Credibility of risk predictions in medical research: concepts, tools, and applications

Babak Oskooei

London Hub for Trials Methodology Research
MRC Clinical Trials Unit at UCL
27 November 2015

Contents

1. Introduction – why risk prediction is important
2. Risk prediction models
3. Concepts underlying the assessment of risk predictions
   a. Discrimination
   b. Calibration
   c. Predictive ability
4. Performance of a survival risk prediction model
   a. In validation setting: transportability & reproducibility
   b. A new measure of predictive ability: total gain (TG) statistics
5. Some real examples
6. Conclusions

Why is risk prediction important?

- It is used in clinical management of patients
  - Selection for surgery
  - Selection for screening/diagnostic tests
  - Determining prognosis

- It can be used to assess the importance/significance of available prognostic factors as well as the new biomarkers

- We use them in the design of clinical trials
  - E.g. RAMPART trial

Clinical management of patients
Example I: breast cancer

MRC Hubs for Trials Methodology Research
MRC Clinical Trials Unit Hub

- Online web-tool PREDICTwww.predict.nhs.uk :
  - to select the most appropriate adjuvant therapy following surgery

PREDICT Tool: Breast Cancer Survival; Input

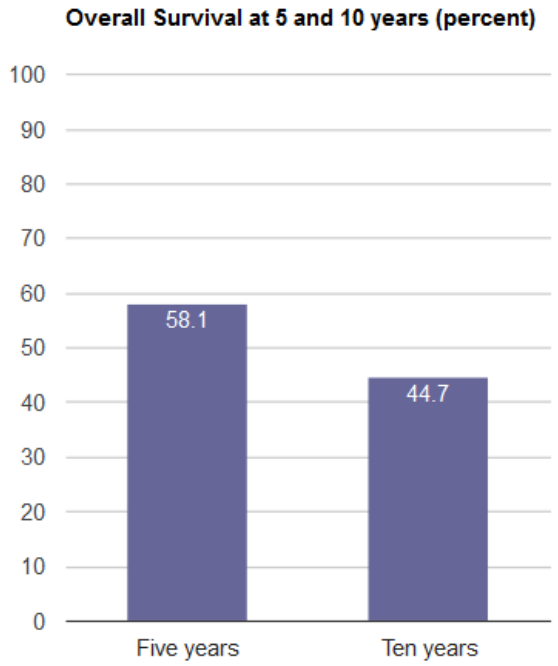
Age at diagnosis:	65		
Mode of detection:	Screen-detected	Symptomatic	Unkn
Tumour size in mm:	11 (blank if unknown)		
Tumour Grade:	1	2	3	Unknown
Number of positive nodes:	2 (blank if unknown)		
ER status:	Positive	Negative	
HER2 status:	Positive	Negative	Unkn
KI67 status:	Positive	Negative	Unkn
Gen chemo regimen:	No chemo	Second	Third

