

Healthcare Study Case 1

Xi Liang

April 5, 2017

Objective

The objective of this assignment is to create an analytic file from a dataset provided by QuantileIMS. An attrition data would be created upon the completion of the the analytic file.

We would like to identify patients that received either antidiabetics or anticoagulants from 1/1/13 to 3/31/13 (index date). Then and we would create pre and post index period (+/- 180 days from the index date), and generate new variables such as drug use, medical claims, diagnoses, medication adherence and medical events, for downstream analysis and modeling, which could help us to understand pre and post-index characteristics and drug adherence for specific patients.

After the data wrangling process, one should be able to answer these three questions:

- How many (unique) patients are in the desired index range that have taken AD or AC drugs?
- How many of those patients have more than one drug on the index date?
- How many have both antidiabetics and anticoagulant drugs on the index date?

```
library(readr)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

Loading both of the calim data and Rx data

```
claim_data <- read.csv('data/uopjan16_claims_in.csv', stringsAsFactors = FALSE)
rx_data <- read.csv('data/uopjanref_rx_lookup.csv', stringsAsFactors = FALSE)

dim(claim_data)
```

```
## [1] 1788544      22
```

```
dim(rx_data)
```

```
## [1] 13557      19
```

Data Wrangling

Filtering NDC list in order to get the NDC numbers for Antidiabetics and Anticoagulants

```
#change the dtypes of the ndc column to numeric.
claim_data$ndc <- as.numeric(claim_data$ndc)
```

```
## Warning: NAs introduced by coercion
```

```
rx_data$National.Drug.Code <- as.numeric(rx_data$National.Drug.Code)
```

```
#Creating an NDC list for AD and AC drugs
NDC_list <- filter(rx_data, rx_group == 'ANTIDIABETICS' | rx_group == 'ANTICOAGULANTS')
```

Merging claim data and NDC list to get record of patients that received either AC or AD medications

```
#Patients in the data given AC or AD medications
claim_w_ACAD <- merge(claim_data, NDC_list, by.x = 'ndc', by.y = 'National.Drug.Code') %>% tbl_df()

claim_w_ACAD$pat_id %>% unique %>% length
```

```
## [1] 5799
```

Narrowing down patients had their medications prescribed within index date range (1/1/13 to 3/31/13).

```
claim_w_ACAD$from_dt <- as.Date(claim_w_ACAD$from_dt)
claim_w_ACAD$to_dt <- as.Date(claim_w_ACAD$to_dt)

claim_w_index_date <- filter(claim_w_ACAD, from_dt >= "2013/01/01" & to_dt <= "2013/03/31")
```

Using the minimum prescription date in order to get the index date of a specific patient

```
#First occurrence of either AC/AD for these patients
first_occurrence <- group_by(claim_w_index_date, pat_id) %>% filter(from_dt == min(from_dt))
first_occurrence$pat_id %>% unique %>% length
```

```
## [1] 5798
```

```
#We have more unique patient more than it suppose to be, because some patients receive same type of drug.
first_occurrence %>% dim()
```

```
## [1] 7034 40
```

```
table(table(first_occurrence$pat_id) >1)
```

```
##
## FALSE TRUE
## 4771 1027
```

```
#Select column that will be used in the final version of the analytic file.
analytic_file_final <- first_occurrence %>% select(pat_id, rx_group, gen_nm, IDX_Date = from_dt)
analytic_file_final <- analytic_file_final %>%
  mutate(IDX_group = ifelse(rx_group == 'ANTIDIABETICS', 'AD', 'AC'))
```

Populate IDX_Group Column: The first claim for an AD, AC, both on same date (DC) or neither (null) during the index period (1/1/13 - 3/31/13)

```
# check which patients who received more than one type of medications on the index date; extract their
tmp_index <- analytic_file_final %>% group_by(pat_id) %>% summarise(len = length(IDX_group)) %>% filter(
tmp_index
```

```
## # A tibble: 1,027 × 2
##       pat_id   len
##       <chr> <int>
## 1 0309AAAAAAAEEDAZ     3
## 2 0309AAAAAAAFZJWC     3
## 3 0309AAAAAAAOZFXU     2
## 4 0309AAAAAATHFQB      2
## 5 0309AAAAAAAWNCPR     2
## 6 0309AAAAAABBIGCP     3
## 7 0309AAAAAABBKDPS     3
## 8 0309AAAAAABFWZRI     3
## 9 0309AAAAAABHLQFJ     2
## 10 0309AAAAAABIXTCA     3
## # ... with 1,017 more rows
```

```
#filter out unrelated patients and narrow the dataframe down to patients who received more than one med
tmp_index_2 <- analytic_file_final %>% filter(pat_id %in% tmp_index$pat_id)
```

