

2018-09-09

## Erik Bulow

- Swedish
- PhD student in medicine (orthopedics)
- Statistician
- Use micro data from official sources (Statistics Sweden)

## Categorization

- Deterministic (defined mapping between codes and categories)
- Not statistical classification

## TOC

- What's the problem?
- How to solve it?
- The coder package.
- Example with comorbidity.
- Details about the package.

## My problem

- Patients from one register
- Previous medical conditions (comorbidity) from another (much bigger) register.

```
## (polygon[GRID.polygon.1], polygon[GRID.polygon.2], polygon[GRID.polygon.3], polygon[GRID.polygon.4], text[GRID.text.5], text[GRID.text.6], lines[GRID.li
```

## How can we do this?

- SAS macro exist (doh ...!)
- Possible with some basic but messy R-script (18 hours to run)
- R-package `comorbidities.icd10` (slow and complicated, not on CRAN)
- R-package `icd` (lacks some of the desired features, was initially slow and only contains specific coding schemes)
- Let's make another package!

## coder

- Separate functionality from classification schemes
- Classification using regular expression
- Some use of matrix algebra when possible
- Rely on `data.table`
- Avoid other dependencies (for internal IT system)
- Keep it simple
- Pipe-friendly (or use just one function)

## Centered around three objects

1. Item data (the patients)
2. Code data for the same items (the patient register)
3. Classification scheme (based on reg-ex and with weighting schemes and additional conditions).

## Codes

- Item data with many units (individuals)
- Code data with many instances (hospital visits) per item
  - Some time frame of relevance (one year before surgery or 30 days after)
- Many codes per code instance
  - Some of them relevant
  - Some of them with additional characteristics (main vs additional diagnostic codes)
- Relevant codes grouped by category
  - One code can belong to more than one category
- Weighting schemes for combined indices (total comorbidity burden based on individual comorbidities)

## Example: Charlson Comorbidity

- Group of patients
- Link to National Patient Register for comorbidity data during the year before surgery
- Identify comorbidities based on the Charlson comorbidity classification scheme
- Calculate comorbidity index due to one of the proposed standards

# The coder package

## Patients

Here by name, otherwise by personal identity numbers (PIN) for ((almost) perfect linkage.

```
head(ex_people)
##           name      surgery
## 1 Beaver, Tristin 2016-09-11
## 2 Maestas, Lilibeth 2016-10-23
## 3           Jung, Derek 2017-02-20
## 4 Hayes, Kylihah 2016-12-31
## 5 el-Riaz, Adam 2016-04-19
## 6 Sanchez, Dominique 2017-02-25
```

## Medical records

```
head(ex_pardata)
##           id variable      code code_date hdia
## 1: el-Haider, Ruwaid  hdia T840F 2012-01-04 TRUE
## 2: Roacho, Marla     bdia2 Z510 1993-01-23 FALSE
## 3: el-Hariri, Huda   hdia A469 2008-10-07 FALSE
## 4: el-Parsa, Ghazaala hdia M205 2010-10-28 FALSE
## 5: Martinez, Jonathan hdia 989,98 2008-02-12 TRUE
## 6: Martinez, Crystal hdia H401C 2006-09-05 TRUE
```

## Charlson comorbidity

```
charlson_icd10[c(5, 19), 1:5]
##           group                regex charlson deyo_ramano dhoore
## 5 dementia ^{F0([0-3]|51)}G3(0|11))          1           1      1
## 19 AIDS/HIV ^{B2[0124]}                    6           1      1
```

## Visualize classification scheme

```
visualize(charlson_icd10, c("congestive heart failure"))
```

## Summarise classification scheme

```
suppressPackageStartupMessages(library(tidyverse))
summary(charlson_icd10) %>%
  mutate(codes = substr(codes, 1, 30))
##           group      n      codes
## 1           AIDS/HIV 22 B200, B201, B202, B203, B204,
## 2 cerebrovascular disease 82 G450, G451, G452, G453, G454,
## 3 chronic pulmonary disease 57 I278, I279, J409, J410, J411,
## 4 congestive heart failure 8 I099, I110, I130, I132, I500,
## 5           dementia 23 F000, F001, F002, F009, F010,
## 6 diabetes complication 71 E102, E102A, E102B, E102C, E10
## 7 diabetes without complication 55 E100, E100A, E100B, E100C, E10
## 8 hemiplegia or paraplegia 22 G041, G114, G801, G801A, G801B
## 9           malignancy 525 C000, C001, C002, C003, C004,
## 10 metastatic solid tumor 29 C770, C771, C772, C773, C774,
## 11 mild liver disease 83 B180, B180A, B180B, B180C, B18
## 12 moderate or severe liver disease 11 I850, I859, I864, I982, K704,
## 13 myocardial infarction 15 I210, I211, I212, I213, I214,
## 14 peptic ulcer disease 36 K250, K251, K252, K253, K254,
## 15 peripheral vascular disease 38 I700, I701, I702, I702A, I702C
## 16           renal disease 27 I120, I131, N032, N033, N034,
## 17 rheumatic disease 63 M050, M051, M052, M053, M058,
```

