Import AQI data

Michael Andreae

Friday, June 19, 2015

Insurance status predicts antiemetic use

We investigate the Hypothesis that insurance status predicets antiemetic use the population in the Public Use File of the Anestehsia Quality Institute with electronic anesthesia records recording antiemetic use.

Data import

we load the original dataset and save as it as *PUF_Q4_2013.Rdata*

Original Data

```
# run only once
PUF_Q4_2013 <- read.csv("Analysis/Data/PUF_Q4_2013_Antimetic.csv")
save(PUF_Q4_2013, file="Analysis/Data/PUF_Q4_2013.Rdata")
```

Load Rdata AQI raw dataset ***PUF_Q4_2013.Rdata***

```
rm(list = ls())
load("Analysis/Data/PUF_Q4_2013.Rdata")
```

Clean Data

Predictor: insurance status

The predictor insurance status (*Payment*) is coded in 4 levels as Commercial, MEDICAID, Medicare, SELF, after we removed 265311 cases without insurance information, (orginally coded as “”), with 176334 unique cases remaining. (False indicating no NAs in Payment: FALSE)

Outcome variables: antiemetic administration

We focus on the antiemetics *ondansetron*, *dexamethason* and *droperidol*, the only agents with convincing evidence for effect.

Table 1: Cases with Ondansetron versus Dexamethason

	no Dex	Dex
no Ondan	79842	4873
Ondan	61191	30428

Table 2: Cases with Ondansetron versus Droperidol

	no Drope	Drope
no Ondan	84398	317
Ondan	89067	2552

Table 3: Cases with Dexamethason versus Droperidol

	no Drope	Drope
no Dex	140010	1023
Dex	33455	1846

The antiemetics *ondansetron* and *dexamethason* were sometimes administered together. This is coded in *ondan_dex_either*

Potential confounders and other variables

practice ID versus facility ID

	193055	691419	5013437	5610264
136085	0	23241	0	0
1116623	36127	0	0	0
9485541	0	0	0	94080
23100212	0	0	79	0
46100453	0	0	15040	0
53228659	0	0	1	0
71100339	0	0	7766	0

The table of facility ID versus practice ID suggests that five practices have only one facility ID and one practice (=5013437) has three (sub) facilities. We will simplify by using practice ID, which has no NA.

case_duration_minutes

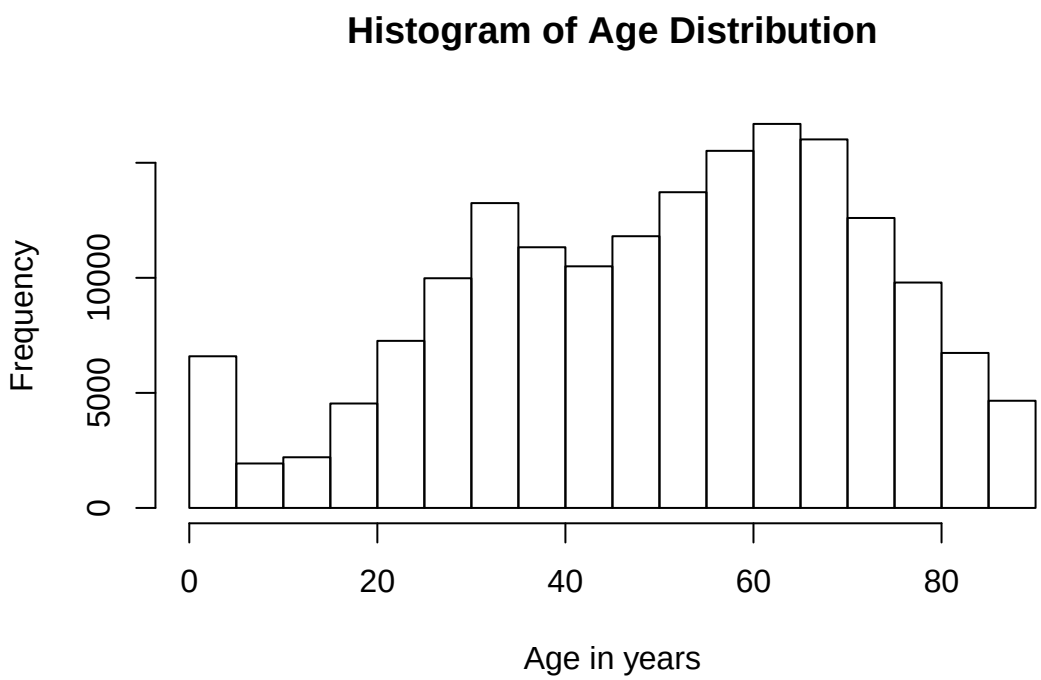
```
PUF_Q4_2013$case_duration_minutes [PUF_Q4_2013$case_duration_minutes == -1] <- NA
missing <- sum(PUF_Q4_2013$case_duration_minutes==-1)
PUF_Q4_2013 <- PUF_Q4_2013 [complete.cases(PUF_Q4_2013$case_duration_minutes),]
```

