Data load

Anthony Staines & Anne O'Farrell

2021-04-19

Load files

There are data files for each year from 2013 to October 2020. The rest of 2020 is awaited.

```
D2013 <- readxl::read_excel('data/2013 head injuries.xls') %>% mutate(SOURCE = 'D2013')
D2014 <- readxl::read_excel('data/2014 head injuries.xls') %>% mutate(SOURCE = 'D2014')
D2015 <- readxl::read_excel('data/2015 head injuries.xls') %>% mutate(SOURCE = 'D2015')
D2016 <- readxl::read_excel('data/2016 head injuries.xls') %>% mutate(SOURCE = 'D2016')
D2017 <- readxl::read_excel('data/2017 head injuries.xls') %>% mutate(SOURCE = 'D2017')
D2018_original <- readxl::read_excel('data/2018 head injuries.xls') %>% mutate(SOURCE = 'D2018_original
D2019 <- readxl::read_excel('data/2019 head injuries.xls') %>% mutate(SOURCE = 'D2019')
D2020 <- readxl::read_excel('data/2020_part head injuries.xls') %>% mutate(SOURCE = 'D2020') # Up to Oc

D2018_revised  <- readxl::read_excel('data/2018 Head Injuries_revised.xls') %>%
  anti_join(D2018_original) %>% mutate(SOURCE = 'D2018_revised') # Extra 5073 entries

names(D2019)
#vtable::vt(D2013)

TBI <- D2020 %>% rbind(D2019) %>% rbind(D2018_original) %>% rbind(D2017) %>% rbind(D2016) %>% rbind(D20

nrow(TBI) #95422
nrow(distinct(TBI)) # SAME
nrow(distinct(TBI %>% select(-SOURCE))) # 93346

nrow(distinct(TBI)) -nrow(distinct(TBI %>% select(-SOURCE))) # 2096 from 2019 October - December.

TBI <- TBI %>% rbind(D2018_revised) # Add 5073 missed from 2018 before.

nrow(TBI) #100515
nrow(distinct(TBI)) # SAME
nrow(distinct(TBI %>% select(-SOURCE))) # 98419

nrow(distinct(TBI)) - nrow(distinct(TBI %>% select(-SOURCE))) # 2096 from 2019 October - December No ex

TBI <- TBI %>% distinct(across(1:173), .keep_all = TRUE) #Some dates overlap 2019 and part 2020
names(TBI) # Last one is SOURCE
nrow(TBI) # 98419

rm(D2013,D2014,D2015,D2016,D2017,D2019,D2020)
rm(D2018_original, D2018_revised)
```

Data

Fix the variable names

Fix the variable types

```
table(NAMES$NewName,NAMES$Type)

#Numbers
Numbers <- NAMES %>% filter(Type == 'Numeric') %>% select(NewName)
  COLS <- Numbers$NewName
TBIt <- TBI %>%
  mutate(across(all_of(COLS), as.numeric))
```

```
## Warning in mask$eval_all_mutate(quo): NAs introduced by coercion

## Warning in mask$eval_all_mutate(quo): NAs introduced by coercion

## Warning in mask$eval_all_mutate(quo): NAs introduced by coercion

## Warning in mask$eval_all_mutate(quo): NAs introduced by coercion

## Warning in mask$eval_all_mutate(quo): NAs introduced by coercion

## Warning in mask$eval_all_mutate(quo): NAs introduced by coercion

## Warning in mask$eval_all_mutate(quo): NAs introduced by coercion

## Warning in mask$eval_all_mutate(quo): NAs introduced by coercion

## Warning in mask$eval_all_mutate(quo): NAs introduced by coercion

## Warning in mask$eval_all_mutate(quo): NAs introduced by coercion

## Warning in mask$eval_all_mutate(quo): NAs introduced by coercion

## Warning in mask$eval_all_mutate(quo): NAs introduced by coercion

## Warning in mask$eval_all_mutate(quo): NAs introduced by coercion

## Warning in mask$eval_all_mutate(quo): NAs introduced by coercion
```

```
  rm(Numbers)

#Dates
Dates <- NAMES %>% filter(Type == 'Date') %>% select(NewName)
  COLS <- Dates$NewName
TBIt <- TBIt %>%
  mutate(across(all_of(COLS), as_date))
  rm(Dates)

#vtable::vt(TBIt)
#warnings()

TBI <- TBIt
rm(TBIt, COLS, NAMES)
```

