UNIVERSITY OF GOTHENBURG

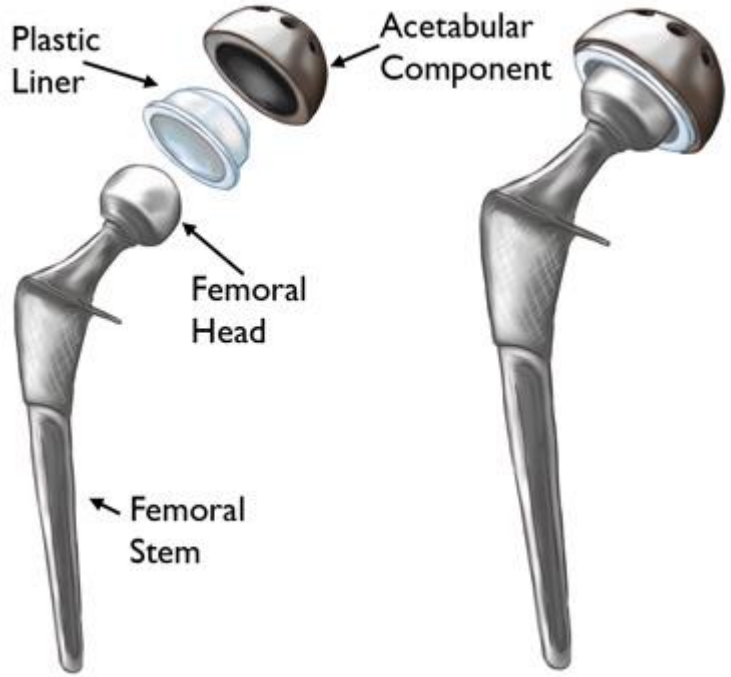
EXPERIENCES FROM A REAL-WORLD PREDICTION MODELLING PROJECT

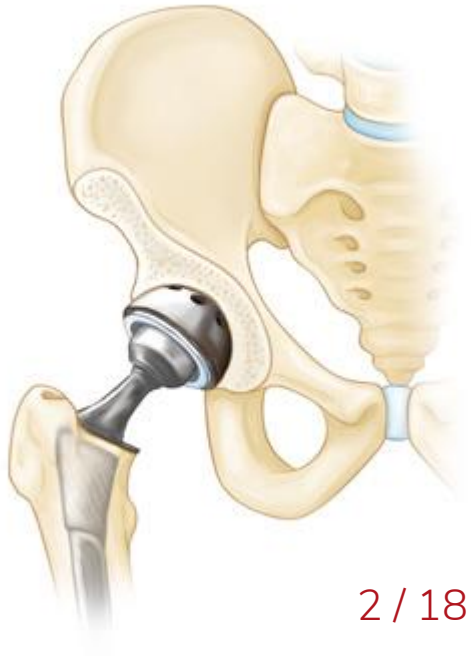
Swedish Hip Arthroplasty Register

ERIK BÜLOW, STATISTICIAN AND PHD STUDENT

The optimists

- Predict 90-day mortality after elective total hip arthroplasty
- Web calculator
- Publish in high impact journal
- Save the world!
- Get the Nobel price!
- Become immortable heroes!





The pessimists

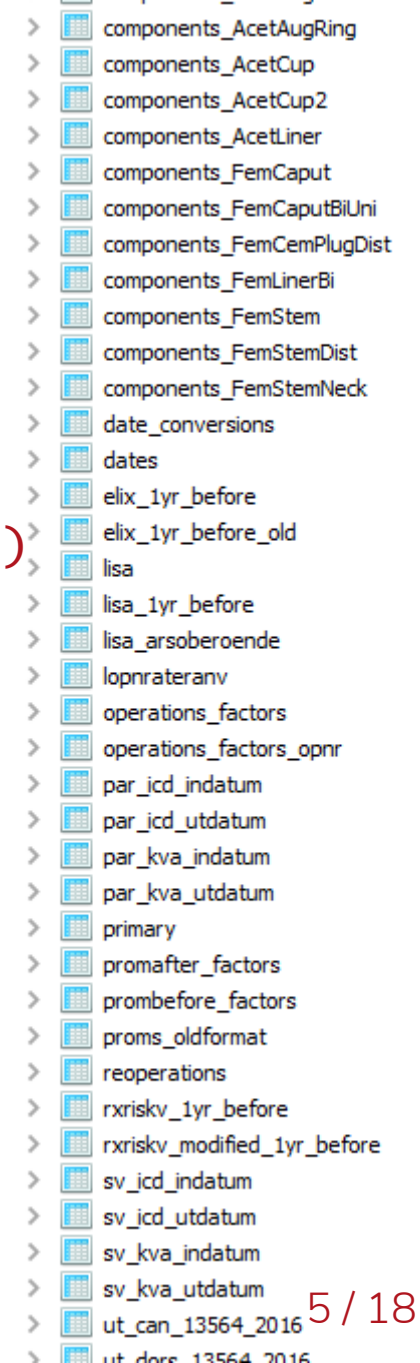
- Observational data (no intervention)
- Imbalanced data (0.3 % dies)
- We already know why patients die!