```
#check which patients received both AC and AD medications on their index date
tmp3 <- tmp_index_2 %>% group_by(pat_id) %>% mutate(tmp = unique(IDX_group)[1], tmp2 = unique(IDX_group)
tmp3 <-filter(tmp3, tmp != is.na(tmp) & tmp2 != is.na(tmp2))
tmp3
```

```
## Source: local data frame [70 x 7]
## Groups: pat_id [28]
##
##       pat_id      rx_group      gen_nm  IDX_Date
##       <chr>      <chr>      <chr>    <date>
## 1 w112AAAAAAAEVCAUN ANTIDIABETICS INSULIN LISPRO (HUMAN) 2013-02-01
## 2 p615AAAAAADWMHYX  ANTIDIABETICS INSULIN LISPRO (HUMAN) 2013-02-28
## 3 a786AAAAAAGJRUQT  ANTIDIABETICS INSULIN LISPRO (HUMAN) 2013-01-01
## 4 s103AAAAAAEORAYZ  ANTIDIABETICS      SAXAGLIPTIN HCL 2013-03-15
```

```
## 5 s103AAAAAAEORAYZ ANTIDIABETICS SAXAGLIPTIN HCL 2013-03-15
## 6 f941AAAAAAETKDSH ANTIDIABETICS SITAGLIPTIN PHOSPHATE 2013-02-24
## 7 mg19AAAAAAFWRCW ANTIDIABETICS SITAGLIPTIN PHOSPHATE 2013-01-03
## 8 s103AAAAAAJDCJBO ANTIDIABETICS SITAGLIPTIN PHOSPHATE 2013-01-14
## 9 mr01AAAAAACGUDMR ANTICOAGULANTS WARFARIN SODIUM 2013-01-14
## 10 s107AAAAABFPLWJY ANTIDIABETICS INSULIN GLARGINE 2013-01-04
## # ... with 60 more rows, and 3 more variables: IDX_group <chr>, tmp <chr>,
## # tmp2 <chr>
```

```
#adding the IDX_Group into the analytic_file_final, label patients that received both drug types as 'DC'
analytic_file_final[analytic_file_final$pat_id %in% tmp3$pat_id,]$IDX_group = 'DC'
```

```
#removing duplicated pat_id
analytic_file_final <- subset(analytic_file_final, !duplicated(analytic_file_final[,1]))
```

Populating YOB, Pat_Age, and Pat_sex for specific patient

```
enroll_data <- read.csv('data/uopjan16_enroll_in.csv') %>% tbl_df
glimpse(enroll_data)
```

```
## Observations: 5,802
## Variables: 13
## $ pat_id      <fctr> 0309AAAAAAAATOHM, 0309AAAAAAAEEEDAZ, 0309AAAAAAAFFHC...
## $ eststring   <fctr> -----...
## $ clm_frst    <fctr> 2001-02-19, 2001-01-08, 2001-01-09, 2001-01-02, 20...
## $ clm_last    <fctr> 2015-09-30, 2015-09-30, 2015-09-30, 2015-09-30, 20...
## $ enr_frst    <fctr> 2001-01-01, 2001-01-01, 2001-01-01, 2001-01-01, 20...
## $ enr_last    <fctr> 2015-09-30, 2015-09-30, 2015-09-30, 2015-09-30, 20...
## $ mxce_fst    <fctr> 2001-01-01, 2001-01-01, 2001-01-01, 2001-01-01, 20...
## $ mxce_lst    <fctr> 2015-09-30, 2015-09-30, 2015-09-30, 2015-09-30, 20...
## $ der_sex     <fctr> M, F, M, M, F, M, F, M, F, M, F, F, M, F, F, M, M,...
## $ der_yob     <int> 1949, 1944, 1932, 1943, 1961, 1947, 1978, 1950, 194...
## $ nbr_clms    <int> 1016, 1091, 1130, 1841, 1370, 1105, 372, 1912, 1286...
## $ mon_totl    <int> 149, 177, 177, 177, 175, 177, 94, 177, 177, 150, 17...
## $ pat_region  <fctr> MW, MW, MW, MW, MW, MW, MW, MW, MW, MW, MW, MW, MW...
```

```
analytic_file_final <- select(enroll_data, pat_id, der_sex, der_yob) %>% merge(analytic_file_final, by =
analytic_file_final %>% tbl_df
```

