2020 to 2023 analysis w/ CR and insurance

Setup

Code to load packages

```
#packages
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.3     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.4.3     v tibble    3.2.1
## v lubridate 1.9.2     v tidyr     1.3.0
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(readxl)
library(readr)
library(parsedate)
```

```
##
## Attaching package: 'parsedate'
##
## The following object is masked from 'package:readr':
##
##     parse_date
```

```
library(janitor)
```

```
##
## Attaching package: 'janitor'
##
## The following objects are masked from 'package:stats':
##
##     chisq.test, fisher.test
```

```
library(lubridate)
library(labelled)
library(overviewR)
```

Code to load data

```
library(readxl)
visit_pull <- read_excel("C_F_AllPatients_2020to2023.xlsx", na = c(" ", "", "NA")) %>%  as_tibble()
```

```
## New names:
## * `Pat Enc Csn ID` -> `Pat Enc Csn ID...1`
## * `Pat Enc Csn ID` -> `Pat Enc Csn ID...32`
## * `Descovy Ordered Num` -> `Descovy Ordered Num...96`
## * `Descovy Ordered Num` -> `Descovy Ordered Num...97`
```

```
chart_review <- C_F_AdultChartReview_2020to2022 <- read_excel("C_F_AdultChartReview_2020to2022.xlsx",
                                                              col_types = c("skip", "skip", "numeric",
path <- "C_F_All_Insurance_Pull.xlsx"
insurance <-path |>
  excel_sheets() |>
  set_names() |> map(read_excel, path = path) |> list_rbind() |> as_tibble()
```

visit_pull has 942 obs of 192 variables chart_review has 691 obs of 36 variables insurance has 415 obs of 25 variables cleaning names:

```
#name cleaning
cn_visit_pull <-clean_names(visit_pull)
cn_visit_pull <- cn_visit_pull %>% rename(pat_enc_csn_id = pat_enc_csn_id_1) %>% mutate(pat_enc_csn_id =
cn_chart_review <-clean_names(chart_review)
cn_chart_review <- cn_chart_review %>% mutate(pat_enc_csn_id = as.character(pat_enc_csn_id))
cn_insurance <-clean_names(insurance)
#optional removing of old DF
rm(chart_review,visit_pull,insurance,path)
```