Reality

- Ethical permission 2013
  - Extended several times 2014-2017
- Data linkage using personal identitiy numbers 2016
- Massive data sets. Slow computing.

Data management

- SQLite/`dplyr`
- `data.table`!
- Regular expressions `^(I(099|1(10|3[02])|255|4(2[05-9]|3)|50)|P290)`
- `profvis`
- `furrr`
- MRO did not help in this case
- Slow local computer infra-structure!



coder

The goal of `coder` is to classify items from one dataset, using codes from a secondary source. Please se vigtnettes with introductionary examples!

Installation

You can (soon) install the released version of coder from CRAN with:

```
# install.packages("coder") # Not yet!
```

And the development version from GitHub with:

```
# install.packages("devtools")
devtools::install_github("eribul/coder")
```

Classification schemes

Classification schemes are used to classify items. These schemas are constructed by regular expressions for computational speed, but their content can be summarized and visualized for clarity.

The package includes several default classification schemes (so called `classcodes` objects). Most of these are related to medical

Links

Browse source code at
https://github.com/eribul/coder

Report a bug at
https://github.com/eribul/coder/issues

License

MIT + file LICENSE

Developers

Erik Bulow
Author, maintainer

Dev status

build failing
build passing
codecov 58%
repo status Active

First modelling attempt

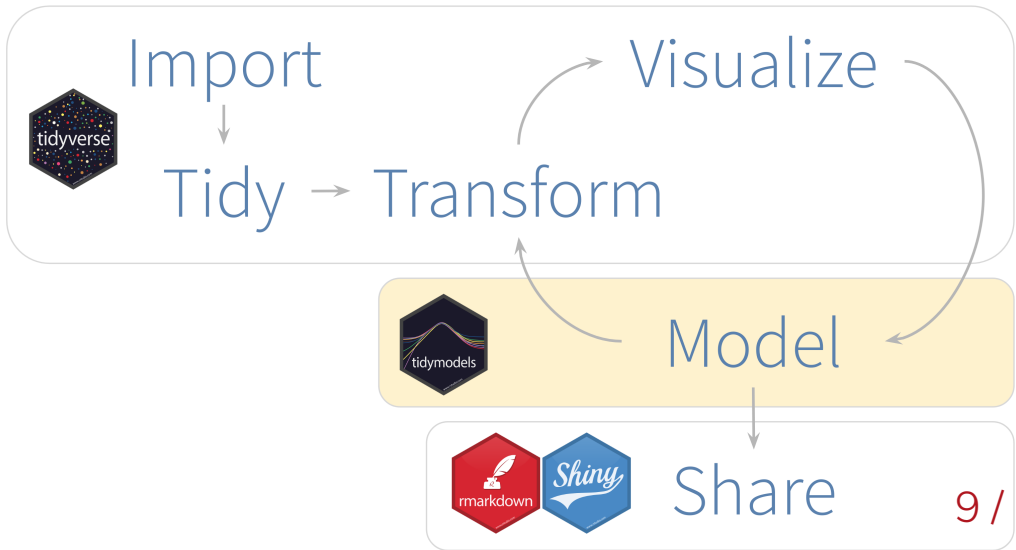
- Data:
  - demography, socio-economics
  - surgical, components, hospital
  - comorbidities (Charlson/Elixhauser)
  - medical prescriptions (RxRisk V).
- Logistic regression (dead/alive after 90 days)
  - Univariable screening
  - Step-wise
- Results
  - Bad sensitivity
  - Good specificity and AUC
  - Clinically reasonable model

Stop!

