Covid Staging Data

Steph Reynolds (Stephanie.Reynolds@ucsf.edu)

Dec 15, 2021 09:22 AM

Background

Project

MINDSCAPE: Modeling of infectious network dynamics for surveillance, control and prevention enhancement

Description

This file imports `bin_clin_scen_df_11.08.2021.csv` which contains all binary clinical scenarios and O2 devices for all patients, and runs a function to generate WHO Clinical Stages of Severity (4-10).

SourceData

- Binary Clinical Scenarios and O2 Devices (`bin_clin_scen_df_11.08.2021.csv`)
  - This file contains data on patients' O2 flow rates, clinical scenarios, receipt of O2 devices and their respective category of respiratory support.

Load required packages

```
library(here)
library(tidyverse)
library(tableone)
```

Import and preview data

```
bin_clin_scen_df <- read_csv(here("data", "binary_clin_scen_df_11.08.21.csv"))
```

```
## Rows: 15790 Columns: 34

## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr   (8): ID, sex, zip, race, ethnicity, smoking, end_in_death, SF_LT_200
## dbl  (24): age, BMI, LOS, VP, ECMO, CRRT, NIV, NIV_per_day, HD, INTUB, O2, L...
## date  (2): date, death_date
```

```
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
# Look at structure of df and variable types
# str(bin_clin_scen_df)
head(bin_clin_scen_df, n=40)
```

```
## # A tibble: 40 x 34
##    ID    date         age sex   zip   race  ethnicity smoking   BMI end_in_death
##    <chr> <date>     <dbl> <chr> <chr> <chr> <chr>     <chr>   <dbl> <chr>
##  1 0055~ 2020-10-19    53 Male  95382 Other Hispanic~ Not Cu~  31.6 No
##  2 0055~ 2020-10-20    53 Male  95382 Other Hispanic~ Not Cu~  31.6 No
##  3 0055~ 2020-10-21    53 Male  95382 Other Hispanic~ Not Cu~  31.6 No
##  4 0055~ 2020-10-22    53 Male  95382 Other Hispanic~ Not Cu~  31.6 No
##  5 0055~ 2020-10-23    53 Male  95382 Other Hispanic~ Not Cu~  31.6 No
##  6 0055~ 2020-10-24    53 Male  95382 Other Hispanic~ Not Cu~  31.6 No
##  7 0055~ 2020-10-25    53 Male  95382 Other Hispanic~ Not Cu~  31.6 No
##  8 0055~ 2020-10-26    53 Male  95382 Other Hispanic~ Not Cu~  31.6 No
##  9 0055~ 2020-10-27    53 Male  95382 Other Hispanic~ Not Cu~  31.6 No
## 10 0055~ 2020-10-28    53 Male  95382 Other Hispanic~ Not Cu~  31.6 No
## # ... with 30 more rows, and 24 more variables: death_date <date>, LOS <dbl>,
## #   VP <dbl>, ECMO <dbl>, CRRT <dbl>, NIV <dbl>, NIV_per_day <dbl>, HD <dbl>,
## #   INTUB <dbl>, SF_LT_200 <chr>, O2 <dbl>, LowO2 <dbl>, HighO2 <dbl>,
## #   NODEV <dbl>, SIMPLEDEV <dbl>, SIMPLE_PER_DAY <dbl>, NIVDEV <dbl>,
## #   IVDEV <dbl>, CPAPDEV <dbl>, NCDEV <dbl>, NC_PER_DAY <dbl>, NC_GT_12 <dbl>,
## #   DEATH <dbl>, STAGE <dbl>
```

Clean `bin_clin_scen_df` – add new variable `STAGE` and set values to NA

```
# Create new variable `STAGE` and set values to NA
bin_clin_scen_df$STAGE <- NA
```

Create function to get COVID stages based on patients' binary clinical scenarios and usage of O2 devices (and their respective level of respiratory support)

```
clin_scen_stg_df <- bin_clin_scen_df %>%
  mutate(STAGE = case_when(DEATH==1 ~ 10,
                           ECMO==1 |
                             ((INTUB==1 | IVDEV==1) & SF_LT_200==1 & (VP==1 | CRRT==1)) ~ 9,      # SHO
                           ((INTUB==1 | IVDEV==1) & SF_LT_200==1) |    # SHOULD I JUST USE 'IVDEV' OR (
                             VP==1 |
                             CRRT==1 ~ 8,
                           (INTUB==1 | IVDEV==1) &
```

```
                              SF_LT_200==0 ~ 7,
                            (NIVDEV==1 & CPAPDEV==0) &
                              IVDEV==0 &
                              ((NIV_per_day > NC_PER_DAY) | NC_GT_12==1 | (NC_PER_DAY>0 & NC_GT_12==1)) ~
                            SIMPLEDEV==1 & (LowO2==1 | HighO2 == 1 | SIMPLE_PER_DAY > 1) ~ 5,
                            NC_PER_DAY>0 & NC_GT_12==0 ~ 5,
                            (NIVDEV==0 & CPAPDEV==0 & INTUB==0 & IVDEV==0 & SIMPLEDEV==0) & LowO2==1 ~ 5,
                            TRUE ~ 4))
```

Create dataset that only contains demographics, death, and stage

```
dm_covid_stg <- clin_scen_stg_df %>%
  select(-c(VP:NC_GT_12))
```

Save and export as .csv files

```
write_csv(clin_scen_stg_df, here("data", "clin_scen_stg_df_11.08.21.csv"))
write_csv(dm_covid_stg, here("data", "dm_covid_stg_11.08.21"))
```

End of Document