Predict Survival | Clear All Fields | Print Results

PREDICT Tool: Breast Cancer Survival; Results

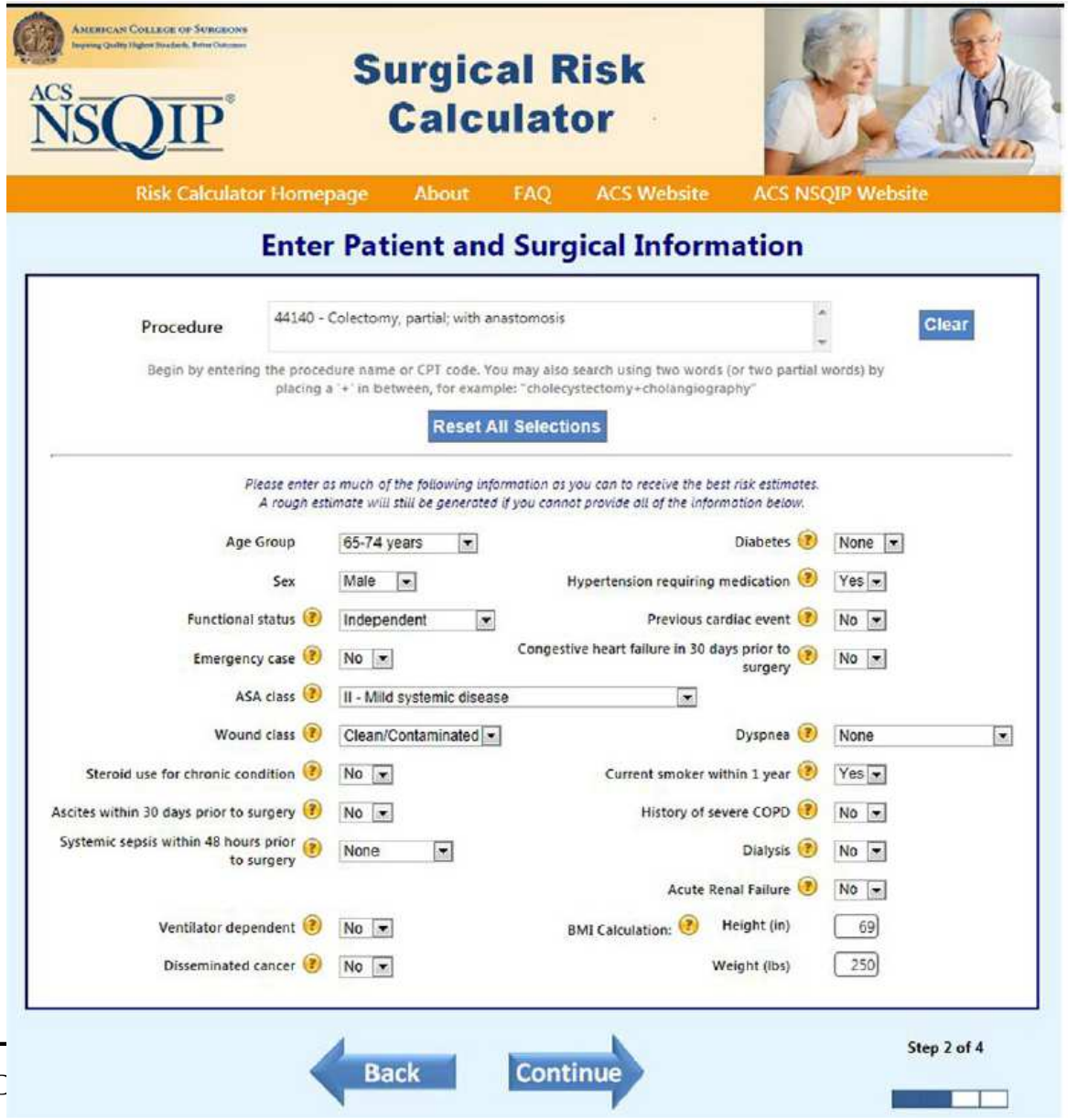
Five year survival
58 out of 100 women are alive at 5 years with no adjuvant therapy after surgery
Ten year survival
45 out of 100 women are alive at 10 years with no adjuvant therapy after surge

To view the numbers in bars hover pointer over each bar-segment
(Or tap segment if using a mobile device)



