data oud;
set alloud;
if site_id=44 then site_id=43; /*Medical Center of the Rockies-crosses with PVH to PVH*/
if site_id=54 then site_id=53; /*Baystate Mary Lane to Baystate Wing*/
if site_id=32 then site_id=31; /*UAB-Highlands to UAB-Main*/
run;

/*covariate constrained randomization*/
proc format;
value system 1="Yale"
    2="UNC"
    3="UAB"
    4="Colorado"
    5="Baystate"
;

data RCT;
set RCT_temp;
if count=1 then site_id=11; /*SRC*/
if count=2 then site_id=12; /*Bridgeport*/
if count=3 then site_id=13; /*L&M*/
if count=4 then site_id=14; /*Greenwich*/
if count=5 then site_id=21;   /*UNC main*/ /*Academic*/
if count=6 then site_id=22; /*Chatham*/
if count=7 then site_id=23; /*Rex*/
if count=8 then site_id=24; /*Smithfield*/
if count=9 then site_id=25; /*UNC-Nash*/
if count=10 then site_id=31; /*UAB Main*/ /*Academic*/
if count=11 then site_id=33; /*UAB Gardendale*/
if count=12 then site_id=41; /*AMC*/        /*Academic*/
if count=13 then site_id=42; /*Memorial central*/
if count=14 then site_id=43; /*Poudre Valley*/
if count=15 then site_id=51; /*Baystate Springfield*/ /*Academic*/
if count=16 then site_id=52; /*Baystate Franklin*/
if count=17 then site_id=53; /*Wing*/
if count=18 then site_id=55; /*Noble*/

if _Other_providers_in_ED__EM_resid="EM residents" then _Other_providers_in_ED__EM_resid="EM Residen

%let RCTlist=
   _EHR_vendor
   _Community_or_academic
   _Rural__Suburban__Urban
   _Other_providers_in_ED__EM_resid
   Annual_Volume_Cat
   DATA_waiver_cat__2__2_or_more_
   Distance_Cat__2__2_
;
proc sort data=rct; by site_id;
proc sort data=oud; by site_id;
data OUDnew;
merge OUD RCT(keep=site_id Randomization_Arm__1__Intervent Site);
by site_id;
run;

proc freq data=oudNew;
table Randomization_Arm__1__Intervent*oud;
run;

;/*3443 vs 3820 = 7263 OUD visits*/
proc freq data=OUDnew;
tables Attending_Of_Record_Prov_ID;
run;

;/*320 OUD vist with missing ID*/
proc sort data=OUDnew nodupkey out=Drs; by Attending_Of_Record_Prov_ID; run;

;/*Step 1: cross over physician*/
/*Attendings on arm1*/
proc sort data=OUDnew(where=(Randomization_Arm__1__Intervent=1)) nodupkey out=DrINT; by Attending_Of_Record_Prov_ID; run;

proc sort data=OUDnew (where=(Randomization_Arm__1__Intervent=2)) nodupkey out=DrCTL; by Attending_Of_Record_Prov_ID; run;

;/*crossover physicians*/
/*How many cross-over*/
data crossDr;
merge DrINT(in=a) DrCTL(in=b);
by Attending_Of_Record_Prov_ID;
if a and b;
crossDr=1;
run;

;/*55 (54 indeed, one is -99999), need to add two more who crossed over after day 540*/
proc sort data=oudNew; by Attending_Of_Record_Prov_ID; run;
data CrossTemp;
merge crossDr(in=a keep=Attending_Of_Record_Prov_ID crossDr) OudNew ;
by Attending_Of_Record_Prov_ID;
run;

data noCrossCTL;
set CrossTemp;
if crossDr=1 and Randomization_Arm__1__Intervent=2 then delete;
if Attending_Of_Record_Prov_ID in ('111035', '110517') and Randomization_Arm__1__Intervent=2 then delete;
run;

;/*6220*/
proc freq; tables  Randomization_Arm__1__Intervent; run;

;/*Step 2: duplicate*/
proc sort data=noCrossCTL; by  Demographics_UID Arrival_Dt Arrival_Hr; run;
proc sort data=noCrossCTL nodupkey out=firstED_Haseena; by Demographics_UID; run;