reformatting before join

```
origin <- "1904-01-01" %>% as_date()
# set TZ to east coast standard
time_zone <- "America/New_York"
#pathway went live on 7/13/21
pathway_start <- mdy("7/13/21", tz = time_zone)
#Story board notification went live on 9/8/21
sb_start <- mdy("9/8/21", tz = time_zone)
# Date CDC put out new STI tx guidelines
CDC_STI_date <- mdy("7/23/2021", tz = time_zone)
rf_visit_pull <- cn_visit_pull %>% mutate(
  arrive_dt = as_datetime(ed_arrival_time, tz = time_zone),
  wtf_departure_dt = as_datetime(ed_departure_time, tz = time_zone),
  # don't use this one for now. Altered by obs/psych/admit pts?
  dispo_dt = as_datetime(ed_disposition_time, tz = time_zone),
  avs_dt = as_datetime(ed_avs_printed_ts, tz = time_zone),
  pt_seen_dt = as_datetime(seen_by_provider_ts, tz = time_zone),
  ed_depart_dt = as_datetime(ed_depart_ts, tz = time_zone),
  #departure TS seems to be the one that is earlier,  departure_time usually later (affected by obs/adm
  patient_birth_date = as_date(patient_birth_date, tz = time_zone)
) %>%
  mutate(ed_arrival_date = date(arrive_dt)) %>%
  mutate(exposure.num = case_when(arrive_dt < pathway_start ~ 0,
```

```
                                arrive_dt > sb_start ~ 1,
                                .default = NA)) %>%
mutate(between = case_when(arrive_dt >= pathway_start &
                           arrive_dt <= sb_start ~ 1,
                         .default = 0)) %>%
mutate(
  exposure.char = case_when(
    exposure.num == 0 ~ "pre intervention",
    exposure.num == 1 ~ "post intervention",
    between == 1 ~ "between period",
    .default = NA
  )
) %>%
mutate(
  ed_disposition = factor(ed_disposition, ordered = FALSE),
  exposure.char = factor(
    exposure.char,
    ordered = FALSE,
    levels = c("pre intervention", "between period", "post intervention")
  ),
  pathway = factor(agile_md_used_yn, exclude = NULL),
  patient_birth_date = as_date(patient_birth_date)
) %>%
add_count(mrn) %>%
rename(num_visits = n) %>%
dplyr::mutate(
  repeater = case_when(num_visits > 1 ~ TRUE,
                       num_visits == 1 ~ FALSE,
                       .default = FALSE),
  race = factor(patient_race, ordered = FALSE),
  age_group = factor(
    age_group,
    levels = c(
      "0-12 yrs",
      "13-16 yrs",
      "17-19 yrs",
      "20-29 yrs",
      "30-39 yrs",
      "40-49 yrs",
      "50-59 yrs",
      "60-69 yrs",
      "70-79 yrs",
      "80-89 yrs",
      "90-99 yrs"
    ),
    ordered = TRUE
  ),
  minor = case_when(age < 18 ~ 1,
                    age >= 18 ~ 0),
  intervention = factor(exposure.char)
) %>%
dplyr::mutate(
  race_bwo = case_when(
```

```
      race == "White" ~ "White",
      race == "Black or African American" ~ "Black",
      race != c("White", "Black or African American") ~ "Other"
    ),
    female = case_when(gender == "Female" ~ 1,
                       gender == "Male" ~ 0),
    age_3_group = case_when(age < 18 ~ "< 18",
                            age >= 18 & age < 55 ~ "18 - 55",
                            age >= 55 ~ "55+")
  ) %>%
  dplyr::mutate(
    female_u55 = ifelse(age < 55 & female == 1, 1, 0),
    white = ifelse(race_bwo == "White", 1, 0),
    black = ifelse(race_bwo == "Black", 1, 0),
    other = ifelse(race_bwo == "Other", 1, 0)
  )

levels(rf_visit_pull$pathway) <-
  list("Didn't Use Pathway" = "N", "Used Pathway" = "Y")
# need to also reformat cn_chart_review ed arrival date to get same format
cn_chart_review <- cn_chart_review %>%
  mutate(ed_arrival_date = as_date(ed_arrival_date))
```

#Joining datasets, excluding patients

```
minors <- rf_visit_pull %>% filter(age < 18)
#adult_visit_pull <- rf_visit_pull %>% filter(age >= 18) %>%  mutate(ed_arrival_date = date(arrive_dt))
#now joining in minors as well with plan to filter them out at the end
DF1 <- left_join(rf_visit_pull, cn_chart_review, by = join_by(pat_enc_csn_id == pat_enc_csn_id, ed_arri
s_insurance <- cn_insurance %>% select(pat_enc_csn_id, age, primary_coverage_payor_name, primary_coverag

DF2 <- left_join(DF1, s_insurance, by = join_by(age.x == age, ed_arrival_month == ed_arrival_month, pat

npts<- n_distinct(DF2$pat_enc_csn_id, na.rm=TRUE) %>% as.character()
nminors <-n_distinct(minors$pat_enc_csn_id, na.rm=TRUE) %>% as.character()
cat("-DF 2 is now the visit pull data with addended chart review data and additional insurance data. To
```

```
## -DF 2 is now the visit pull data with addended chart review data and additional insurance data. Total
```

```
cat("-", nminors, "minor patients are to be excluded")
```

```
## - 264 minor patients are to be excluded
```

```
rm(npts, nminors)
```