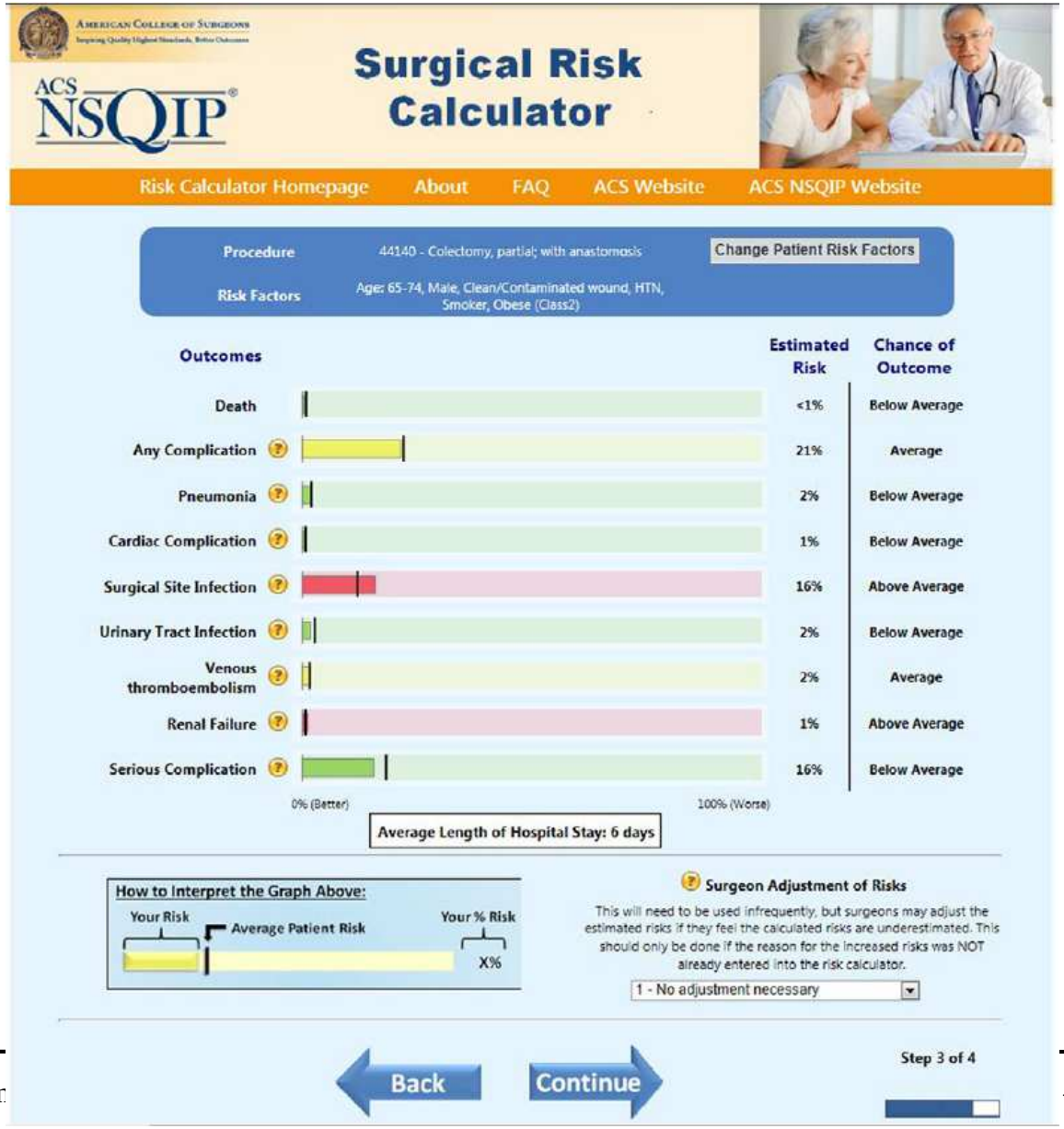
Clinical management of patients

Example II: American college of surgeons surgical risk calculator - http://riskcalculator.facs.org/



Clinical management of patients

Example II: American college of surgeons surgical risk calculator - http://riskcalculator.facs.org/