Add new variables

```
#UniqueID
TBI$ID = 1:nrow(TBI)

#Correctly ordered age groups
TBI <- TBI %>%
  mutate(Age_group = as_factor(Age_by_5_year)) %>%
  mutate(Age_group = fct_reorder(Age_group, Age_by_1_year)) %>%
  mutate(Age_group = fct_other(Age_group,
                               drop = c('95-99','100-104','110-114'),
                               other_level='95+')) # Merge top 3 levels into 1

table(TBI$Age_group)
#  #order by median age
#  mutate(Age_group = fct_relevel(Age_group, "5-9", after=1)) #
#  mutate(Age_group = fct_relevel(Age_group, "100-104", after=Inf)) #

TBI %>% select(Age_group) %>% group_by(Age_group) %>% summarise(N = n())
```

```
names(TBI)
```

```
##   [1] "E_MRN"                   "Hospital_name"           "Area_of_residence"
##   [4] "Address_county_ROI"      "Address_Dub_postal"      "Eircode_route_key"
##   [7] "Admission_date"          "Admission_day"           "Admission_week"
##  [10] "Admission_month"         "Admission_year"          "Admission_time"
##  [13] "Discharge_date"          "Discharge_day"           "Discharge_week"
##  [16] "Discharge_month"         "Discharge_year"          "Discharge_time"
##  [19] "Med_fit_discharge_date" "Med_fit_discharge_day"  "Admission_source"
##  [22] "Admission_type"          "Discharge_destination"   "Medical_card"
##  [25] "NTPF"                    "Public_or_private"       "Specialty_pr_group"
##  [28] "Specialty_dis_group"     "Specialty_principal"     "Specialty_discharge"
##  [31] "Team_code_admission"     "Team_code_discharge"     "Team_code_principal"
##  [34] "Team_code_Dx_1"          "Team_code_Dx_2"          "Team_code_Dx_3"
##  [37] "Team_code_Dx_4"          "Team_code_Dx_5"          "Team_code_Dx_6"
##  [40] "Team_code_Dx_7"          "Team_code_Dx_8"          "Team_code_Dx_9"
##  [43] "Team_code_Dx_10"         "Team_code_Dx_11"         "Team_code_Dx_12"
##  [46] "Team_code_Dx_13"         "Team_code_Dx_14"         "Team_code_Dx_15"
##  [49] "Team_code_Dx_16"         "Team_code_Dx_17"         "Team_code_Dx_18"
##  [52] "Team_code_Dx_19"         "Team_code_Dx_20"         "Team_code_Dx_21"
##  [55] "Team_code_Dx_22"         "Team_code_Dx_23"         "Team_code_Dx_24"
##  [58] "Team_code_Dx_25"         "Team_code_Dx_26"         "Team_code_Dx_27"
##  [61] "Team_code_Dx_28"         "Team_code_Dx_29"         "Team_code_Dx_30"
##  [64] "Team_code_Proc_1"        "Team_code_Proc_2-20"     "Transfer_from"
##  [67] "Transfer_to"             "Ward_AAU"                "Ward_admit"
##  [70] "Ward_discharge"          "Age_by_1_year"           "Age_by_5_year"
##  [73] "Gender"                  "Dx_1_group_CCS_IM"       "Dx_1_name_CCS_IM"
##  [76] "Dx_1_ICD_name"           "Dx_2_ICD_name"           "Dx_3_ICD_name"
##  [79] "Dx_4_ICD_name"           "Dx_5_ICD_name"           "Dx_6_ICD_name"
##  [82] "Dx_7_ICD_name"           "Dx_8_ICD_name"           "Dx_9_ICD_name"
##  [85] "Dx_10_ICD_name"          "Dx_11_ICD_name"          "Dx_12_ICD_name"
##  [88] "Dx_13_ICD_name"          "Dx_14_ICD_name"          "Dx_15_ICD_name"
##  [91] "Dx_16_ICD_name"          "Dx_17_ICD_name"          "Dx_18_ICD_name"
##  [94] "Dx_19_ICD_name"          "Dx_20_ICD_name"          "Dx_21_ICD_name"
```

```