#Main Dataframe (to be re-uploaded to onedrive)

```
variables1 <- colnames(cn_chart_review)
variables2 <-
  DF2 %>% select(arrive_dt:primary_coverage_benefit_plan_name.y) %>% colnames()
variables3 <-
```

```
  DF2 %>% select(ends_with(c(
    "ts", "num", "id", "yn", "name", "dt", "time", "zip", "name"
  ))) %>% colnames()
vars <- c(variables1, variables2, variables3)
#r stands for reduced
r_DF1 <- DF2 %>%
  select(matches(vars)) %>%
  mutate(
    insurance_cov = case_when(
      !is.na(primary_coverage_benefit_plan_name.x) ~ primary_coverage_benefit_plan_name.x,
      is.na(primary_coverage_benefit_plan_name.x) & !is.na(primary_coverage_benefit_plan_name.y) ~ prima
      .default = NA
    )
  ) %>%
  mutate(
    insurance_pay = case_when(
      !is.na(primary_coverage_payor_name.x) ~ primary_coverage_payor_name.x,
      is.na(primary_coverage_payor_name.x) & !is.na(primary_coverage_payor_name.y) ~ primary_coverage_pa
    )
  ) # %>%
  #no longer need: mutate(reason_to_exclude = case_when(is.na(reason_to_exclude) ~ 0, .default = reason
```

insurance categories:

```
Medicare <-("MCR|MEDICARE|CONNECTICARE")
Medicaid <-("MCD|HUSKY|MEDICAID" )
SA_insurance <-("SEXUAL|ASSAULT")
private<-("UNITED HEALTHCARE|AETNA|HARVARD PILGRIM|BCBS|CENTURY PREFERRED|OXFORD|CIGNA|COMMERCIAL GENER
r_DF2 <- r_DF1 %>% mutate(
  insurance = case_when(
    str_detect(insurance_pay, paste(Medicare)) ~ "Medicare",
    str_detect(insurance_pay, paste(Medicaid)) ~ "Medicaid",
    str_detect(insurance_pay, paste(SA_insurance)) ~ "Sexual Assault",
    str_detect(insurance_pay, paste(private)) ~ "Private Insurance",
    is.na(insurance_pay) ~ "Uninsured/Self-Pay",
    insurance_pay == "0" ~"Uninsured/Self-Pay",
    .default = insurance_pay
  )
)
cat("The insurance categories are:", unique(r_DF2$insurance))
```

```
## The insurance categories are: Sexual Assault Medicaid Private Insurance Medicare Uninsured/Self-Pay
```

trauma patient data:

need to first re-format so single CSN per row, multiple dx columns doesn't work (too many) plan: make new variable for dx category, then consolidate by grouping by ICD10 codes, ESI level, Procedure (imaging)

```
library(readr)
C_F_Trauma <- read_csv("C_F_Trauma.csv",
    col_types = cols(order_time = col_skip(),
```

```
        exam_begin_time = col_skip(), authorizing_provider_type = col_skip(),
        final_dx_poa_c = col_skip(), dx_poa_flag = col_skip()))
```

ICD10 dx categories:

```
#import dataset w/ classifiers:
DY_ICD <- read_excel("DY_ICD10_Classifications.xlsx")
```

```
## New names:
## * `` -> `...9`
## * `` -> `...10`
## * `` -> `...11`
## * `Other` -> `Other...12`
## * `Other` -> `Other...13`
```

```
ICD_names <- DY_ICD %>%
  pivot_longer(
    everything(),
    cols_vary = "slowest",
    names_to = "category",
    values_to = "dx",
    values_drop_na = TRUE
  ) %>%
  relocate(dx, .before = category) %>%
  mutate(category = if_else(
    category %in% c("Other...12", "Other...13"), "Other", category
  )) %>%
  distinct()
#creating dx category vectors
etoh <- ICD_names %>% filter(category == "Alcohol") %>% pull(var = dx, name = category)
pain <-  ICD_names %>% filter(category == "Pain") %>% pull(var = dx, name = category)
psych <-  ICD_names %>% filter(category == "Psychiatry") %>% pull(var = dx, name = category)
drugs <-  ICD_names %>% filter(category == "Illicit Drug Use") %>% pull(var = dx, name = category)
minor_injury <- ICD_names %>% filter(category == "Minor Injury") %>% pull(var = dx, name = category)
major_injury <- ICD_names %>% filter(category == "Minor Injury") %>% pull(var = dx, name = category)
sdoh <- ICD_names %>% filter(category == "Homelessness/SDOH") %>% pull(var = dx, name = category)
```