```
## # A tibble: 5,798 × 7
##       pat_id der_sex der_yob rx_group
##       <fctr> <fctr>   <int>   <chr>
## 1 0309AAAAAAAATOHM      M    1949 ANTIDIABETICS
## 2 0309AAAAAAAEEEDAZ      F    1944 ANTIDIABETICS
## 3 0309AAAAAAAFFHCZL      M    1932 ANTIDIABETICS
## 4 0309AAAAAAAFFZJWC      M    1943 ANTIDIABETICS
## 5 0309AAAAAAAIFEJH      F    1961 ANTIDIABETICS
## 6 0309AAAAAAAAKFIHG      M    1947 ANTIDIABETICS
## 7 0309AAAAAAAAXRBA      F    1978 ANTIDIABETICS
## 8 0309AAAAAAAALOFMS      M    1950 ANTICOAGULANTS
## 9 0309AAAAAAAALUHNG      F    1941 ANTIDIABETICS
```

```
## 10 0309AAAAAAAALUQDC      M      1938 ANTIDIABETICS
## # ... with 5,788 more rows, and 3 more variables: gen_nm <chr>,
## #   IDX_Date <date>, IDX_group <chr>
```

```
#calculate patients' ages as today
```

```
analytic_file_final <- analytic_file_final %>% mutate(age = parse_number(2013 - der_yob))
analytic_file_final %>% tbl_df
```

```
## # A tibble: 5,798 × 8
##           pat_id der_sex der_yob      rx_group
##           <fctr>  <fctr>   <int>      <chr>
## 1 0309AAAAAAAATOHM      M     1949 ANTIDIABETICS
## 2 0309AAAAAAAEEEDAZ      F     1944 ANTIDIABETICS
## 3 0309AAAAAAAFHCZL      M     1932 ANTIDIABETICS
## 4 0309AAAAAAAFZJWC      M     1943 ANTIDIABETICS
## 5 0309AAAAAAAIFEJH      F     1961 ANTIDIABETICS
## 6 0309AAAAAAAKFIHG      M     1947 ANTIDIABETICS
## 7 0309AAAAAAAKXRBA      F     1978 ANTIDIABETICS
## 8 0309AAAAAAAALOFMS      M     1950 ANTICOAGULANTS
## 9 0309AAAAAAAALUHNG      F     1941 ANTIDIABETICS
## 10 0309AAAAAAAALUQDC      M     1938 ANTIDIABETICS
## # ... with 5,788 more rows, and 4 more variables: gen_nm <chr>,
## #   IDX_Date <date>, IDX_group <chr>, age <dbl>
```

Creating Pre and Post Index Date for each patient

```
#creating pre and post date index
```

```
pre_post_filter <- select(analytic_file_final, pat_id, IDX_Date) %>% mutate(pre_index = IDX_Date - 180,
pre_post_filter %>% dim()
```

```
## [1] 5798      4
```

```
pre_post_filter %>% head
```

```
##           pat_id   IDX_Date pre_index post_index
## 1 0309AAAAAAAATOHM 2013-02-03 2012-08-07 2013-08-02
## 2 0309AAAAAAAEEEDAZ 2013-01-05 2012-07-09 2013-07-04
## 3 0309AAAAAAAFHCZL 2013-02-10 2012-08-14 2013-08-09
## 4 0309AAAAAAAFZJWC 2013-01-18 2012-07-22 2013-07-17
## 5 0309AAAAAAAIFEJH 2013-01-07 2012-07-11 2013-07-06
## 6 0309AAAAAAAKFIHG 2013-02-22 2012-08-26 2013-08-21
```