- External validation!
- Newer data!
- Statistician quit!

Second modelling attempt

- New linkage data base (1999-2015)
- New ethical permissions
- New statistician (me)
- `ProjectTemplate`
- `tidymodels`
- Downsampling, Bagging and model averaging



STOP!

- Don't do down sampling!
- Don't use medical prescription data!
- Don't use all those ad hoc methods!

Third modelling attempt

- Bootstrap ranking with penalized regression (LASSO) and piecewice linear regression to find break points
- Adjust for optimism (overfitting)
- Calibration belt plots
- Transparent reporting
- Only for $\hat{p} < .03$

PLOS ONE

RESEARCH ARTICLE

Improved Variable Selection Algorithm Using a LASSO-Type Penalty, with an Application to Assessing Hepatitis B Infection Relevant Factors in Community Residents

Pi Guo[1,2], Fangfang Zeng[1,2], Xiaomin Hu[1,2], Dingmei Zhang[1,2], Shuming Zhu[1,2], Yu Deng[1,2], Yuantao Hao[1,2]*

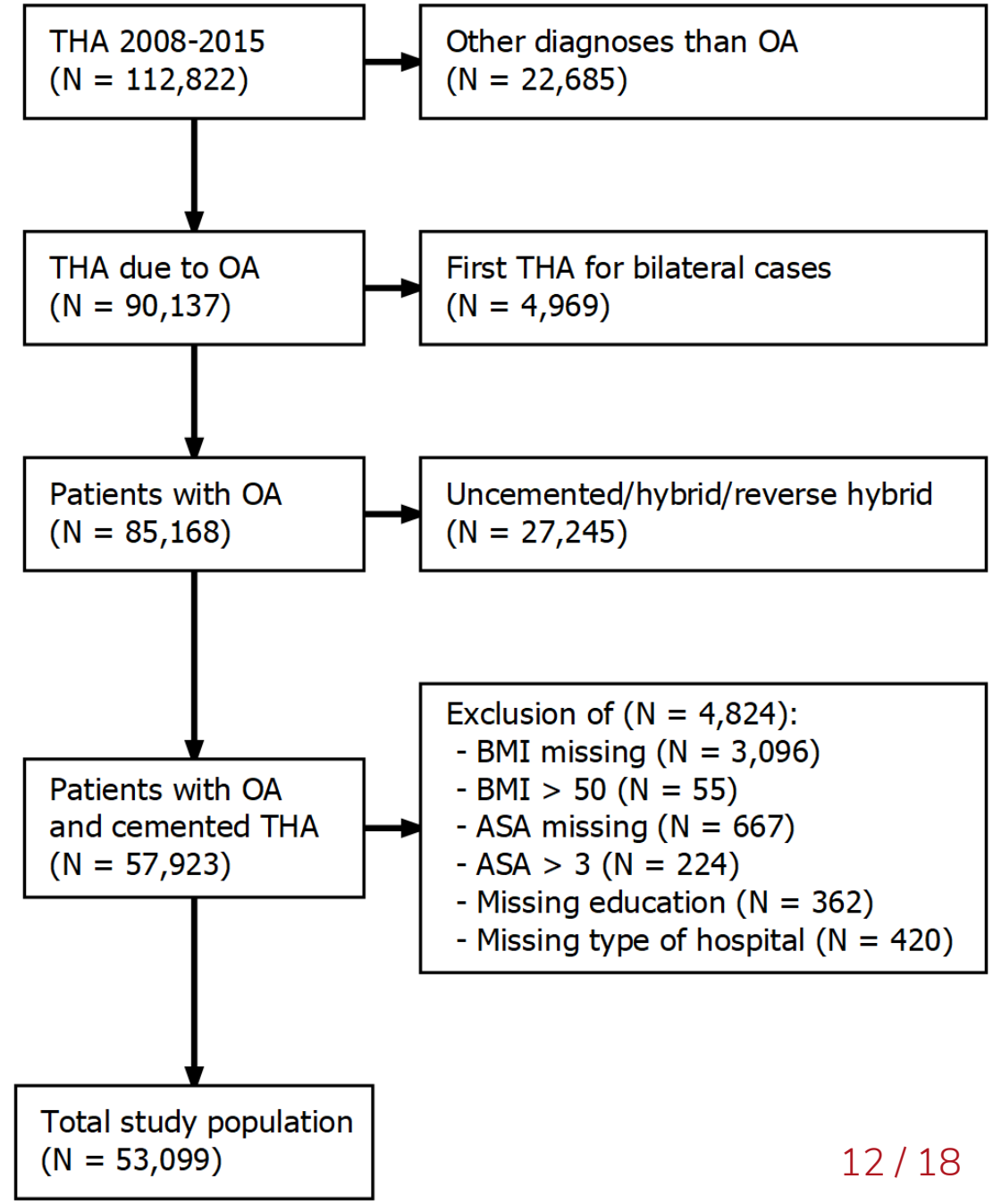
1 Department of Medical Statistics and Epidemiology and Health Information Research Center, School of Public Health, Sun Yat-sen University, Guangzhou, Guangdong, 510080, China, **2** Laboratory of Health Informatics, Guangdong Key Laboratory of Medicine, Sun Yat-sen University, Guangzhou, Guangdong, 510080, China