;/*5126: same number as Harini's subject level data*/
proc freq; tables  Randomization_Arm__1__Intervent; run;
OPTIONS FORMCHAR="|----|+|---+=|-/<>*";

proc format;
  value randfor 1="Intervention" 2="Usual Care";
  value hcsfor 40='Colorado' 50='Baystate' 30='UAB' 10= 'Yale' 20='UNC';
run;

proc format;
  value rangewk
    0-6='WK 01'
    7-13='WK 02'
    14-20='WK 03'
    21-27='WK 04'
    28-34='WK 05'
    35-41='WK 06'
    42-48='WK 07'
    49-55='WK 08'
    56-62='WK 09'
    63-69='WK 10'
    70-76='WK 11'
    77-83='WK 12'
    84-90='WK 13'
    91-97='WK 14'
    98-104='WK 15'
    105-111='WK 16'
    112-118='WK 17'
    119-125='WK 18'
    126-132='WK 19'
    133-139='WK 20'
    140-146='WK 21'
    147-153='WK 22'
    154-160='WK 23'
    161-167='WK 24'
    168-174='WK 25'
    175-181='WK 26'
    182-188='WK 27'
    189-195='WK 28'
    196-202='WK 29'
    203-209='WK 30'
    210-216='WK 31'
    217-223='WK 32'
    224-230='WK 33'
    231-237='WK 34'
    238-244='WK 35'
    245-251='WK 36'
    252-258='WK 37'
    259-265='WK 38'
    266-272='WK 39'
    273-279='WK 40'
    280-286='WK 41'
    287-293='WK 42'
    294-300='WK 43'
    301-307='WK 44'
    308-314='WK 45'
    315-321='WK 46';
value rangemonth
0-30 = 'Month 1'
31-60 = 'Month 2'
61-90 = 'Month 3'
91-120 = 'Month 4'
121-150 = 'Month 5'
151-180 = 'Month 6'
181-210 = 'Month 7'
211-240 = 'Month 8'
241-270 = 'Month 9'
271-300 = 'Month 10'
301-330 = 'Month 11'
331-360 = 'Month 12'
361-390 = 'Month 13'
391-420 = 'Month 14'
421-450 = 'Month 15'
451-480 = 'Month 16'
481-510 = 'Month 17'
value quarter
0-90 = 'Quarter 1'
91-180 = 'Quarter 2'
181-270 = 'Quarter 3'
271-360 = 'Quarter 4'
361-450 = 'Quarter 5'
451-570 = 'Quarter 6'
;

value split
0-95 = 'Q1'
96-190 = 'Q2'
191-285 = 'Q3'
286-380 = 'Q4'
381-475 = 'Q5'
476-570 = 'Maintenance'
;
run;
/*Wk 69 only has 5 days*/

/**********Subject level data will be used for primary analysis******/
data subjectLevel;
set x.keepsubject.aug2021;
if Arrival_Dt > 540 then delete; /*up to 18 months*/
if race="other pacific islander" then race="native hawaiian or other pacific islander";
else if race="other/not listed" then race="other";
    else if race="patient refused" then race="Unknown or Not Reported";
    else if race in ("unknown","patient refused", "-99999") then race="Unknown or Not Reported";
if ethnicity in ("unknown","patient refused", "-99999") then ethnicity="Unknown or Not Reported";
if insurance_class = ("-99999") then insurance_class="Unknown or Not Reported";
if Narcan_Inpatient=-99999 then Narcan_Inpatient=.;
if Narcan24Mo not in (1,2) then Narcan24Mo=.;
if On_CS_Med = -99999 then On_CS_Med=.;

label OUD_Dx="Phenotype";
label Narcan_Inpatient="Naloxone prescribed during encounter as inpatient medication";
label Narcan24Mo="Prescribed Naloxone within past 24 months";
label PL_OUD_Dx="OUD Diagnosis on Problems List at time of encounter";
run;
/*5126 unique OUD patients*/
proc contents data=subjectLevel varnum; run;
proc sort data=subjectLevel nodupkey; by Demographics_UId Arrival_Dt /*Arrival_Dt*/; run;
/*no duplicates: 5126*/