```
#filter the data with post and pre index date
```

```
tmp_df<-merge(claim_data, pre_post_filter, by = 'pat_id') %>% tbl_df
```

```
tmp_df$from_dt = as.Date(tmp_df$from_dt)
```

```
tmp_df$to_dt = as.Date(tmp_df$to_dt)
```

```
tmp_df %>% head %>% print.data.frame(width = Inf)
```

```
##          pat_id rec_spec rectype ALLOWED DAYSSUP QUAN  rec_ix
## 1 0309AAAAAAAATOHM  INTERN      A    0.00      NA   NA 1200812
## 2 0309AAAAAAAATOHM  INTERN      A    0.00      NA   NA 1200814
## 3 0309AAAAAAAATOHM  INTERN      A    0.00      NA   NA 1200815
## 4 0309AAAAAAAATOHM   GP_FP      A   15.38      NA   NA 1200801
## 5 0309AAAAAAAATOHM   GP_FP      A   24.46      NA   NA 1200802
## 6 0309AAAAAAAATOHM  INTERN      A    0.00      NA   NA 1200813
##      from_dt      to_dt          prov_id  paid charge ndc proc_cde diag1
## 1 2012-03-23 2012-03-23 0309AAAAAAAAB  0.00 219.70 NA    93005 78659
## 2 2012-03-23 2012-03-23 0309AAAAAAAAB  0.00  35.20 NA    36415 78659
## 3 2012-03-23 2012-03-23 0309AAAAAAAAB  0.00 255.90 NA    80053 78659
## 4 2012-03-14 2012-03-14 0309AAAAAAAAGEP 15.38  21.00 NA    80048 25000
## 5 2012-03-14 2012-03-14 0309AAAAAAAAGEP 24.46  32.00 NA    80061 25000
## 6 2012-03-23 2012-03-23 0309AAAAAAAAB  0.00  47.17 NA      272 78659
##      diag2 diag3 diag4 ptypeflg pos          cluspvld conf_num  IDX_Date
## 1  7295 25000 V5869      1 ZZ              NA 2013-02-03
## 2  7295 25000 V5869      1 ZZ 0309AAAAAAAAB      NA 2013-02-03
## 3  7295 25000 V5869      1 ZZ 0309AAAAAAAAB      NA 2013-02-03
## 4  4011  2724              0 11 0309AAAAAAAAGEP      NA 2013-02-03
## 5  4011  2724              0 11 0309AAAAAAAAGEP      NA 2013-02-03
## 6  7295 25000 V5869      1 ZZ 0309AAAAAAAAB      NA 2013-02-03
##      pre_index post_index
## 1 2012-08-07 2013-08-02
## 2 2012-08-07 2013-08-02
## 3 2012-08-07 2013-08-02
## 4 2012-08-07 2013-08-02
## 5 2012-08-07 2013-08-02
## 6 2012-08-07 2013-08-02
```

```
analytic_file_tmp <- filter(tmp_df, from_dt >= pre_index & from_dt <= post_index)

analytic_file_tmp %>% tbl_df()
```

```
## # A tibble: 694,209 × 25
##          pat_id rec_spec rectype ALLOWED DAYSSUP  QUAN  rec_ix
##          <chr>   <chr>   <chr>   <dbl>   <int> <dbl>   <int>
## 1 0309AAAAAAAATOHM  INTERN      A    0.00      NA   NA 1200904
## 2 0309AAAAAAAATOHM  INTERN      A   22.67      NA   NA 1200905
## 3 0309AAAAAAAATOHM  INTERN      A  734.08      NA   NA 1200906
## 4 0309AAAAAAAATOHM  INTERN      A  331.25      NA   NA 1200907
## 5 0309AAAAAAAATOHM  RADIOL      A  165.83      NA   NA 1200908
## 6 0309AAAAAAAATOHM  RADIOL      A   59.58      NA   NA 1200909
## 7 0309AAAAAAAATOHM   GP_FP      M  160.48      NA   NA 1200910
## 8 0309AAAAAAAATOHM   GP_FP      A   14.60      NA   NA 1200911
## 9 0309AAAAAAAATOHM   GP_FP      A    1.92      NA   NA 1200912
## 10 0309AAAAAAAATOHM   GP_FP      A    1.92      NA   NA 1200902
## # ... with 694,199 more rows, and 18 more variables: from_dt <date>,
## #   to_dt <date>, prov_id <chr>, paid <dbl>, charge <dbl>, ndc <dbl>,
## #   proc_cde <chr>, diag1 <chr>, diag2 <chr>, diag3 <chr>, diag4 <chr>,
## #   ptypeflg <int>, pos <chr>, cluspvld <chr>, conf_num <int>,
## #   IDX_Date <date>, pre_index <date>, post_index <date>
```