##  [97] "Dx_22_ICD_name"         "Dx_23_ICD_name"         "Dx_24_ICD_name"
## [100] "Dx_25_ICD_name"         "Dx_26_ICD_name"         "Dx_27_ICD_name"
## [103] "Dx_28_ICD_name"         "Dx_29_ICD_name"         "Dx_30_ICD_name"
## [106] "Dx_1_ICD_code"          "Dx_2_ICD_code"          "Dx_3_ICD_code"
## [109] "Dx_4_ICD_code"          "Dx_5_ICD_code"          "Dx_6_ICD_code"
## [112] "Dx_7_ICD_code"          "Dx_8_ICD_code"          "Dx_9_ICD_code"
## [115] "Dx_10_ICD_code"         "Dx_11_ICD_code"         "Dx_12_ICD_code"
## [118] "Dx_13_ICD_code"         "Dx_14_ICD_code"         "Dx_15_ICD_code"
## [121] "Dx_16_ICD_code"         "Dx_17_ICD_code"         "Dx_18_ICD_code"
## [124] "Dx_19_ICD_code"         "Dx_20_ICD_code"         "Dx_21_ICD_code"
## [127] "Dx_22_ICD_code"         "Dx_23_ICD_code"         "Dx_24_ICD_code"
## [130] "Dx_25_ICD_code"         "Dx_26_ICD_code"         "Dx_27_ICD_code"
## [133] "Dx_28_ICD_code"         "Dx_29_ICD_code"         "Dx_30_ICD_code"
## [136] "Proc_1_ACHI_name"       "Proc_2_20_ACHI_name"    "Proc_1_group_RCs"
## [139] "Proc_1_name_RCs"        "Proc_1_surg"            "Proc_1_ACHI_code"
## [142] "Proc_2_20_ACHI_code"    "Proc_1_date"            "Proc_1_day"
## [145] "Proc_2_20_dates"        "DRG_name"               "DRG_code"
## [148] "Hospital_acq_code"      "Hospital_acq_name"      "ASA_score_proc_1"
## [151] "ASA_score_all"          "Charlson_score_group"   "Charlson_score_value"
## [154] "Discharge_alive_dead"   "Emerg_adm_12m"          "MAIS_score"
## [157] "Palliative_care"        "Sepsis"                 "VTE"
## [160] "COVID_19"               "LOS_total"              "LOS_pre_proc_1"
## [163] "LOS_post_proc_1"        "LOS_pre_med_fit_dis"    "LOS_post_med_fit_dis"
## [166] "ITU_or_CCU_bed_days"    "DOSA"                   "Discharge_same_day"
## [169] "Elective_day_case"      "Statistical_day_case"   "Readm_7d"
## [172] "Readm_30d"              "Readm_info"             "SOURCE"
## [175] "ID"                     "Age_group"
```

```
table(TBI$SOURCE)
```

```
##
##          D2013          D2014          D2015          D2016          D2017
##           7064          10701          11411          11826          15909
## D2018_original  D2018_revised          D2019          D2020
##          11676           5073          10703          14056
```

```
head(TBI$ID)
```

```
## [1] 1 2 3 4 5 6
```

```
#Lots of Repeated MRNs
TBI %>% group_by(E_MRN) %>% summarise(Count = n()) %>% filter(Count > 1) %>% summarise(N=n())
```

```
## # A tibble: 1 x 1
##       N
##   <int>
## 1  5740
```

```
TBI %>% group_by(SOURCE,E_MRN) %>% summarise(Count = n()) %>% filter(Count > 1) %>% summarise(N=n())
```

```
## # A tibble: 9 x 2
##   SOURCE             N
##   <chr>          <int>
## 1 D2013            219
## 2 D2014            336
## 3 D2015            362
## 4 D2016            390
```

```
## 5 D2017            660
## 6 D2018_original   400
## 7 D2018_revised    152
## 8 D2019            387
## 9 D2020            584
```

```
TBI %>% group_by(ID) %>% summarise(Count = n()) %>% filter(Count > 1) %>% summarise(N=n())
```

```
## # A tibble: 1 x 1
##       N
##   <int>
## 1     0
```

```
TBI <- TBI %>% dplyr::select(ID,SOURCE,E_MRN:Readm_info,Age_group)
```