New variable of dx categories for trauma set

```
C_F_Trauma_dxcat <- C_F_Trauma %>%
  mutate(
    Etoh = if_else(dx_name %in% etoh, 1, NA),
    Pain = if_else(dx_name %in% pain, 1, NA),
    Psych = if_else(dx_name %in% psych, 1, NA),
    Drugs = if_else(dx_name %in% drugs, 1, NA),
    Minor_injury = if_else(dx_name %in% minor_injury, 1, NA),
    Major_injury = if_else(dx_name %in% major_injury, 1, NA),
    SDOH = if_else(dx_name %in% sdoh, 1, NA),
    )
```

Now shorten and condense to prep for join

```
short_trauma <- C_F_Trauma_dxcat %>%
  select(pat_enc_csn_id, esi_level, trauma_case_yn) %>% distinct()
```

```
a<- C_F_Trauma_dxcat %>%  select(pat_enc_csn_id, Etoh) %>% filter(!is.na(Etoh)) %>% distinct()
b<- C_F_Trauma_dxcat %>%  select(pat_enc_csn_id, Pain) %>% filter(!is.na(Pain)) %>% distinct()
c<-C_F_Trauma_dxcat %>%  select(pat_enc_csn_id, Psych) %>% filter(!is.na(Psych)) %>% distinct()
d<-C_F_Trauma_dxcat %>%  select(pat_enc_csn_id, Drugs) %>% filter(!is.na(Drugs)) %>% distinct()
e<-C_F_Trauma_dxcat %>%  select(pat_enc_csn_id, Minor_injury) %>% filter(!is.na(Minor_injury)) %>% dist
f<-C_F_Trauma_dxcat %>%  select(pat_enc_csn_id, Major_injury) %>% filter(!is.na(Major_injury)) %>% dist
g<-C_F_Trauma_dxcat %>%  select(pat_enc_csn_id, SDOH) %>% filter(!is.na(SDOH)) %>% distinct()
```

join them all back

```
short_trauma_cat<-left_join(short_trauma, a, by = join_by(pat_enc_csn_id))
short_trauma_cat<-left_join(short_trauma_cat, b, by = join_by(pat_enc_csn_id))
short_trauma_cat<-left_join(short_trauma_cat, c, by = join_by(pat_enc_csn_id))
short_trauma_cat<-left_join(short_trauma_cat, d, by = join_by(pat_enc_csn_id))
short_trauma_cat<-left_join(short_trauma_cat, e, by = join_by(pat_enc_csn_id))
short_trauma_cat<-left_join(short_trauma_cat, f, by = join_by(pat_enc_csn_id))
short_trauma_cat<-left_join(short_trauma_cat, g, by = join_by(pat_enc_csn_id)) %>% mutate(pat_enc_csn_i
rm(a,b,c,d,e,f,g)
```

joining trauma/dx/ esi data with other joined DF

```
full_merge_DF <- left_join(r_DF2, short_trauma_cat, by = join_by(pat_enc_csn_id)) %>% distinct()
```

creating cohorts

```
excluded_patients <- full_merge_DF %>% filter(!is.na(exclude))
r_DF3 <- full_merge_DF %>% filter(is.na(exclude))
excluded_patients <- excluded_patients %>% add_value_labels(reason_to_exclude = c( "Seen earlier" = "1"
excluded_patients %>% group_by(reason_to_exclude) %>% summarise(n=n())
```

```
## # A tibble: 4 x 2
##   reason_to_exclude                       n
##   <fct>                               <int>
## 1 Seen earlier                           14
## 2 Patient reports not being assaulted    12
## 3 not excluded                            2
## 4 Eloped                                  9
```

```
n_excluded <-n_distinct(excluded_patients$pat_enc_csn_id, na.rm=TRUE) %>% as.character()
cat("-", n_excluded, "patients were excluded based on chart review.")
```

```
## - 37 patients were excluded based on chart review.
```

```r
rm(n_excluded)
```

```r
post_WO <- full_merge_DF %>% filter(ed_arrival_date > sb_start) %>% filter(age >= 18)
```