

R Coding for State Injury Report

2014 TBI Death in Washington State

Mamadou Ndiaye - Washington State Department of Health

August 26, 2016

Abstract

Steps to provide count of traumatic brain injury (TBI) deaths using R: (1) Define the injury with regular expressions, (2) Specify the columns with the icd-10 codes of interest, (3) Use a predefined function to match the regular expression patterns with the icd10 codes, (4) Create a new variable assigning '1' to observation matching the defined regular expression in any of the selected field, assigning '0' otherwise. (5) Make the tables of counts after the new variable for the defined injury is created.

Introduction

R libraries and custom functions

The R libraries **dplyr**, **tidyr** and **Hmisc** were used in this analysis. They can be loaded with the following codes (assuming they are already installed with the function *"install.packages"*):

```
library(dplyr)
library(tidyr)
library(Hmisc)
```

The functions *"create.diag"* and *"create_cond_diag"* can be loaded from my github page with:

```
source("https://raw.githubusercontent.com/epinotes/InjuryEpi/master/create_diag")
source("https://raw.githubusercontent.com/epinotes/InjuryEpi/master/create_cond_diag")
source("https://raw.githubusercontent.com/epinotes/InjuryEpi/master/onclip")
```

The Death dataset

The variables used in the death dataset (**dea2014f**) were:

- The diagnosis code columns: *underly*, *mltcse1* to *mltcse20*
- The age group column (11 age groups): *agegrp11*
- The sex column: *sex*

Injury Definitions with Regular Expressions

all injuries

Definition from CDC: The ICD-10 codes **V01–Y36**, **Y85–Y87**, **Y89**, **U01–U03** in the underlying cause of death field. They are re-written as regular expressions in the following that save them in the R object **.cdc_inj**.

```
.cdc_inj <- "~V[0-9][1-9]|~[W-X][0-9][0-9]|~Y[0-2][0-9]|~Y3[0-6]|~Y8[5-7|9]|U0[1-3]"
```

Select the column of interest (the underlying cause of death column):

```
colx <- grep("^underly$", names(dea2014f))
```

Create the variable for any injury, *cdc_inj*:

```
dea2014f$cdc_inj <- new.diag_m(data=dea2014f, expr=.cdc_inj, colvec= colx)
```

Check the results with the function *describe()* (get information on the function by running *?Hmisc::describe*)

```
describe(dea2014f$cdc_inj)
```

Show the requested CDC table of count of injury deaths by the 11 age groups (assigned to the object *dea2014f_inj*):

```
dea2014f_inj <- dea2014f %>% filter(cdc_inj == 1) %>% group_by(agegrp11) %>% summarise(deaths = n()) %>%  
print(dea2014f_inj)
```

we can copy the table in a clipboard then paste it on the CDC injury table templates:

```
onclip(dea2014f_inj)
```

TBI Injuries

Once a case is defined as injury (as previously done); it is TBI if any field of the multiple cause of death file has one of these ICD-10 code:

S01.0–S01.9, S02.0, S02.1, S02.3, S02.7–S02.9, S04.0, S06.0–S06.9, S07.0, S07.1, S07.8, S07.9, S09.7–S09.9, T01.0, T02.0, T04.0, T06.0, T90.1, T90.2, T90.4, T90.5, T90.8, T90.9

or re-written as a regular expression:

```
.cdc_tbi <- "S01[0-9]|S02[0|1|3|7|8|9]|S040|S06[0-9]|S07[0|1|8|9]|S09[7-9]|T0[1|2|4|6]0|T90[1|2|4|5|8|9]"
```

Then we select the underlying cause column (“underly”) and the 20 multiple cause columns (“mltcsi1”, “mltcsi20”):

```
col_m <- grep("underly$|mltcsi", names(dea2014f), value = F)
```

Create the TBI variable

```
dea2014f <- dea2014f %>% mutate(cdc_tbi = create_cond_diag(., expr=.cdc_tbi, colvec= col_m, cond.var = c(
```

Check the results:

```
describe(dea2014f$cdc_tbi)
```

Select the subset of TBI data subset for the other tables:

```
sel_inj <- c(1, 3, 51, 52, grep("underly$|ecode|mltcse", names(dea2014f ), value = F), 169:180)  
dea2014_tbi <- dea2014f[sel_inj] %>% filter(cdc_tbi == 1)
```

More

[Injury Epi GitHub](#)