Basic tables

```
Gender <- gt(data=TBI %>%
               select(Gender) %>%
               group_by(Gender) %>%
               summarise(N = n())
             ) %>%
  tab_header(
    title = "Gender",
    subtitle = "Only Male and Female are recorded"
  )
Gender
```

Gender

Only Male and Female are recorded

Gender	N
Female	37837
Male	60582

```
Age <- gt(data=TBI %>%
            select(Age_group) %>%
            group_by(Age_group) %>%
            summarise(N = n())
             ) %>%
  tab_header(
    title = "Age (5 year age groups)",
    subtitle = "All S00-S09 admissions"
  )
Age
```

Age (5 year age groups)

All S00-S09 admissions

Age_group	N
0-4	12161
5-9	5424
10-14	4395
15-19	6121

20-24	5687
25-29	4759
30-34	4394
35-39	4125
40-44	3648
45-49	3564
50-54	3667
55-59	3849
60-64	3852
65-69	4293
70-74	5271
75-79	6281
80-84	7136
85-89	6122
90-94	2904
95+	766

```
Source <- gt(data=TBI %>%
             select(Admission_source) %>%
             group_by(Admission_source) %>%
             summarise(N = n())
              ) %>%
  tab_header(
    title = "Source of admission",
    subtitle = "All S00-S09 admissions"
  )
Source
```

Source of admission

All S00-S09 admissions

Admission_source	N
Home	90089
New born	31
Other	42
Prison	84
Temporary place of residence	143
Transfer from Acute Hospital	5607
Transfer from hospice not in HIPE hospital listing	1
Transfer from Non-Acute Hospital not in HIPE hospital listing	80
Transfer from nursing home/convalescent home or other long stay accommodation	2264
Transfer from psychiatric hospital/unit	78

```
Type <- gt(data=TBI %>%
             select(Admission_type) %>%
             group_by(Admission_type) %>%
             summarise(N = n())
              ) %>%
  tab_header(
    title = "Type of admission",
    subtitle = "All S00-S09 admissions"
  )
Type
```

Type of admission
All S00-S09 admissions

Admission_type	N
Elective	9736
Elective Readmission	747
Emergency	86963
Emergency Readmission	435
Maternity	195
New born	343

```
Destination <- gt(data=TBI %>%
            select(Discharge_destination) %>%
            group_by(Discharge_destination) %>%
            summarise(N = n())
             ) %>%
  tab_header(
    title = "Discharge destination",
    subtitle = "All S00-S09 admissions"
  )
Destination
```

Discharge destination
All S00-S09 admissions

Discharge_destination	N
Absconded	360
Died no post mortem	1872
Died with post mortem	766
Home	79735
Hospice (not in HIPE Hospital Listing)	132
Nursing home, convalescent home or long stay accommodation	7084
Other (e.g. Foster care)	150
Prison	129
Self discharge	1722
Temporary place of residence (e.g. hotel)	191
Transfer to external rehabilitation facility (not in HIPE Hospital Listing)	535
Transfer to Hospital - Emergency	1573
Transfer to Hospital - Non Emergency	3847
Transfer to Non-Acute Hospital not in HIPE Hospital Listing - Emergency	8
Transfer to Non-Acute Hospital not in HIPE Hospital Listing - Non Emergency	134
Transfer to psychiatric hospital/unit	181

```
Group.db <- TBI %>%
  select(Medical_card, NTPF, Public_or_private) %>%
  group_by(Medical_card, NTPF, Public_or_private) %>%
  summarise(N=n())

Group.db %>%
  kbl() %>%
  kable_classic(full_width = F, html_font = "Cambria")
```

```
rm(Age, Destination, Gender, Group.db, Source, Type)
```

