
dat$d <- dat$fail2_binary

fitX.LZ <- glm(formula = X ~ Z, data = dat)
fitT.LX <- coxph(formula = Surv(T, d) ~ X, data = dat)

fitIV <- ivcoxph(
  estmethod = "ts",
  fitX.LZ = fitX.LZ,
  fitT.LX = fitT.LX,
  data = dat,

```

```

    ctrl = TRUE
  )
  summary(fitIV)

# then with covariates:

fitX.LZ <- glm(formula = X ~ Z + age + age2 + race2 + race3 + race4 + female + drug_cat2 +
drug_cat3 + drug_cat4 + drug_cat5 + drug_cat6 + prior_biologics_cat2_2 + prior_biologics_cat2_3 +
ssa_0 + charlson_score + cov_chronic_pain + outpatient_visits + outpatient_rheum + ed_visit_cat2
+ ed_visit_cat3 + ed_visit_cat4 + infection_baseline2 + infection_baseline3 + infection_baseline4
+ region2 + region3 + region4 + region5 + region6 + region7 + region8 + region9 + income2 +
income3 + income4 + income5 + income6 + residence_urban + mtx_0 + year + disability +
cancer_scr_breast_prior + cancer_scr_colon_prior + dual2 + dual3, data = dat)
fitT.LX <- coxph(formula = Surv(T, d) ~ X + age + age2 + race2 + race3 + race4 + female +
drug_cat2 + drug_cat3 + drug_cat4 + drug_cat5 + drug_cat6 + prior_biologics_cat2_2 +
prior_biologics_cat2_3 + ssa_0 + charlson_score + cov_chronic_pain + outpatient_visits +
outpatient_rheum + ed_visit_cat2 + ed_visit_cat3 + ed_visit_cat4 + infection_baseline2 +
infection_baseline3 + infection_baseline4 + region2 + region3 + region4 + region5 + region6 +
region7 + region8 + region9 + income2 + income3 + income4 + income5 + income6 + residence_urban +
mtx_0 + year + disability + cancer_scr_breast_prior + cancer_scr_colon_prior + dual2 + dual3,
data = dat)

fitIV <- ivcoxph(
  estmethod = "ts",
  fitX.LZ = fitX.LZ,
  fitT.LX = fitT.LX,
  data = dat,
  ctrl = TRUE
)
summary(fitIV)

**** non-obsexp IV primary dx:
dat$X <- dat$steroids_baseline
dat$Z <- dat$gc_90_30d_bi_high
dat$T <- dat$time2_primary
dat$d <- dat$fail2_primary_binary

fitX.LZ <- glm(formula = X ~ Z, data = dat)
fitT.LX <- coxph(formula = Surv(T, d) ~ X, data = dat)

fitIV <- ivcoxph(
  estmethod = "ts",
  fitX.LZ = fitX.LZ,
  fitT.LX = fitT.LX,
  data = dat,
  ctrl = TRUE
)
summary(fitIV)

# then with covariates:

fitX.LZ <- glm(formula = X ~ Z + age + age2 + race2 + race3 + race4 + female + drug_cat2 +
drug_cat3 + drug_cat4 + drug_cat5 + drug_cat6 + prior_biologics_cat2_2 + prior_biologics_cat2_3 +
ssa_0 + charlson_score + cov_chronic_pain + outpatient_visits + outpatient_rheum + ed_visit_cat2
+ ed_visit_cat3 + ed_visit_cat4 + infection_baseline2 + infection_baseline3 + infection_baseline4
+ region2 + region3 + region4 + region5 + region6 + region7 + region8 + region9 + income2 +
income3 + income4 + income5 + income6 + residence_urban + mtx_0 + year + disability +
cancer_scr_breast_prior + cancer_scr_colon_prior + dual2 + dual3, data = dat)
fitT.LX <- coxph(formula = Surv(T, d) ~ X + age + age2 + race2 + race3 + race4 + female +
drug_cat2 + drug_cat3 + drug_cat4 + drug_cat5 + drug_cat6 + prior_biologics_cat2_2 +
prior_biologics_cat2_3 + ssa_0 + charlson_score + cov_chronic_pain + outpatient_visits +
outpatient_rheum + ed_visit_cat2 + ed_visit_cat3 + ed_visit_cat4 + infection_baseline2 +
infection_baseline3 + infection_baseline4 + region2 + region3 + region4 + region5 + region6 +
region7 + region8 + region9 + income2 + income3 + income4 + income5 + income6 + residence_urban +
mtx_0 + year + disability + cancer_scr_breast_prior + cancer_scr_colon_prior + dual2 + dual3,
data = dat)

fitIV <- ivcoxph(
  estmethod = "ts",
  fitX.LZ = fitX.LZ,

```

```
fitT.LX = fitT.LX,  
data = dat,  
ctrl = TRUE  
)  
summary(fitIV)
```

```
*/  
** to convert coefficients from R:  
di exp(coef)  
di exp(coef - invnormal(0.975)*SD)  
di exp(coef + invnormal(0.975)*SD)
```