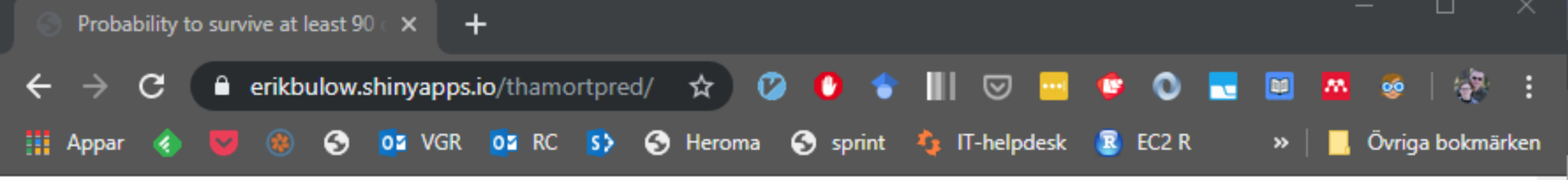
* haoyt@mail.sysu.edu.cn

CrossMark click for updates

Automated flowchart

- `rlang` (quosures)
- `DiagrammeR`





PROBABILITY TO SURVIVE AT LEAST 90 DAYS AFTER THA SURGERY?

Bla bla bla ...

Age

35 72 99

35 42 49 56 63 70 77 84 91 98 99

Sex

Male Female

ASA grade

1 2 3

Do you have any of the following:

Cancer

CNS disease

Prediction About

Around
99.94 %

September 23, 2019

Software Open Access

eribul/rmst: Submitted to Clinical Epidemiology

Erik Bülow

Article on Restricted Mean Survival Time

Preview

rmst-v0.1.zip

- eribul-rmst-ff805dc
  - .gitignore 56 Bytes
  - 2019-01702.Rproj 257 Bytes
  - README.md 560 Bytes
  - config
    - README.md 122 Bytes
    - global.dcf 591 Bytes
  - data
    - linkage.R 452 Bytes
  - graphs
    - flowchart.gv 881 Bytes
    - flowchart.png 14.9 kB

Edit

New version

5 views 0 downloads

See more details...

Available in

GitHub

Indexed in

Binder

mybinder.org/v2/gh/eribul/thamortpred/master?urlpath=rstudio

Appar VGR RC Heroma sprint IT-helpdesk EC2 R SODA Databasdiagram Cryptshare Trip Epistemonikos JIRA Rocen Övriga bokmärken

Thanks to Google Cloud and OVH for sponsoring our computers!

binder

Starting repository: eribul/thamortpred/master

New to Binder? Check out the Binder Documentation for more information

Build logs show

Here's a non-interactive preview on nbviewer while we start a server for you. Your binder will open automatically when it is ready.

jupyter nbviewer

JUPYTER FAQ

thamortpred

Manuscript

- `bookdown`
- Mendeley + `.bib`
- Word template (line spacing, paginatoin, formatting)
- And of course ... manual editing through e-mails ...

External validation

- Adjustments
  - Model as is
  - re-calibration
- Validation
  - Discrimination: AUC
  - Calibration belts

shpr.registercentrum.se
gu.se