## Add Charlson comorbidity

```
ex <-
ex_people %>%
  categorize(
    ex_icd10,
    "charlson_icd10",
    id = "name",
    date = "surgery",
    days = c(-365, -1),
    ind = "quan_original"
  )
# Patient data from SHAR
# Categorize this data using ...
# ... data from NPR
# based on Charlson comorbidity from ICD-10
# Identify id variable from SHAR
# Identify date variable to relate to
# Only include comorbidity during this period
# Calculate comorbidity index with specified weights
```

## Result

```
ex[, c(1, 7, 16, 21:22)]
##           name dementia malignancy AIDS/HIV quan_original
## 1: Abzari, Joseph FALSE TRUE FALSE 2
## 2: Alexis, Dayveon FALSE FALSE FALSE 0
```

## 3:	Anderson, Laquesh	FALSE	FALSE	FALSE	2
## 4:	Armijo Jr, Kalyne	FALSE	FALSE	FALSE	0
## 5:	Babcock, Landon	FALSE	TRUE	FALSE	2
## 6:	Barela, Marco	FALSE	FALSE	FALSE	0
## 7:	Barger, Willanna	FALSE	FALSE	FALSE	0
## 8:	Barnes, Amos	FALSE	FALSE	FALSE	0
## 9:	Bauder, Taylor	FALSE	FALSE	FALSE	0
## 10:	Beasley, Michael	FALSE	FALSE	FALSE	0
## 11:	Beaver, Tristin	FALSE	FALSE	FALSE	0
## 12:	Bille, Cheyenne	FALSE	FALSE	FALSE	0
## 13:	Boffill, Jordan	FALSE	FALSE	FALSE	0
## 14:	Bradshaw, Noah	FALSE	FALSE	FALSE	0
## 15:	Breshears, Kayla	FALSE	FALSE	FALSE	0
## 16:	Bruce, Bryan	FALSE	FALSE	FALSE	0
## 17:	Bruno, Jorge	FALSE	TRUE	FALSE	3
## 18:	Burton, Danyalle	FALSE	FALSE	FALSE	0
## 19:	Charlie, Mareno	FALSE	FALSE	FALSE	0
## 20:	Deshazor, Kordell	FALSE	FALSE	FALSE	0
## 21:	Dominguez Enriquez, David	FALSE	FALSE	FALSE	0
## 22:	Edmondson, Christian	FALSE	TRUE	FALSE	2
## 23:	Elder, Kevyn	FALSE	FALSE	FALSE	1
## 24:	Gallegos, Diana	NA	NA	NA	NA
## 25:	Ghimire, Chelsea	FALSE	FALSE	FALSE	0
## 26:	Hall, Samorra	FALSE	FALSE	FALSE	0
## 27:	Hayes, Kylihah	FALSE	FALSE	FALSE	0
## 28:	Headdden, Devon	FALSE	FALSE	FALSE	0
## 29:	Henricks, Donovan	FALSE	FALSE	FALSE	0
## 30:	Her, Lia	FALSE	FALSE	FALSE	0
## 31:	Hicks, Camren	FALSE	FALSE	FALSE	0
## 32:	Hoang, Elijah	FALSE	FALSE	FALSE	0
## 33:	Jiles-Wright, Bralynd	FALSE	FALSE	FALSE	1
## 34:	Jung, Derek	FALSE	FALSE	FALSE	0
## 35:	Kane, Chapin	FALSE	FALSE	FALSE	0
## 36:	Kim, Alisa	FALSE	FALSE	FALSE	0
## 37:	King, Hali	FALSE	FALSE	FALSE	0
## 38:	Laulu, Kajtshiab	FALSE	TRUE	FALSE	8
## 39:	Lopez, Chiara	FALSE	TRUE	FALSE	2
## 40:	Maestas, Lilibeth	FALSE	TRUE	FALSE	2
## 41:	Mamet, A D	FALSE	FALSE	FALSE	0
## 42:	Marquez, Jacqueline	FALSE	FALSE	FALSE	0
## 43:	Martin, Elizario	FALSE	FALSE	FALSE	0
## 44:	Martinez, David	FALSE	FALSE	FALSE	0
## 