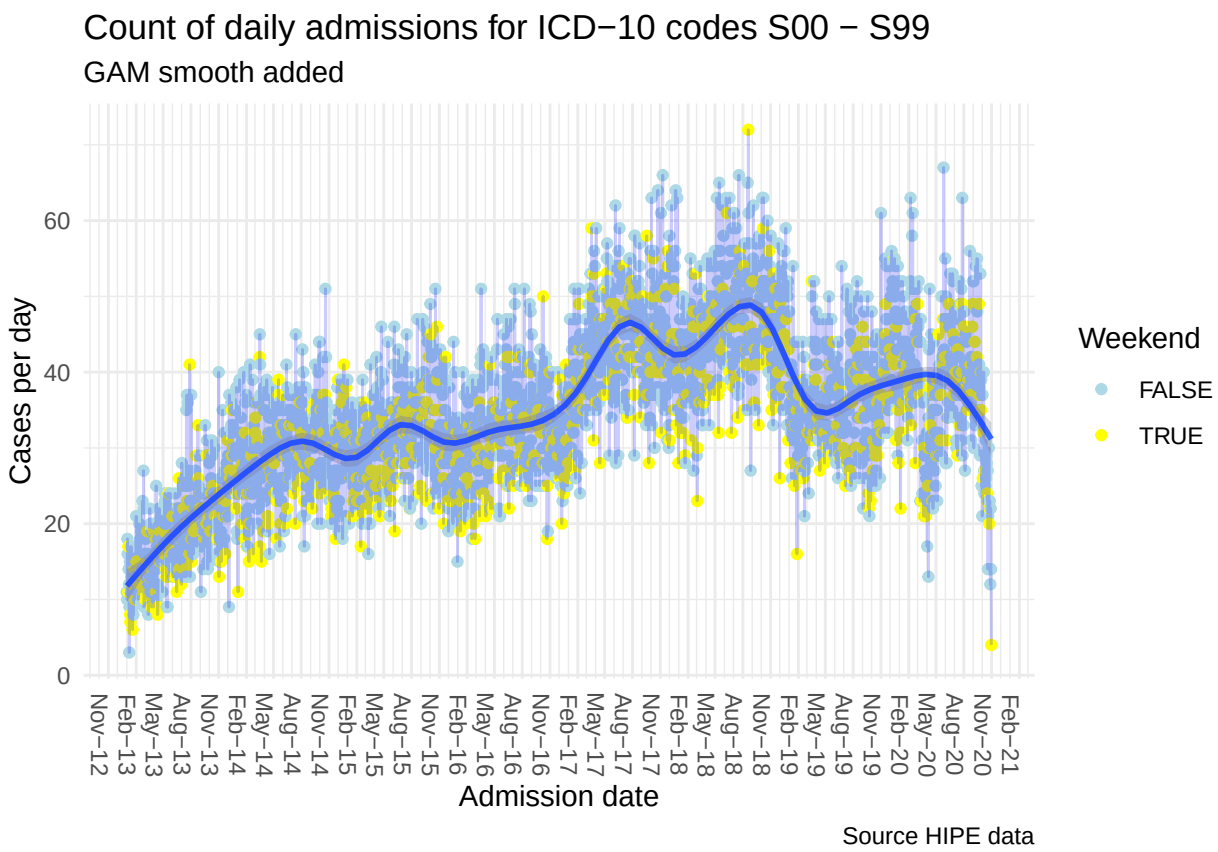
Medical_card	NTPF	Public_or_private	N
No	No	Private	13501
No	No	Public	35751
No	Yes	Public	27
Unknown	No	Private	31
Unknown	No	Public	162
Unknown	Yes	Public	1
Yes	No	Private	3178
Yes	No	Public	45761
Yes	Yes	Public	7

