

~ / NIAMS / Ward / USRDS2015 - master - RStudio

Go to file / function

Addins

teTo2015.R × DataMunging.R × 2014_15_report.R × medevid.R × discontinued.R × hosp >>

Run Source

```
#' ---
#' title: "Inconsistencies between 2014 and 2015 data"
#' author: "Abhijit Dasgupta"
#' output: html_document
#' ---
#'
#' # Background
#'
#' We want to update 2014 data on which we generated competing risk models for death while
#' on dialysis with 2015 data that should update the timelines of existing individuals
#' in our study. However, there appear to be several discrepant observations between the
#' 2014 and 2015 datasets.
#'
#+ preamble, echo=FALSE, cache=TRUE, message=FALSE
library(tidyverse)
ProjTemplate::reload()
dbdir <- verifyPaths()
dbdir14 <- verifyPaths(2014)
con <- dbConnect(SQLite(), file.path(dbdir, 'USRDS.sqlite3'))
analytic_data <- readRDS('data/rda/analytic14.rds')

P1 <- tbl(con, 'StudyIDs') %>% left_join(tbl(con, 'patients')) %>%
  select(USRDS_ID, FIRST_SE, TX1DATE, DIED) %>% collect(n=Inf)
P1 <- P1 %>%
  mutate_at(vars(FIRST_SE:DIED), as.Date)
analytic_data <- analytic_data %>%
  mutate_at(vars(FIRST_SE, TX1DATE, DIED), as.Date, origin = '1960-01-01')
dbDisconnect(con)

d <- analytic_data %>% select(USRDS_ID, FIRST_SE, TX1DATE, DIED) %>%
  rename(FIRST_SE14 = FIRST_SE, TX1DATE14 = TX1DATE, DIED14 = DIED) %>%
  left_join(P1 %>% rename(FIRST_SE15 = FIRST_SE, TX1DATE15 = TX1DATE, DIED15 = DIED))
#'
#' We find that there are `r sum(is.na(P1$FIRST_SE))` individuals who have FIRST_SE missing
#' in the 2015 data who had this value in the 2014 data. Moreover, we see some disagreement
#' in the values of this variable, as well. `r sum(d$FIRST_SE15 != d$FIRST_SE14, na.rm=T)`
#' individuals have discrepant FIRST_SE observations, and
#' `r sum(d$FIRST_SE15 < '2003-01-01', na.rm=T)`
#' individuals now have FIRST_SE values before 2013, which was the cutoff used in the 2014
#' analysis.
```

33:1 (Top Level) ↕ R Script ↕

Environment

History

Connections

Git

Console

Files

Plots

Packages

Help

Viewer

Publish

Inconsistencies between 2014 and 2015 data

Abhijit Dasgupta

Tue Nov 28 09:31:31 2017

Background

We want to update 2014 data on which we generated competing risk models for death while on dialysis with 2015 data that should update the timelines of existing individuals in our study. However, there appear to be several discrepant observations between the 2014 and 2015 datasets.


We find that there are 395 individuals who have FIRST_SE missing in the 2015 data who had this value in the 2014 data. Moreover, we see some disagreement in the values of this variable, as well. 4094 individuals have discrepant FIRST_SE observations, and 17 individuals now have FIRST_SE values before 2013, which was the cutoff used in the 2014 analysis.













**We'll talk
about this
next week**

We'll talk about this next week

teTo2015.R × DataMunging.R × 2014_15_report.R × medevid.R × discontinued.R × hosp >> 

   ☐ Source on Save    Run  Source 

```
#' ---
#' title: "Inconsistencies between 2014 and 2015 data"
#' author: "Abhijit Dasgupta"
#' output: html_document
#' ---
#'
#' # Background
#'
#' We want to update 2014 data on which we generated competing risk models for death while
#' on dialysis with 2015 data that should update the timelines of existing individuals
#' in our study. However, there appear to be several discrepant observations between the
#' 2014 and 2015 datasets.
#'
#+ preamble, echo=FALSE, cache=TRUE, message=FALSE
library(tidyverse)
ProjTemplate::reload()
dbdir <- verifyPaths()
dbdir14 <- verifyPaths(2014)
con <- dbConnect(SQLite(), file.path(dbdir, 'USRDS.sqlite3'))
analytic_data <- readRDS('data/rda/analytic14.rds')





P1 <- tbl(con, 'StudyIDs') %>% left_join(tbl(con, 'patients')) %>%
  select(USRDS_ID, FIRST_SE, TX1DATE, DIED) %>% collect(n=Inf)
P1 <- P1 %>%
  mutate_at(vars(FIRST_SE:DIED), as.Date)
analytic_data <- analytic_data %>%
  mutate_at(vars(FIRST_SE, TX1DATE, DIED), as.Date, origin = '1960-01-01')
dbDisconnect(con)

d <- analytic_data %>% select(USRDS_ID, FIRST_SE, TX1DATE, DIED) %>%
  rename(FIRST_SE14 = FIRST_SE, TX1DATE14 = TX1DATE, DIED14 = DIED) %>%
  left_join(P1 %>% rename(FIRST_SE15 = FIRST_SE, TX1DATE15 = TX1DATE, DIED15 = DIED))
#'
```

33:1 (Top Level) ↕ R Script ↕

Console

Files Plots Packages Help Viewer

   Publish 

Inconsistencies between 2014 and 2015 data

Abhijit Dasgupta
Tue Nov 28 09:31:31 2017

Background

We want to update 2014 data on which we generated competing risk models for death while on dialysis with 2015 data that should update the timelines of existing individuals in our study. However, there appear to be several discrepant observations between the 2014 and 2015 datasets.

We find that there are 395 individuals who have FIRST_SE missing in the 2015 data who had this value in the 2014 data. Moreover, we see some disagreement in the values of this variable, as well. 4094 individuals have discrepant FIRST_SE observations, and 17 individuals now have FIRST_SE values before 2013, which was the cutoff used in the 2014 analysis.

Weaving documents programmatically