45:	Martinez, Jessica	FALSE	FALSE	FALSE	0
## 46:	Martinez, Juan	FALSE	FALSE	FALSE	0
## 47:	Martinez, Nelson	FALSE	FALSE	FALSE	0
## 48:	Mccarthy, Gabriel	FALSE	FALSE	FALSE	0
## 49:	Mcginnis, Adunola	FALSE	FALSE	FALSE	0
## 50:	Mcknight, Jaymes	FALSE	FALSE	FALSE	0
## 51:	Miller, Darshay	FALSE	FALSE	FALSE	0
## 52:	Moktader, Sameer	FALSE	FALSE	FALSE	0
## 53:	Mosely, Jereek	FALSE	FALSE	FALSE	0
## 54:	Nam, John	FALSE	FALSE	FALSE	0
## 55:	Nelson, Phillip	FALSE	FALSE	FALSE	0
## 56:	Okoye, Vanessa	FALSE	FALSE	FALSE	0
## 57:	Parker, Pharo	FALSE	FALSE	FALSE	0
## 58:	Perez Jr, Hannah	NA	NA	NA	NA
## 59:	Perez, Kimberly	FALSE	FALSE	FALSE	2
## 60:	Pham, Kyle	FALSE	FALSE	FALSE	0
## 61:	Pronteau, Colton	FALSE	FALSE	FALSE	0
## 62:	Reyes, Sylvia	NA	NA	NA	NA
## 63:	Rodriguez, Alma	FALSE	FALSE	FALSE	0
## 64:	Rowley Booneel, Hailee	FALSE	FALSE	FALSE	0
## 65:	Rusnak, Kyle	NA	NA	NA	NA
## 66:	Sanburg, Cassandra	FALSE	FALSE	FALSE	0
## 67:	Sanchez, Dominique	FALSE	FALSE	FALSE	0
## 68:	Santee, Holly	FALSE	FALSE	FALSE	0
## 69:	Schell, Jack	FALSE	FALSE	FALSE	2
## 70:	Schubert, Patrick	FALSE	FALSE	FALSE	0
## 71:	Simpson, King	FALSE	FALSE	FALSE	0
## 72:	Smith, John	FALSE	FALSE	FALSE	0
## 73:	Smith, Ruasha	FALSE	FALSE	FALSE	0
## 74:	Spoerl, John	FALSE	TRUE	FALSE	2
## 75:	Stephan, Scott	FALSE	FALSE	FALSE	2
## 76:	Suriwong, Sarah	FALSE	FALSE	FALSE	1
## 77:	Thompson, Stephen	FALSE	FALSE	FALSE	0
## 78:	Thwaites, Travis	FALSE	FALSE	FALSE	0
## 79:	Todacheene, Erminio	FALSE	FALSE	FALSE	1
## 80:	Tomlinson, Amanda	FALSE	FALSE	FALSE	0
## 81:	Tuitele-Britton, Janet	FALSE	FALSE	FALSE	0
## 82:	Valencia, Johnny	FALSE	TRUE	FALSE	2
## 83:	Vallie, Mitchellle	FALSE	FALSE	FALSE	1
## 84:	Vigo, Shi Hyung	FALSE	TRUE	FALSE	2
## 85:	Ward, Harrison	NA	NA	NA	NA
## 86:	Wilkerson, Teesa	FALSE	FALSE	FALSE	0
## 87:	Wood, Elijah	FALSE	FALSE	FALSE	0
## 88:	al-Abid, Asad	FALSE	FALSE	FALSE	0
## 89:	al-Ahsan, Umar	FALSE	FALSE	FALSE	0
## 90:	al-Eid, Abdus Samad	FALSE	TRUE	FALSE	2
## 91:	al-Kassem, Humaidaan	NA	NA	NA	NA
## 92:	al-Mansur, Kawkab	FALSE	FALSE	FALSE	2
## 93:	el-Assad, Jubair	FALSE	FALSE	FALSE	0
## 94:	el-Bina, Muntaha	FALSE	TRUE	FALSE	2
## 95:	el-Kazi, Najeeba	FALSE	FALSE	FALSE	0
## 96:	el-Masood, Nasreen	FALSE	FALSE	FALSE	0
## 97:	el-Nagi, Waddaah	NA	NA	NA	NA
## 98:	el-Riaz, Aadam	FALSE	FALSE	FALSE	0
## 99:	el-Wakim, Nuha	FALSE	FALSE	FALSE	0

```
## 100:      el-Zakaria, Saajid  FALSE      TRUE  FALSE      2
##          name dementia malingnancy AIDS/HIV quan_original
```