```
analytic_file_tmp %>% head %>% print.data.frame(width = Inf)
```

```
##           pat_id rec_spec rectype ALLOWED DAYSSUP QUAN  rec_ix
## 1 0309AAAAAAAATOHM  INTERN      A    0.00      NA   NA 1200904
## 2 0309AAAAAAAATOHM  INTERN      A   22.67      NA   NA 1200905
## 3 0309AAAAAAAATOHM  INTERN      A  734.08      NA   NA 1200906
## 4 0309AAAAAAAATOHM  INTERN      A  331.25      NA   NA 1200907
## 5 0309AAAAAAAATOHM  RADIOL      A  165.83      NA   NA 1200908
## 6 0309AAAAAAAATOHM  RADIOL      A   59.58      NA   NA 1200909
##      from_dt   to_dt      prov_id   paid charge ndc  proc_cde diag1
## 1 2013-03-26 2013-03-26 0309AAAAAAAAB  0.00   135  NA    Q9967  7295
## 2 2013-03-26 2013-03-26 0309AAAAAAAAB 22.67    91  NA    82565  7295
## 3 2013-03-26 2013-03-26 0309AAAAAAAAB 589.58  1579  NA    71275  7295
## 4 2013-03-26 2013-03-26 0309AAAAAAAAB 331.25   845  NA    93970  7295
## 5 2013-03-26 2013-03-26 0309AAAAAAAICK 165.83   272  NA    71275  78650
## 6 2013-03-26 2013-03-26 0309AAAAAAAICK  59.58   101  NA    93970  78650
##      diag2 diag3 diag4 ptypeflg pos      cluspid conf_num  IDX_Date
## 1 78605  7212 78650      1 ZZ 0309AAAAAAAAGEP      NA 2013-02-03
## 2 78605  7212 78650      1 ZZ      NA 2013-02-03
## 3 78605  7212 78650      1 ZZ      NA 2013-02-03
## 4 78605  7212 78650      1 ZZ 0309AAAAAAAAGEP      NA 2013-02-03
## 5 78605  7295      0 22      NA 2013-02-03
## 6 78605  7295      0 22 0309AAAAAAAAGEP      NA 2013-02-03
##      pre_index post_index
## 1 2012-08-07 2013-08-02
## 2 2012-08-07 2013-08-02
## 3 2012-08-07 2013-08-02
## 4 2012-08-07 2013-08-02
## 5 2012-08-07 2013-08-02
## 6 2012-08-07 2013-08-02
```

Populating Pre_hosp

A rectype = 'F', where to_dt - from_dt > 0, and the to_dt is prior to the IDX_Date

```
#extract patients' index whom met the requiriment of Pre_hosp
Pre_hosp_index <- analytic_file_tmp %>% filter(rectype == 'F', to_dt - from_dt > 0, to_dt < IDX_Date)
Pre_hosp_index <- unique(Pre_hosp_index$pat_id)

#labeling those who met the Pre_hosp requiriment = 1, else = 0
analytic_file_final <- analytic_file_final %>% mutate(Pre_hosp = ifelse(pat_id %in% Pre_hosp_index, 1, 0))
```

Populating Pre-CHF

Claim must be a rectype M or F in diag1-2-3-4

```
#these are IDH-9 codes that locate in diag1-2-3-4 columns that indicate specific patients were diagnose
diag_code = c(428, 4280, 4281, 4282, 42820, 42821, 42822, 42823, 42830, 42831, 42832, 42833,
              42840, 42841, 42842, 42843, 4289)

Pre_CHF_index <- analytic_file_tmp %>% filter(rectype == 'M' | rectype == 'F' | rectype == 'S', diag1 %in% diag_code)
```

```

diag2 %in% diag_code | diag3 %in% diag_code | diag4 %in% diag_code)
from_dt <= IDX_Date)

Pre_CHF_index <- unique(Pre_CHF_index$pat_id)

analytic_file_final <- analytic_file_final %>% mutate(Pre_CHF = ifelse(pat_id %in% Pre_CHF_index, 1, 0))

```

Populating Beta-Blocker

```

b_blockers_index <- filter(rx_data, rx_group == 'BETA BLOCKERS')
b_blockers_index <- b_blockers_index$National.Drug.Code

patient_b_blocker <- filter(analytic_file_tmp, ndc %in% b_blockers_index, from_dt < IDX_Date)
patient_b_blocker <- unique(patient_b_blocker$pat_id)

analytic_file_final <- analytic_file_final %>% mutate(Pre_B_Blocker = ifelse(pat_id %in% patient_b_blocker$pat_id, 1, 0))

```

Populating Post_IDX_Rx: Number of prescription claims for the index drug (at the generic name) from the index date to IDX-Date +180days

```

patient_ACAD_pres <- filter(analytic_file_tmp, ndc %in% NDC_list$National.Drug.Code, from_dt >= IDX_Date)
Post_IDX_Rx_Index <- patient_ACAD_pres %>% group_by(pat_id) %>% summarise(Post_IDX_Rx = length(ndc))

analytic_file_final <- merge(analytic_file_final, Post_IDX_Rx_Index, by = 'pat_id', all = TRUE)

```

Populating Post_IDX_DS: The sum of the DAYSSUP for Post_IDX_Rx

```

Post_IDX_DS_index <- patient_ACAD_pres %>% group_by(pat_id) %>% summarise(Post_IDX_DS = sum(DAYSSUP))
analytic_file_final <- merge(analytic_file_final, Post_IDX_DS_index, by = 'pat_id', all = TRUE)

```

Populating IDX_PDC: Proportion of days covered for index drug

```

analytic_file_final$IDX_PDC <- analytic_file_final$Post_IDX_DS/180

```

Populating PDC>80

```

analytic_file_final['IDX_PDC>80'] = ifelse(analytic_file_final$IDX_PDC >= 0.8, 1, 0)

```


Populating Post_hosp: Claim for hospitalization after the index date (binary)

A rectype "F" where the to_dt - from_dt > 0, and the to_dt is after the IDX-Date