Clinical risk predictions:
Example III: Cancer prognosis

MRC Hubs for Trials Methodology Research
MRC Clinical Trials Unit Hub

Name of the web-tool	Web address
Adjuvant Online	http://www.adjuvantonline.com/
AJCC—individualized melanoma patient outcome prediction tools	http://www.melanomaprognosis.org/
Artificial neural networks in prostate cancer	http://www.prostatecalculator.org/
Biochemical recurrence-free survival prediction model	http://eurology.surgery.duke.edu/Aspx/PredictionModel/NomogramsModel.aspx
CancerMath	http://www.lifemath.net/cancer/
UCSF—capra Score	http://urology.ucsf.edu/patientGuides/uroOncPt_Assess.html#capra
Cancer survival query system	http://www.csqs.cancer.gov/
DFS calculator for EBRT, brachytherapy and combinations of the two	http://www.prostate-cancer-radiotherapy.org.uk/calculator.htm
FinProg online	http://www.finprog.org/CM/CM2.asp?pi = 1
Nomograms for predictiong survival of GBM patients	http://www.eortc.be/tools/gbmcalculator/model1.aspx
The Han tables	http://urology.jhu.edu/prostate/hanTables.php
IBTR—breast cancer module version 2.0	http://160.109.101.132/ibtr/
Knight Cancer Institute—survival prediction tools	http://skynet.ohsu.edu/nomograms/
Lerner Research Institute—risk calculators	http://www.lerner.ccf.org/qhs/risk_calculator/
MAASTRO prediction website	http://www.predictcancer.org/
MD Anderson clinical calculators	http://www.mdanderson.org/education-and-research/resources-for-professionals/clinical-tools-and-resources/clinical-calculators/index.html
Memorial Sloan-Kettering—prediction tools	http://www.mskcc.org/cancer-care/prediction-tools
University of Montreal—nomograms	http://nomogram.org/
Mayo clinic adjuvant tool (numeracy)	http://www.mayoclinic.com/calcs/
Prognostigram	http://otooutcomes.wustl.edu/research/topics/cancer/Pages/Prognostigram.aspx
QxMD—calculate	http://www.qxmd.com/apps/calculate-by-qxmd
Calculator for estimating overall life expectancy and lifetime risk for prostate cancer death in newly diagnosed men managed without definitive local therapy	http://www.roswellpark.org/apps/prostate_cancer_estimator/

Ref: Rabin BA, Gaglio B, Sanders T, et al. (2013), *Cancer Epidemiol. Biomarkers Prev.*, 1645–1656 DOI: 10.1158/1055-9965.EPI-13-0513

A risk prediction model

- Aim of a risk prediction model:
  - to assess the prognostic ability of risk factors or the model.
- Prognosis: prediction of the course or outcome of disease
  - The course is about the disease at the population level
  - The outcome is at the individual level
- A risk prediction model is:
  - A formal combination of multiple predictors
  - Converts predictor values to an estimate of risk
  - Other names: prognostic model; prognostic index (PI)/rule
- Developmental phases:
  1. Design and model building – i.e. sample size; selection bias
     - Statistical modeling: the two cultures - Breiman L. (2001)
  2. model assessment – focus of this talk
  3. Clinical impact – i.e. utility analysis

Linear regression model:

- In linear model $Y = \beta X + \varepsilon$ where $\varepsilon \sim N(0, \sigma^2)$
  - Y : outcome, e.g. weight, X: covariates, e.g. age, sex, height
- The outcome is usually expressed in terms of:
  - Parameter estimates: $\hat{\beta}$
  - Confidence intervals (CI)
  - Model fits statistics, e.g. Chi-squared statistic
  - P-values – it can be interpreted as "a measure of surprise"
- The P-value fallacy:
  - It only answers one question: "Does an observed difference exceed that which might reasonably be expected solely as a result of sampling error and/or random allocation of individuals?" (*Colquhoun - 2014, DOI: 10.1098/rsos.140216*)
- Classical statistics tells us how to allow for uncertainty in the data. But what about uncertainty in the model?
- None of these measures provide information about the worth of the model or about the credibility of model based predictions.

Linear regression model:
Predictive ability

- In linear model $Y = \beta X + \varepsilon$ where $\varepsilon \sim N(0, \sigma^2)$
  - Y : outcome, e.g. weight, X: covariates, e.g. age, sex, height
- R^2 measures the amount of prognostic information (i.e. reduction in uncertainty):
- Uncertainty can be measured using: variance, likelihood, etc.

$$R^2 = \frac{Var(Y) - E[Var(Y|X)]}{Var(Y)}$$

$$r^2 = \frac{Var(\hat{\beta}x)}{Var(\hat{\beta}x) + {\sigma_\varepsilon}^2}$$

- R^2 properties: I) $R^2 \in [0,1]$; II) $\beta \uparrow \therefore R^2 \uparrow$
- Variance of $\hat{\beta}x$ (PI) provides vital information.
- Some only consider $Var(\hat{\beta}x)$ or functions of it, Crager (2012) or D-statistic