## More control

Function categorize takes care of everything. Alternative workflow:

```
ex_people %>%
  codify(ex_icd10, "name", "surgery", days = c(-365, -1)) %>%
  classify(charlson_icd10) %>%
  index("quan_original")
```

```
##      Alexis, Dayveon      Bauder, Taylor
##      0
##      Beasley, Michael    Deshazor, Kordell
##      0
##      Henricks, Donovan  Marquez, Jacqueline
##      0
##      Martinez, David    Mccarthy, Gabriel
##      0
##      Simpson, King      Gallegos, Diana
##      0                    NA
##      Perez Jr, Hannah   Reyes, Sylvia
##      NA                    NA
##      Rusnak, Kyle       Ward, Harrison
##      NA                    NA
##      al-Kassem, Humaidan el-Nagi, Waddaah
##      NA                    NA
##      Abzari, Joseph     al-Abid, Asad
##      2                      0
##      al-Ahsan, Umar    al-Eid, Abdus Samad
##      0                      2
##      al-Mansur, Kawkab  Anderson, Laquesh
##      2                      2
##      Armijo Jr, Kalyne  Babcock, Landon
##      0                      2
##      Barela, Marco     Barger, Willanna
##      0                      0
##      Barnes, Amos      Beaver, Tristin
##      0                      0
##      Bille, Cheyenne   Boffill, Jordan
##      0                      0
##      Bradshaw, Noah    Breshears, Kayla
##      0                      0
##      Bruce, Bryan      Bruno, Jorge
##      0                      3
##      Burton, Danyalle  Charlie, Mareno
##      0                      0
##      Dominguez Enriquez, David Edmondson, Christian
##      0                      2
##      el-Assad, Jubair  el-Bina, Muntaha
##      0                      2
##      el-Kazi, Najeeba  el-Masood, Nasreen
##      0                      0
##      el-Riaz, Aadam    el-Wakim, Nuha
##      0                      0
##      el-Zakaria, Saajid Elder, Kevyn
##      2                      1
##      Ghimire, Chelsea  Hall, Samorra
##      0                      0
##      Hayes, Kylihah    Headden, Devon
##      0                      0
##      Her, Lia          Hicks, Camren
##      0                      0
##      Hoang, Elijah     Jiles-Wright, Bralynd
##      0                      1
##      Jung, Derek      Kane, Chapin
##      0                      0
##      Kim, Alisa       King, Hali
##      0                      0
##      Lulu, Kajtshiab   Lopez, Chiara
##      8                      2
##      Maestas, Lilibeth Mamet, A D
##      2                      0
##      Martin, Elizario  Martinez, Jessica
##      0                      0
##      Martinez, Juan    Martinez, Nelson
##      0                      0
##      Mcginnis, Adunola Mcknight, Jaymes
##      0                      0
##      Miller, Darshay   Mktader, Sameer
##      0                      0
##      Mosely, Jereek    Nam, John
##      0                      0
##      Nelson, Phillip   Okoye, Vanessa
##      0                      0
##      Parker, Pharo     Perez, Kimberly
##      0                      2
##      Pham, Kyle       Pronteau, Colton
##      0                      0
##      Rodriguez, Alma  Rowley Booneel, Hailee
##      0                      0
##      Sanburg, Cassandra Sanchez, Dominique
##      0                      0
##      Santee, Holly    Schell, Jack
##      0                      2
```

##	Schubert, Patrick	Smith, John
##	0	0
##	Smith, Ruasha	Spoerl, John
##	0	2
##	Stephan, Scott	Suriwong, Sarah
##	2	1
##	Thompson, Stephen	Thwaites, Travis
##	0	0
##	Todacheene, Erminio	Tomlinson, Amanda
##	1	0
##	Tuitele-Britton, Janet	Valencia, Johnny
##	0	2
##	Vallie, Mitchellle	Vigo, Shi Hyung
##	1	2
##	Wilkerson, Teesa	Wood, Elijah
##	0	0

## Classification schemes

- Comorbidity
  - Charlson (ICD-10)
  - Elixhauaser (ICD-10)
  - comorbidity-polypharmacy score (ICD-10)
  - RxRiskV (based on ATC codes) (ICD-10)
- Adverse events
  - After hip arthroplasty (ICD-10 and KVÅ)
  - After knee arthroplasty (ICD-10)
- Example scheme for car brands
- S3 class mechanism to make tailored classification schemes

## Where to find it?

- Github: [www.github.com/eribul/coder](https://www.github.com/eribul/coder)
- Documented by pkgdown: [eribul.github.io/coder](https://eribul.github.io/coder)
  - Some vignettes exist and more are planned
- Plan to increase test coverage (now 65 %)
- Plan to release on CRAN

## So ... how fast is it?

- It's quite fast!
- What took 18 hours before now takes 30 seconds!
- The newly updated `icd` package is even faster though, but is not as generic and.

## Thanks!

Questions?