/****************************Care team****************************/
data careTeam;
set x.educaare_teams;
if Arrival_Dt > 540 then delete; /*up to 18 months*/
run;
/*18337*/
/*Find team_relation for each attending_id*/
data teamRole;
set careTeam;
where prov_id ne "-99999";
rename prov_id=Attending_Of_Record_Prov_ID;
keep prov_id Team_Relation;
run;
proc sort data=TeamRole nodupkey; by Attending_Of_Record_Prov_ID; run; /*1425 unique provide ID with a role (title)*/

/*how many encounters are by nurse*/
proc sort data=subjectLevel; by Attending_Of_Record_Prov_ID; run;

Data FirstEncounters noTeamRol;
merge subjectLevel(in=a) teamRole(in=b);
by Attending_Of_Record_Prov_ID;
if a then output FirstEncounters; /*5126*/
if a and not b then output noTeamRol; /*650 provider ID (include duplicates) not have a team r
run;
proc freq data=FirstEncounters;
tables team_relation/missing;
run; /*From Ted: exclude 31+27 encounters*/

proc sort data=noTeamROL nodupkey; by Attending_Of_Record_Prov_ID; run; /*118 (exclue -99999) attending ID without role*/

/*provider data*/
proc format;
*value randfor 1="Intervention" 2="Control";
value hcsfor 40='Colorado' 50='Baystate' 30='UAB' 10= 'Yale' 20='UNC';
value provgen 1='Female' 2='Male' 0='Unknown';
value provage
0='Unknown' 1='<35' 2='35-44' 3='45-54' 4='55-64' 5='>64';
value newProvage
0='Unknown' 1='<35' 2='35-44' 3='45+' 4='55-64' 5='>64';
value ptAge
1='18-25' 2='26-55'
/*Check to confirm whether weived physician could be PA/NP*****/
proc sort data=provider; by Prov_Id;
proc sort data=FirstEncounters; by Attending_Of_Record_Prov_ID;

data ckk;
merge provider(in=a rename=(prov_id=Attending_Of_Record_Prov_ID) drop=gender)
FirstEncounters (in=b where=(team_relation in ('Nurse Practitioner','Physician Assistant')));
by Attending_Of_Record_Prov_ID;
if b;
run;
/*68*/
proc freq data=ckk;
tables team_relation*XWaiver_Stat_Code/list missing;
run;
/*9 "no" are PA. No "yes" are PA*/
/*No problem*/

/*Create dataset for attending provider*/
/*Number of attending*/
proc sort data=FirstEncounters (where=(Attending_Of_Record_Prov_ID ne "-99999" and team_relation not in ('Nurse Practitioner','Physician Assistant'))) nodupkey out=attendCK; by Health /*599*/

/*number of ED OUD patients*/
proc freq data=FirstEncounters;
where Attending_Of_Record_Prov_ID ne "-99999" and team_relation not in ('Nurse Practitioner','Physician Assistant');
tables Rand_Arm*Healthcare_System_UId*Attending_Of_Record_Prov_ID/list out=ckatt(drop=PERCENT);
run;
/*variable count in data ckatt is number of OUD visits each ID had*/

data Attend599;
set ckatt;
/*
if Healthcare_System_Name="Yale" then sysID=10;
if find(Healthcare_System_Name,"Carolina") then sysID=20;
if find(Healthcare_System_Name,"Alabama") then sysID=30;
if Healthcare_System_Name="Colorado" then sysID=40;
if find(Healthcare_System_Name,"Baystate") then sysID=50;
if find(Healthcare_System_Name,"Universal Access") then sysID=100;
*/
newID=Healthcare_System_UId||"-"||Attending_Of_Record_Prov_ID;
run;
proc freq data=attend599; tables Healthcare_System_UId newID ; run;
proc sort data=provider; by newID;
proc sort data=Attend599; by newID; run;

data attendings;
merge Attend599(keep=newID Healthcare_System_UId Attending_Of_Record_Prov_ID count Rand_Arm in=a) provider(in=b)
  ; by newID;
if a;
if XWaiver_Stat_Code=-99999 then XWaiver_Stat_Code= .;
else if XWaiver_Stat_Code in (1,2) then xwaiver= XWaiver_Stat_Code;
  else if XWaiver_Stat_Code > 2 then xwaiver=3; /*waiverd during study*/
if count <5 then patient_num=1;
else if count <21 then patient_num=2;
  else if count > 20 then patient_num=3;
format patient_num numb. xwaiver xw.;
run;

