socialrisk R Package

Wyatt P. Bensken
PhD Candidate, Epidemiology and Biostatistics
Case Western Reserve University
wyattbensken.com
wpb27 [at] case.edu

Agenda

- Overview of the Package
- Installing the Package
- Included Data
- Included Data Cleaning Function
- Included Taxonomies
- The package can be found at:
  - https://CRAN.R-project.org/package=socialrisk

Overview of the Package

Why Do We Need The `socialrisk` package?

- Social determinants of health and their accompanying social risks and social needs are increasingly recognized as important parts of health and health care
- With the introduction of ICD-10 in 2015, there was a new class of codes to capture social risk
  - Start with the prefix "Z" and thus are called "Z-codes"
- There is notable divergence in the literature around which Z codes, and their conceptual domains, are part of social risk/social need analyses
- This package allows one to implement any of three approaches and taxonomies in their data using ICD-10 Z-codes

What Does The Package Do?

- The `socialrisk` package includes functions to:
  - Clean the data
  - Run three different taxonomies
- The package also includes a number of sample datasets which can help demonstrate the package's functions

Installing the Package

Installing `socialrisk`

- `socialrisk` is available on CRAN
  - https://CRAN.R-project.org/package=socialrisky
  - It can be installed with the following code:

```
install.packages("socialrisk")
```

- The development version of `socialrisk` can be found on my GitHub
  - https://github.com/WYATTBENSKEN/socialrisk

Included Data

Claims Data (i10_wide)

- A dataset with fake patient data for 5 patients
  - Both inpatient and outpatient data, as well as HCPCS codes, and ICD10 codes
  - Only ICD-10 codes are needed for the package
- 29 rows, and 11 variables

```
 [1] "date_of_serv" "dx1"          "dx2"
 [4] "dx3"          "dx4"          "dx5"
 [7] "hcpcs"        "icd_version"  "patient_id"
[10] "sex"          "visit_type"
```

Claims Data (i10_wide)

A few rows as an example.

patient_id	sex	date_of_serv	visit_type	dx1	dx2	dx3
1001	male	2020-02-14	ip	E876	Z560	Z6372
1001	male	2021-05-15	ip	J189	Z644	A408
1001	male	2021-01-10	ot	I119	Z628	I10

dx4	dx5	hcpcs	icd_version
Z654	E440	E2201	10
I10	G309	E2201	10
NA	NA	E2201	10

Included Data Cleaning Function

clean_data()

- Takes our raw claims data, in a number of different forms, and prepares it in a way which allows the other function in this package to easily work with it
  - It is recommended to run this function on all data

clean_data()

```
data <- clean_data(dat = , id = ,
                   style = , prefix_dx = "dx")
```

- dat, our dataset
- id, the variable name of our our patient id
- style, “wide” or “long”
- prefix_dx, the prefix for all of our diagnoses variables

clean_data()

```
clean_data(dat = i10_wide, id = patient_id,
           style = "wide", prefix_dx = "dx")
```

```
# A tibble: 5 x 2
  patient_id dx
  <fct>      <chr>
1 1001       E876
2 1001       Z560
3 1001       Z6372
4 1001       Z654
5 1001       E440
```

socialrisk() Function

socialrisk()

- All of the taxonomies are run with the same `socialrisk()` function. The general structure of this function is:

```
socialrisk(dat = , id = , dx = , taxonomy = )
```

- dat, our dataset
- id, the variable name of our our patient id
- dx, the diagnosis variable name
- taxonomy, the social risk taxonomy to implement, either “cms”, “mha”, or “siren”

Included Taxonomies

“cms”

- The Centers for Medicare and Medicaid Services (CMS) has identified Z55-Z65 as the Z-codes for social risk
  - Source (hyperlink): “Using Z Codes”
- However they have not placed them into conceptual categories
- In this package we just use the parent code to place them into categories

```
cms <- socialrisk(dat = data, id = patient_id,
                  dx = dx, taxonomy = "cms")
```

“cms”

```
# A tibble: 5 x 12
  patient_id any_social_risk number_domains
  <fct>                <dbl>          <dbl>
1 1001                     1              7
2 1002                     1              2
3 1003                     1              2
4 1004                     0              0
5 1005                     0              0
# ... with 9 more variables: z55_education <dbl>,
#   z56_employment <dbl>, z57_occupation <dbl>,
#   z59_housing <dbl>, z60_social <dbl>,
#   z62_upbringing <dbl>, z63_family <dbl>,
#   z64_psychosocial <dbl>,
#   z65_psychosocial_other <dbl>
```

“mha”

- The Missouri Hospital Association uses the same set of codes as CMS, but has put them into 5 conceptually meaningful categories, separate from the parent codes
  - Source (hyperlink): “Decoding the Social Determinants of Health”

```
socialrisk(dat = data, id = patient_id,
           dx = dx, taxonomy = "mha")
```

“mha”

```
# A tibble: 5 x 8
  patient_id any_social_risk number_domains employment
  <fct>                <dbl>          <dbl>      <dbl>
1 1001                     1              5          1
2 1002                     1              2          0
3 1003                     1              1          1
4 1004                     0              0          0
5 1005                     0              0          0
# ... with 4 more variables: family <dbl>,
#   housing <dbl>, psychosocial <dbl>, ses <dbl>
```

"siren"

- The Social Interventions Research & Evaluation Network (SIREN) at the University of California, San Francisco (UCSF) compendium uses a broader range of codes (beyond Z-codes), has a larger number of conceptual categories, but notably excludes occupational exposures
  - Source (hyperlink): "Compendium of medical terminology codes for social risk factors"
- The conceptual domains are **not** mutually exclusive, so a single Z-code may be assigned to multiple domains

```
socialrisk(dat = data, id = patient_id,
           dx = dx, taxonomy = "siren")
```

“siren”

```
# A tibble: 5 x 19
  patient_id any_social_risk number_domains access
  <fct>                <dbl>          <dbl>  <dbl>
1 1001                     1              5      0
2 1002                     1              6      1
3 1003                     1              1      0
4 1004                     0              0      0
5 1005                     0              0      0
# ... with 15 more variables: education <dbl>,
#   employment <dbl>, finances <dbl>, food <dbl>,
#   housing <dbl>, immigration <dbl>,
#   incarceration <dbl>, language <dbl>,
#   race_eth <dbl>, safety <dbl>, soc_connect <dbl>,
#   stress <dbl>, transportation <dbl>,
#   utilities <dbl>, veteran <dbl>
```

Questions?

Contact Information

- Wyatt P. Bensken
- wyattbensken.com
- wpb27 [at] case.edu