```
Post_hosp_index <- filter(analytic_file_tmp, rectype == 'F', to_dt - from_dt > 0, to_dt > IDX_Date)
Post_hosp_index <- Post_hosp_index$pat_id
```

```
Post_hosp_index %>% length
```

```
## [1] 1382
```

```
analytic_file_final <- analytic_file_final %>% mutate(Post_hosp = ifelse(pat_id %in% Post_hosp_index, 1
```

Removing patient with missing data (Requested by the professor)

```
for (i in 1:ncol(analytic_file_final)){
  print(which(is.na(analytic_file_final[,i])))
}
```

```
## integer(0)
## integer(0)
## [1] 1282 4556
## integer(0)
## integer(0)
## integer(0)
## integer(0)
## integer(0)
## [1] 1282 4556
## integer(0)
## integer(0)
## integer(0)
## integer(0)
## integer(0)
## integer(0)
## integer(0)
## integer(0)
```

#row 1282 and 4556 have missing vlaue, we would like to remove them

```
analytic_file_final <- analytic_file_final[-c(1282,4556),]
```

Constructing attrition table

```
filter(analytic_file_final, IDX_group == 'AD') %>% dim
```

```
## [1] 4962 16
```

```
filter(analytic_file_final, IDX_group == 'AD' | IDX_group == 'AC' | IDX_group == 'DC') %>% dim
```

```
## [1] 5796 16
```

```
filter(analytic_file_final, IDX_group == 'AD', der_sex == 'M') %>% dim
```

```
## [1] 2404 16
```

```
filter(analytic_file_final, IDX_group == 'AD', age >= 76) %>% dim
```

```
## [1] 363 16
```

```
filter(analytic_file_final, IDX_group == 'AD', age >= 66 & age <= 75) %>% dim
```

```
## [1] 683 16
```

```
filter(analytic_file_final, IDX_group == 'AD', age >= 51 & age <= 65) %>% dim
```

```
## [1] 2420 16
```

```
filter(analytic_file_final, IDX_group == 'AD', age >= 19 & age <= 50) %>% dim
```

```
## [1] 1360 16
```

```
filter(analytic_file_final, IDX_group == 'AD', age > 0 & age <= 18) %>% dim
```

```
## [1] 136 16
```

```
filter(analytic_file_final, IDX_group == 'AC') %>% dim
```

```
## [1] 806 16
```

```
filter(analytic_file_final, IDX_group == 'AC', der_sex == 'M') %>% dim
```

```
## [1] 434 16
```

```
filter(analytic_file_final, IDX_group == 'AC', age >= 76) %>% dim
```

```
## [1] 221 16
```

```
filter(analytic_file_final, IDX_group == 'AC', age >= 66 & age <= 75) %>% dim
```

```
## [1] 148 16
```

```
filter(analytic_file_final, IDX_group == 'AC', age >= 51 & age <= 65) %>% dim
```

```
## [1] 310 16
```

```
filter(analytic_file_final, IDX_group == 'AC', age >= 19 & age <= 50) %>% dim
```

```
## [1] 122 16
```

```
filter(analytic_file_final, IDX_group == 'AC', age > 0 & age <= 18) %>% dim
```

```
## [1] 5 16
```

```
filter(analytic_file_final, IDX_group == 'DC') %>% dim
```

```
## [1] 28 16
```

```
filter(analytic_file_final, IDX_group == 'DC', der_sex == 'M') %>% dim
```

```
## [1] 14 16
```

```
filter(analytic_file_final, IDX_group == 'DC', age >= 76) %>% dim
```

```
## [1] 3 16
```

```
filter(analytic_file_final, IDX_group == 'DC', age >= 66 & age <= 75) %>% dim
```

```
## [1] 5 16
```

```
filter(analytic_file_final, IDX_group == 'DC', age >= 51 & age <= 65) %>% dim
```

```
## [1] 13 16
```

```
filter(analytic_file_final, IDX_group == 'DC', age >= 19 & age <= 50) %>% dim
```

```
## [1] 7 16
```

```
filter(analytic_file_final, IDX_group == 'DC', age > 0 & age <= 18) %>% dim
```

```
## [1] 0 16
```

```
enroll_data$age = 2013 - enroll_data$der_yob  
enroll_data$pat_id %>% length() - length(analytic_file_final$pat_id)
```

```
## [1] 6
```

```
(filter(enroll_data, der_sex == 'M') %>% nrow) - (filter(analytic_file_final, der_sex == 'M') %>% nrow)
```