/*two different system*/
proc print data=provider;
where prov_id in ("34928","8916");
run;

/*two providers are not from first visit*/

Obs prov_id Prov_Demographics_UId Gender Healthcare_System_UId Age_Code XWaiver_Stat_Code Entry_Date
1 110598 152 Female University of North Carolina <35 Yes, received Aug-Oct 2020 05JUL2021:11:37:57.4
2 20159442 199 Male University of North Carolina 35-44 Yes, received Aug-Oct 2020 05JUL2021:11:38:09
3 45051 334 Male Colorado 55-64 No 05JUL2021:11:39:38.617

/*779*/

**********************************************************************************Use new data:3/3/2022**********************************************************************************
/*BUP admin and RX*/
data bup_inED;
set x.edata_bup_in_ed;
if Arrival_Dt > 540 then delete; /*up to 18 months*/
BUP_ED=1;
if site_Id=44 then site_id=43; /*Medical Center of the Rockies-crosses with PVH to PVH*/
if site_id=54 then site_id=53; /*Baystate Mary Lane to Baystate Wing*/
if site_id=32 then site_id=31; /*UAB-Highlands to UAB-Main*/
if ED_BUP_Description="nalOXONE" then delete;
run;
/*779*/
proc freq data=bup_inEd;
tables
Arrival_Dt
ED_BUP_Description
ED_BUP_Mar_Action
ED_BUP_Time
;
run;
proc sort data=BUP_inED; by Demographics_UId Arrival_Dt Arrival_Hr; run;
proc print data=bup_inED; where Demographics_UId=865; run;
proc sort data=BUP_inED nodupkey out=bup_inED_noDup; by Demographics_UId Arrival_Dt /*Arrival_Hr*/; /*592: same arrival hour may have more than one BUP admin*/
data bup_rx;
set x.edata_bup_rx;
if Arrival_Dt > 540 then delete; /*up to 18 months*/
BUP_rx_new=1;
if site_Id=44 then site_id=43; /*Medical Center of the Rockies-crosses with PVH to PVH*/
if site_id=54 then site_id=53; /*Baystate Mary Lane to Baystate Wing*/
if site_id=32 then site_id=31; /*UAB-Highlands to UAB-Main*/
if BUP_Rx_Description in ("Acetaminophen","Albuterol","Cephalexin","Gabapentin","Methadone","Tramadol","Clonidine","Docusate","Loratadine","Lorazepam","MethylPREDNISolone","Neomycin/Colistin/Hydrocortisone","Ondansetron","Oxymetazoline Nasal")
    then delete;
run;
/*885*/
proc sort data=BUP_Rx; by Demographics_UId Arrival_Dt Arrival_Hr; run;
proc print data=BUP_rx; where Demographics_UId in (865,309, 192); run;
proc sort data=BUP rx nodupkey out=BUP_rx_noDup; by Demographics_UId Arrival_Dt Arrival_Hr;run; /*66*/
Data BUP_comb_temp;
merge BUP_rx_noDUP(drop=Entry_date -- Deleted_reason in=a)
    BUP_inED_noDUP(drop=Entry_date -- Deleted_reason in=b)
    ;
by Demographics_UId Arrival_Dt Arrival_Hr;
*if a and b;
run;
/*353 overlap (may have duplicates)/900*/
proc print; run;
proc sort data=Firstencounters; by Demographics_UId Arrival_Dt Arrival_Hr;