Basic plots

```
Daily_Admissions <- TBI %>%
  select(Admission_date) %>%
  group_by(Admission_date) %>%
  summarise(N = n()) %>%
  mutate(Day_of_week = wday(Admission_date,
                            label = TRUE)) %>%
  mutate(Weekend = ifelse((Day_of_week %in% c('Sat','Sun')), TRUE, FALSE))

str(Daily_Admissions)

ggplot(Daily_Admissions, aes(x=Admission_date, y=N)) +
  geom_point(aes(colour=Weekend)) +
  geom_smooth( method = "gam", formula = y ~ s(x, k=20, bs = "cs"))  +
  geom_line(alpha=0.2,colour='blue') +
  scale_colour_manual(values=c('lightblue','yellow')) +
  scale_x_date(date_minor_breaks='1 month',
               date_breaks = '3 months',
               date_labels = '%b-%y') +
  labs(title = 'Count of daily admissions for ICD-10 codes S00 - S99',
       subtitle = 'GAM smooth added',
       x = 'Admission date',
       y = 'Cases per day',
       caption='Source HIPE data') +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = -90))
```




```
rm(Daily_Admissions)
```

This is all head injuries with codes in S00 to S99.

Dictionary of codes

This is set of all the ICD10 codes used in the data which were associated with names. Some of these are blank and were filled in from where they were given in the dataset, some completed from the US codes and a few were completed manually.

```
Dictionary <- TBI %>%
  select(ID,
         Dx_1_ICD_code:Dx_30_ICD_code,
         Dx_1_ICD_name:Dx_30_ICD_name) %>%  #All diagnoses
  pivot_longer(-ID,
               names_to=c("Dx",".value"),
               names_pattern="(Dx_\\d+)_(.*)") %>% # Generate one row per diagnosis
  filter(!is.na(ICD_code)) %>% # Lose blank codes
  distinct(ICD_code,ICD_name) %>%
  mutate(Name_missing = is.na(ICD_name)) %>%
  mutate(Data_source = ifelse(Name_missing, NA, 'HIPE'))
```

ICD-10 Codes used

Codes used for diagnosis 1 only

```
Codes_used_Dx_1 <- TBI %>%
  select(ID,Dx_1_ICD_code) %>%  #Principal diagnosis
  pivot_longer(-ID,
               names_to=c("Dx","Type"),
               names_pattern="(Dx_\\d+)_(.*)",
               values_to="Code") %>% # Generate one row per diagnosis
  select(-Type) %>% # This just says ICD_code
  filter(!is.na(Code)) %>% # Lose blank codes
  left_join(Dictionary, by = c('Code' = 'ICD_code')) %>%
  group_by(Code, ICD_name) %>% # Group by code
  summarize(N=n()) %>% # Count every code for principal diagnosis
  arrange(desc(N)) # Put from most common to least common

pander::pander(head(Codes_used_Dx_1,n=10))
```

Code	ICD_name	N
S099	Unspecified injury of head	11610
S0188	Open wound of other parts of head	6258
S022	Fracture of nasal bones	6088
S098	Other specified injuries of head	5034
S065	Traumatic subdural haemorrhage	4074
S0602	LOC brief dur [less than 30 minutes]	3777
R55	Syncope and collapse	3244
S010	Open wound of scalp	2635
S0151	Open wound of lip	2580
S024	Fracture of malar and maxillary bones	2334

The commonest code used in the first diagnosis was 'Unspecified injury of head', with 'Open wound of other parts of the head', 'Fracture of nasal bones' and 'Other specified injuries of head' roughly equally common.