Case duration in minutes (*case_duration_minutes*) is an integer and has NA missing values coded as -1, which we recoded as NA and removed from the dataset, leaving 175123 unique cases.

patient age

```
PUF_Q4_2013$patient_age[PUF_Q4_2013$patient_age==-1] <- NA
hist(PUF_Q4_2013$patient_age,
```

```
     main = "Histogram of Age Distribution",
     xlab = "Age in years")
```




```
missing <- sum(patient_age==-1)
PUF_Q4_2013 <- PUF_Q4_2013[complete.cases(PUF_Q4_2013$patient_age),]
```

Patient age (*patient_age*) is an integer with a distribution above and has 8 missing values coded as -1, which we recoded as NA and removed from the dataset, leaving 175115 unique cases.

patient_age_group

```
levels(PUF_Q4_2013$patient_age_group)[2] <- "1-18"
levels(PUF_Q4_2013$patient_age_group)[1] <- NA
missing <- sum(is.na(PUF_Q4_2013$patient_age_group))
which(is.na(PUF_Q4_2013$patient_age_group))
```

```
## integer(0)
```

```
PUF_Q4_2013 <- PUF_Q4_2013[complete.cases(PUF_Q4_2013$patient_age_group),]
```

Patient age group (*patient_age_group*) is a factor with 6 levels: 1-18, 19 - 49, 50 - 64, 65 - 79, 80+, Under 1; it has 0 missing values, leaving 175115 unique cases.

(Missing values were initially coded as -1, which we recoded as NA and removed as a level; we corrected the miscoding from “18-Jan”" to “1-18”).

patient_sex

Patient gender (*patient_sex*) is recoded as factor with the two levels female, male and 18 NAs, which are removed from the dataset, leaving 175097 unique cases.

in_or_out_patient

in- or outpatient status (*in_or_out_patient*) is recoded as a factor with the two levels Outpatient, Inpatient and 62432 NAs, which are too numerous to include this variable in the dataset.

Save cleaned dataset

AQI_4_14.Rdata saved as a clean dataframe

```
## 'data.frame':    175097 obs. of  10 variables:
##  $ Payment              : Factor w/ 4 levels "Commercial","MEDICAID",..: 1 1 1 3 1 3 2 1 3 1 ...
##  $ ondansetron          : Factor w/ 2 levels "no Ondan","Ondan": 2 1 2 1 1 2 2 2 1 2 ...
##  $ dexamethason         : Factor w/ 2 levels "no Dex","Dex": 1 1 1 1 1 1 1 1 1 1 ...
##  $ droperidol           : Factor w/ 2 levels "no Drope","Drope": 1 1 1 1 1 1 1 1 1 1 ...
##  $ ondan_dex_either     : Factor w/ 2 levels "neither","either": 2 1 2 1 1 2 2 2 1 2 ...
##  $ practiceID           : Factor w/ 4 levels "193055","691419",..: 2 2 2 2 2 2 2 2 2 2 ...
##  $ case_duration_minutes: int  59 43 190 56 37 116 93 108 70 93 ...
##  $ patient_age          : int  50 53 58 73 64 73 19 27 85 59 ...
##  $ patient_age_group    : Factor w/ 6 levels "1-18","19 - 49",..: 3 3 3 4 3 4 2 2 5 3 ...
##  $ patient_sex          : Factor w/ 2 levels "female","male": 2 2 1 2 2 1 1 1 2 2 ...
```