An evaluation of the c-statistic for benefit

Thomas Debray
University Medical Center Utrecht

Acknowledgements

Jeroen Hoogland

University Medical Center
Utrecht
The Netherlands

Orestis Efthimiou

University of Bern
Switzerland

Background

Prediction of absolute treatment effect

- aims at individualized assessment of treatment benefit (or harm)
- estimates outcomes under counterfactual treatment conditions
- involves risk modelling strategies (e.g. regression, machine learning)
- adjusts for baseline risk, relative treatment effect and, if relevant, treatment-covariate interactions

Background

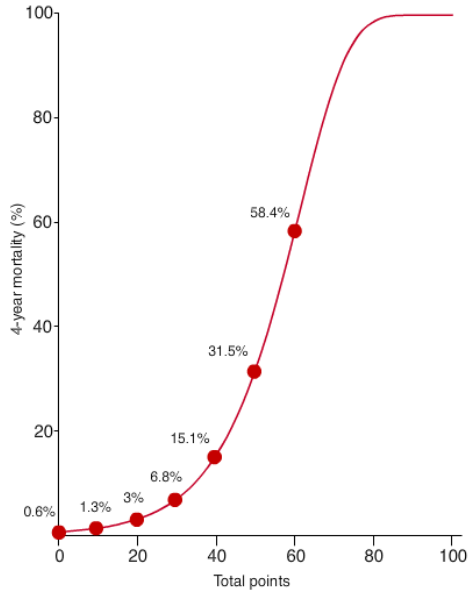
An example: **The SYNTAX score II**

"The SYNTAX score II is a clinical tool that combines clinical variables with the anatomical SYNTAX score, providing expected 4-year mortality for both coronary artery bypass grafting (CABG) and percutaneous coronary intervention (PCI) — thus recommending either PCI only, CABG only or equipoise in treatment based on long-term mortality."

DOI: 10.21037/acs.2018.07.02

Background

SYNTAX SCORE II 4-year mortality



Nomogram depicting predicted 4-year mortality as a function of the SYNTAX II Score for patients proposed to undergo myocardial revascularization (CABG or PCI).

Adapted from Farooq et al., The Lancet. 2013 Feb 23;381(9867):639-50

SYNTAX Score II questions

SYNTAX Score I

Age (years)

CrCl mL/min

LVEF (%)

Left Main ○ no ○ yes

Gender ○ male ○ female

COPD ○ no ○ yes

PVD ○ no ○ yes

SYNTAX Score II Calculate

Background

SYNTAX Score II

SYNTAX II

Decision making -between CABG and PCI- guided by the SYNTAX Score II to be endorsed by the Heart Team.

PCI	
SYNTAX Score II:	46.6
PCI 4 Year Mortality:	24.7 %
CABG	
SYNTAX Score II:	26.8
CABG 4 Year Mortality:	5.3 %
Treatment recommendation ⓘ*:*	*CABG or PCI*

Absolute treatment effect is 19.4% in favor of CABG

Concordance-for-benefit

How to assess the performance of absolute treatment effect predictions?

- Van Klaveren et al. proposed the concordance between predicted and observed treatment benefit (c-for-benefit)
- A regular c-statistic applied to pairs of patients that underwent different treatments but had similar predicted treatment benefit.
- The c-for-benefit has been recommended for comparing prediction models that are used for treatment decision-making

Models with a high c-for-benefit (close to 1) should be preferred

Objective

To evaluate the key properties of the c-for-benefit and their implications for practical application

Toy example

- logit(P(Y = 1)) = $\beta_0 + x - t$

where

- x is a prognostic factor
- t is the received treatment (0 for control , 1 for the alternative treatment)

Note that the treatment effect is constant (i.e. absence of HTE)

Issue #1: Sensitivity to outcome incidence

Variability in treatment benefit (and thus c-for-benefit) is affected by outcome incidence

β_0	x	Pr(Y=1|t=0)	Pr(Y=1|t=1)	Abs treatment effect	C-for-benefit
0	0	0.50	0.27	-0.23	0.5
	1	0.73	0.50	-0.23	
-2	0	0.12	0.05	-0.07	> 0.5
	1	0.27	0.12	-0.15	

Issue #2: Sensitivity to variability in control outcome incidence

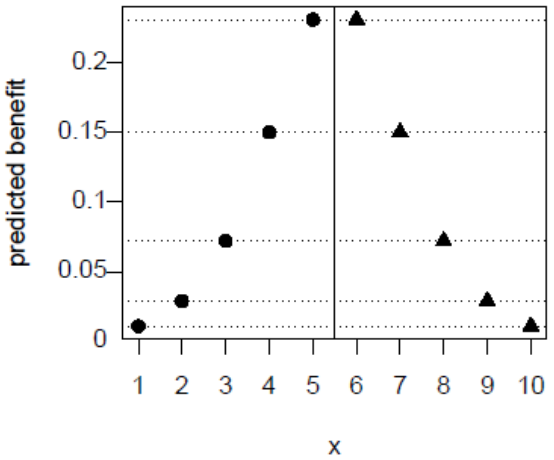
Variability in treatment benefit (and thus c-for-benefit) is affected by variability in prognostic factors.

β_0	x	Pr(Y=1|t=0)	Pr(Y=1|t=1)	Abs treatment effect	C-for-benefit
0	1	0.73	0.5	-0.23	> 0.5
	2	0.88	0.73	-0.09	
	3	0.95	0.88	-0.07	
	4	0.98	0.95	-0.03	
	5	0.99	0.98	-0.01	

Issue #3: Sensitivity to matching procedure

Calculation of the c-for-benefit requires matching of individuals under alternative treatments

- It was suggested to match on absolute predicted treatment benefit.
- Estimates of predicted benefit may be similar despite differences in control outcome risk.
- This generates noise in the comparison of interest and may therefore lead to attenuation bias of the c-for-benefit



Issue #3: Sensitivity to matching procedure

We generated data with uniform probability for x and 1:1 treatment control allocation for 500 patients, repeated for 500 simulations.

Matching procedure	C-for-benfit	95% CI
x (individual covariates)	0.63	0.54 – 0.72
Absolute treatment benefit	0.55	0.51 – 0.60

In the simulations, $\beta_0 = -5$ and $x \in \{1,2,\ldots,10\}$ with equal probability 0.1.

Issue #4: Lack of statistical power

- We only have 3 possible outcomes of treatment benefit: +1, 0 and -1. Hence, there are few eligible pairs to evaluate c-for-benefit.
- There is irreducible error in the difference between the predicted treatment effect (probability) and its manifestation as a (binomial) outcome. Hence, estimation of the c-for-benefit is affected by noise.

Previous simulations showed that obtaining a c-for-benefit > 0.65 is difficult even in the presence of strong treatment-covariate interaction.

https://jhoogland.shinyapps.io/data_exploration/

Final thoughts

- Estimation of the c-for-benefit is difficult (bias & precision)
- The magnitude and interpretation of the c-for-benefit can greatly depend on specific implementation choices.
- The conditions under which the c-for-benefit can be used effectively for model selection are as of yet unclear
- For the logistic model, capturing discriminative power on the patient relevant absolute effect scale and on the parameter level may be best seen as two separate goals. Different implementations of the c-for-benefit may serve either purpose

Simulations planned to evaluate the impact of aforementioned issues