Codes used for all diagnoses

```
Codes_used_all_Dx <- TBI %>%
  select(ID,Dx_1_ICD_code:Dx_30_ICD_code) %>%  #All diagnoses
  pivot_longer(-ID,
               names_to=c("Dx","Type"),
               names_pattern="(Dx_\\d+)_(.*)",
               values_to="Code") %>% # Generate one row per diagnosis
  select(-Type) %>% # This just says ICD_code
  filter(!is.na(Code)) %>% # Lose blank codes
  left_join(Dictionary, by = c('Code' = 'ICD_code')) %>%
  group_by(Code,ICD_name) %>% # Group by code
  summarize(N=n()) %>% # Count every code used for any diagnosis
  arrange(desc(N)) # Put from most common to least common

pander::pander(head(Codes_used_all_Dx,n=10))
```

Code	ICD_name	N
U739	Unspecified activity	65543
Y929	Unspecified place of occurrence	38942
W19	Unspecified fall	21992
Y9209	Other and unspecified place in home	17923
S099	Unspecified injury of head	14533
S0188	Open wound of other parts of head	12992
Z720	Tobacco use current	8980
U738	Other specified activity	8955
S022	Fracture of nasal bones	7975
U732	While rest sleep eat engaging vtl act	7436

Across all diagnoses, the commonest codes used were 'Unspecified activity', 'Unspecified place of occurrence' and 'Unspecified fall', suggesting that in many medical records these data were not given.

Codes used and the number of the associated diagnosis

```
Codes_used_by_Dx <- TBI %>%
  select(ID,Dx_1_ICD_code:Dx_30_ICD_code) %>%  #All diagnoses
  pivot_longer(-ID,
               names_to=c("Dx","Type"),
               names_pattern="(Dx_\\d+)_(.*)",
               values_to="Code") %>% # Generate one row per diagnosis
  select(-Type) %>% # This just says ICD_code
  filter(!is.na(Code)) %>% # Lose blank codes
  left_join(Dictionary, by = c('Code' = 'ICD_code')) %>%
  group_by(Dx,Code,ICD_name) %>% # Group by diagnosis, and code
  summarize(N=n()) %>% # Count every combination of diagnosis and code
  arrange(desc(N)) # Put from most common to least common

pander::pander(head(Codes_used_by_Dx, n=10))
```

Dx	Code	ICD_name	N
Dx_4	U739	Unspecified activity	23831
Dx_5	U739	Unspecified activity	14324
Dx_4	Y929	Unspecified place of occurrence	14001
Dx_3	Y929	Unspecified place of occurrence	12699
Dx_1	S099	Unspecified injury of head	11610
Dx_3	U739	Unspecified activity	8525
Dx_2	W19	Unspecified fall	7592
Dx_6	U739	Unspecified activity	7209
Dx_1	S0188	Open wound of other parts of head	6258
Dx_1	S022	Fracture of nasal bones	6088

This analysis by diagnosis number and code suggests that these very common codes were, as would be expected, subsidiary codes. Only 'Unspecified injury of head' 'Open wound of other parts of head' and 'Fracture of nasal bones' were common first diagnoses in this set.

```
rm(Codes_used_all_Dx, Codes_used_by_Dx, Codes_used_Dx_1)
```

Working file

Pivot TBI to long - one row per non-missing diagnosis

Each row is one diagnosis for one HIPE episode. Patients with many diagnoses have many rows.

```
TBI_l <- TBI %>%
  select(ID,E_MRN,SOURCE,Hospital_name,starts_with('Dx')) %>%
  pivot_longer(cols = c(Dx_1_ICD_code:Dx_30_ICD_code,
         Dx_1_ICD_name:Dx_30_ICD_name),
               names_to=c("Dx",".value"),
               names_pattern="(Dx_\\d+)_(.*)") %>%
  filter(!is.na(ICD_code))
```

Save the data files for further work

```
saveRDS(Dictionary, file='data/Dictionary.Rds')
saveRDS(TBI, file='data/TBI.Rds')
saveRDS(TBI_l, file='data/TBI_l.Rds')
```