```
## [1] 2
```

```

filter(enroll_data, age > 0 & age <= 18) %>% nrow - filter(analytic_file_final, age > 0 & age <= 18)

## [1] 0

filter(enroll_data, age >= 19 & age <= 50) %>% nrow - filter(analytic_file_final, age >= 19 & age <= 50)

## [1] 2

filter(enroll_data, age >= 51 & age <= 65) %>% nrow - filter(analytic_file_final, age >= 51 & age <= 65)

## [1] 1

filter(enroll_data, age >= 66 & age <= 75) %>% nrow - filter(analytic_file_final, age >= 66 & age <= 75)

## [1] 1

filter(enroll_data, age >= 76) %>% nrow - filter(analytic_file_final, age >= 76) %>% nrow

## [1] 0

filter(analytic_file_final, IDX_group == 'AD') %>% summarise(summation = sum(Pre_hosp))

##      summation
## 1           282

filter(analytic_file_final, IDX_group == 'AD') %>% summarise(summation = sum(Pre_CHF))

##      summation
## 1           140

filter(analytic_file_final, IDX_group == 'AD') %>% summarise(summation = sum(Pre_B_Blocker))

##      summation
## 1          1226

filter(analytic_file_final, IDX_group == 'AC') %>% summarise(summation = sum(Pre_hosp))

##      summation
## 1           156

filter(analytic_file_final, IDX_group == 'AC') %>% summarise(summation = sum(Pre_CHF))

##      summation
## 1             77

```

```

filter(analytic_file_final, IDX_group == 'AC') %>% summarise(summation = sum(Pre_B_Blocker))

##      summation
## 1          386

filter(analytic_file_final, IDX_group == 'DC') %>% summarise(summation = sum(Pre_hosp))

##      summation
## 1           4

filter(analytic_file_final, IDX_group == 'DC') %>% summarise(summation = sum(Pre_CHF))

##      summation
## 1           0

filter(analytic_file_final, IDX_group == 'DC') %>% summarise(summation = sum(Pre_B_Blocker))

##      summation
## 1          15

filter(analytic_file_final, IDX_group == 'AD') %>% summarise(avg = mean(Post_IDX_Rx, na.rm = TRUE))

##      avg
## 1 6.044538

filter(analytic_file_final, IDX_group == 'AD') %>% summarise(avg = mean(Post_IDX_DS, na.rm = TRUE))

##      avg
## 1 253.7733

filter(analytic_file_final, IDX_group == 'AD') %>% summarise(avg = mean(IDX_PDC, na.rm = TRUE))

##      avg
## 1 1.409852

filter(analytic_file_final, IDX_group == 'AD') %>% summarise(std = sd(Post_IDX_Rx, na.rm = TRUE))

##      std
## 1 4.46851

filter(analytic_file_final, IDX_group == 'AD') %>% summarise(std = sd(Post_IDX_DS, na.rm = TRUE))

##      std
## 1 169.8366

```

```

filter(analytic_file_final, IDX_group == 'AD') %>% summarise(std = sd(IDX_PDC, na.rm = TRUE))

##          std
## 1 0.9435367

filter(analytic_file_final, IDX_group == 'AC') %>% summarise(avg = mean(Post_IDX_Rx, na.rm = TRUE))

##          avg
## 1 5.01737

filter(analytic_file_final, IDX_group == 'AC') %>% summarise(avg = mean(Post_IDX_DS, na.rm = TRUE))

##          avg
## 1 203.7754

filter(analytic_file_final, IDX_group == 'AC') %>% summarise(avg = mean(IDX_PDC, na.rm = TRUE))

##          avg
## 1 1.132086

filter(analytic_file_final, IDX_group == 'AC') %>% summarise(std = sd(Post_IDX_Rx, na.rm = TRUE))

##          std
## 1 3.796003

filter(analytic_file_final, IDX_group == 'AC') %>% summarise(std = sd(Post_IDX_DS, na.rm = TRUE))

##          std
## 1 134.1332

filter(analytic_file_final, IDX_group == 'AC') %>% summarise(std = sd(IDX_PDC, na.rm = TRUE))

##          std
## 1 0.7451847

filter(analytic_file_final, IDX_group == 'DC') %>% summarise(avg = mean(Post_IDX_Rx, na.rm = TRUE))

##          avg
## 1 12.64286

filter(analytic_file_final, IDX_group == 'DC') %>% summarise(avg = mean(Post_IDX_DS, na.rm = TRUE))

##          avg
## 1 460.7857

```

```

filter(analytic_file_final, IDX_group == 'DC') %>% summarise(avg = mean(IDX_PDC, na.rm = TRUE))