Data BUP_comb_temp1;
merge BUP_comb_temp Firstencounters(in=a keep=Demographics_UId Arrival_Dt Arrival_Hr);
by Demographics_UId Arrival_Dt Arrival_Hr;
if a;
run;
/*5126*/
proc freq data=bup_comb_temp1;
tables BUP_ED*BUP_rx_new/missing;
run;
proc sort data=Firstencounters; by Demographics_UId; run;
data one5126;
merge /*subjectLevel*/ Firstencounters
    bup_comb_temp1(keep=Demographics_UId BUP_ED BUP_rx_new BUP_Rx_Ord_Prov_Id
    BUP_Rx_Authrzing_Prov_Id ED_BUP_Presc_Prov_Id ED_BUP_Ord_Prov_Id BUP_Rx_Description ED_BUP_Description
    /*one(keep=Demographics_UId tot_bupRx any_bupRx)*/
    ;
by Demographics_UId;
if first.Demographics_UId then count=0;
count+1;
month=arrival_dt;
quarter=arrival_dt;
split=arrival_dt;
*if Attending_Of_Record_Prov_ID = "-99999" then delete; /*missing attending ID*/
if site_Id=44 then site_id=43; /*Medical Center of the Rockies-crosses with PVH to PVH*/
if site_id=54 then site_id=53; /*Baystate Mary Lane to Baystate Wing*/
if site_id=32 then site_id=31; /*UAB-Highlands to UAB-Main*/
format arrival_dt rangewk.;
format month rangemonth.;
format quarter quarter.;
format split split.;
run;
/*5126 patients/encounters: each patient counted once */
proc sort data=one5126; by Healthcare_System_UId Attending_Of_Record_Prov_ID; run;
proc sort data=attendings; by Healthcare_System_UId Attending_Of_Record_Prov_ID; run;
data final_temp;
merge one5126 attendings(rename=(gender=provider_gender));
by Healthcare_System_UId Attending_Of_Record_Prov_ID;
if Attending_Of_Record_Prov_ID = "-99999" or
team_relation in ('Nurse Practitioner','Physician Assistant') then delete;
run;
/*Merge with RCT to include constrinted covariates*/
proc sort data=RCT; by site_id;
proc sort data=final_temp; by site_id;
data final;
length race_pt $50. insurance $50.;
merge final_temp RCT;
by site_id;
if race='white or caucasian' then race_pt="White";
else if race='black or african american' then race_pt="Black";
else race_pt="Others";
if Insurance_class in ("Other", "Unknown or Not Reported", "") then Insurance="Other";
else Insurance=Insurance_class;
if team_relation = "" then team_relation="Unknown";
if provider_gender = . then provider_gender=0;
if age_code= . then age_code=0;
if xwaiver=. then xwaiver=0;
Narcan=(Narcan_Outpatient=1);/*<1% missing, treated as no*/
EMBED=(intervention=1);
if bup_rx_new=. then bup_rx_new=0;
if bup_ed=. then bup_ED=0;
if BUP_ed=1 or BUP_rx_new=1 then bupRX=1;
else bupRX=0;
if Attending_Of_Record_Prov_ID in ("34928","8916") then Attending_Of_Record_Prov_ID=Healthcare_Sys
run;
/*5047*/
proc sort data=colorado; by Demographics_UId;
proc sort data=final; by Demographics_UId; run;
data final_update_1;
update final colorado(keep=Demographics_UId intervention);
by Demographics_UId;
PROC IMPORT OUT= colorado_new
   DATAFILE= "Y:\PI Folder\Melnick Ted\EMBED\Analysis 2021, Fangyong\Documentation and emai\Haseena 4_6_2022 CO resaved.xlsx"
   DBMS=EXCEL REPLACE;
   sheet="sheet1$";
   GETNAMES=YES;
   MIXED=NO;
   SCANTEXT=YES;
   USEDATE=YES;
   SCANTIME=YES;
RUN;
/*5047 107*/

/*only need to updated colorado ethnicity*/
data final_update;
length ethn_new $25.;
merge final_update_1 colorado_new(keep=Demographics_UId Updated_Ethnicity);
by Demographics_UId;
if Healthcare_System_UId=40 then
   do;
      if updated_ethnicity in ("Hispanic", "hispanic or latino") then ethn_new="hispanic or latino";
      else if updated_ethnicity in ("Unknown", "unknown") then ethn_new="Unknown or Not Reported";
      else ethn_new="non-hispanic";
   end;
else ethn_new=ethnicity;
run;