##          avg
## 1 2.559921

filter(analytic_file_final, IDX_group == 'DC') %>% summarise(std = sd(Post_IDX_Rx, na.rm = TRUE))

##          std
## 1 5.716864

filter(analytic_file_final, IDX_group == 'DC') %>% summarise(std = sd(Post_IDX_DS, na.rm = TRUE))

##          std
## 1 219.9861

filter(analytic_file_final, IDX_group == 'DC') %>% summarise(std = sd(IDX_PDC, na.rm = TRUE))

##          std
## 1 1.222145

filter(analytic_file_final, IDX_group == 'AD') %>% summarise(summation = sum(`IDX_PDC>80`, na.rm = TRUE))

##    summation
## 1         3771

filter(analytic_file_final, IDX_group == 'AD') %>% summarise(summation = sum(Post_hosp, na.rm = TRUE))

##    summation
## 1         283

filter(analytic_file_final, IDX_group == 'AC') %>% summarise(summation = sum(`IDX_PDC>80`, na.rm = TRUE))

##    summation
## 1         562

filter(analytic_file_final, IDX_group == 'AC') %>% summarise(summation = sum(Post_hosp, na.rm = TRUE))

##    summation
## 1         109

filter(analytic_file_final, IDX_group == 'DC') %>% summarise(summation = sum(`IDX_PDC>80`, na.rm = TRUE))

##    summation
## 1          27

```

```
filter(analytic_file_final, IDX_group == 'DC') %>% summarise(summation = sum(Post_hosp, na.rm = TRUE))
```

```
## summation
## 1 1
```

Attrition Table

| | Variable | IDX_Group | | | | | | | |
|------------|---------------------------|---------------|---------|---------------|--------|--------|--------|---------|-------|
| | | Anti-Diabetic | | Anticoagulant | | Both | | Neither | |
| | | n | % | n | % | n | % | n | % |
| At Index | N | 4962 | | 806 | | 28 | | 4 | |
| | Male (n, %) | 2404 | 48.45 | 434 | 53.85 | 14 | 50.00 | 2 | 50.00 |
| | Age 0-18 at index (n, %) | 136 | 2.74 | 5 | 0.62 | 0 | 0.00 | 0 | 0.00 |
| | Age 19-50 at index (n, %) | 1360 | 27.41 | 122 | 15.14 | 7 | 25.00 | 2 | 50.00 |
| | Age 51-65 at index (n, %) | 2420 | 48.77 | 310 | 38.46 | 13 | 46.43 | 1 | 25.00 |
| | Age 66-75 at index (n, %) | 683 | 13.76 | 148 | 18.36 | 5 | 17.86 | 1 | 25.00 |
| | Age 76+ at index (n, %) | 363 | 7.32 | 221 | 27.42 | 3 | 10.71 | 0 | 0.00 |
| Pre-Index | Pre_hosp (n, %) | 300 | 6.05 | 186 | 23.08 | 4 | 14.29 | | |
| | Pre_CHF (n, %) | 140 | 2.82 | 77 | 9.55 | 0 | 0.00 | | |
| | Pre_B-Blocker (n, %) | 1226 | 24.7078 | 386 | 47.89 | 15 | 53.57 | | |
| Post-Index | Mean SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD |
| | Post_IDX_Rx (mean, SD) | 6.04 | 4.47 | 5.02 | 3.80 | 12.64 | 5.72 | | |
| | Post_IDX_DS (Mean, SD) | 253.77 | 169.84 | 203.78 | 134.13 | 460.79 | 219.99 | | |
| | IDX_PDC (Mean, SD) | 1.41 | 0.94 | 1.13 | 0.75 | 2.56 | 1.22 | | |
| | IDX_PDC_true (Mean, SD) | | | | | | | | |
| | n % | n | % | n | % | n | % | n | % |
| | IDX_PDC>80 (n, %) | 3771 | 76.00 | 562 | 69.73 | 27 | 96.43 | | |
| | Post_hosp (n, %) | 283 | 5.70 | 109 | 13.52 | 1 | 3.57 | | |

Answer Some Questions

```
dim(analytic_file_final)
```

```
## [1] 5796 16
```

```
tmp <- first_occurance %>% group_by(pat_id) %>% summarise(len = length(rx_group), both = length(unique(
filter(tmp, both > 1) %>% nrow
```

```
## [1] 28
```

```
tmp2<-first_occurance %>% group_by(pat_id) %>% summarise(len = length(rx_group))
filter(tmp2, len >1) %>% nrow
```

```
## [1] 1027
```

- How many (unique) patients are in the desired index range that have taken AD or AC drugs?

5796

- How many of those patients have more than one drug on the index date?

1027

- How many have both antidiabetics and anticoagulant drugs on the index date?

28