Assessment of a risk prediction model

Different Facets of a risk prediction model:

- Discrimination – when the outcome is event
  - The ability of model to distinguish between the high and low risk
- Calibration
  - The agreement between the observed & predicted outcomes
- Predictive ability
  - What is the amount of prognostic information that the model provide
  - Accuracy of prediction at individual level: clinical decision making

Assessment of a risk prediction model

Tools to assess a risk prediction model:

- Discrimination – both rank bases measures
  - The c-index, $c \in [0.5,1]$ (see Berrar & Flach (2011) for pitfalls)
  - The D-statistic, $D \cong \sqrt{Var(PI)}$

- Calibration
  - Calibration plot: agreement bet. observed/predic. Outcomes
  - H-L Chi-squared test

- Predictive ability - R^2-type measure
  - At the population level: disease-related
  - At individual level: clinical decision making

Predictive ability in logistic regression:

- The outcome is a binary variable $Y = [0,1]$
- The mean of Y is $\mathrm{E}(Y) = \Pr(Y = 1) = \pi$
- The model is represented by $logit(\pi|X) = \beta X$
- In a logistic regression, assessment of the predictive ability can be summarised in different ways:
- Discrimination measures
  - AUC or the c-statistic
  - D-statistic
- R^2-type measure:
  - On the probability scale: the Brier score
  - On a "latent" variable scale, i.e. $Y^* = logit(\pi \mid X)$
  - On the likelihood scale
- Each of these approaches answer different research questions.

Predictive ability in logistic regression:

- On the probability scale
  - Brier score: the squared difference between a patient's status and the predicted probability (p_i) for this patient

$$(average)\ Brier\ score = \frac{1}{n}\sum_{i=1}^{n}(Y_i - p_i)^2$$

- One can write the model as a GLM

$$Y^* = \beta X + \varepsilon$$

and $Y^* = logit(p|X)$, ε has a symmetric distribution around 0.

  - One candidate is:

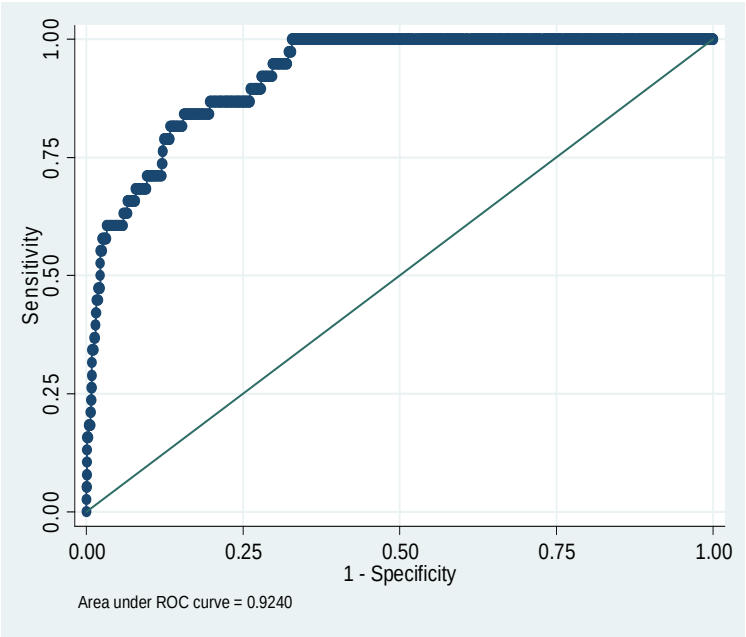
$$R^2{}_{LG} = \frac{Var(\beta X)}{Var(\beta X) + {\pi^2}/{3}}$$

  - In a Probit model ${\pi^2}/{3}$ is replaced with 1.
  - $R^2{}_{LG}$ is commonly used in social sciences

Example: child mortality for children with congenital heart disease

- Population cohort study of all children with CHD in Paris
- Outcome: Death
- Patients: 1166
- Deaths: 40
- Prognostic model: ACC-CHD, gestational age, sex, and birth weight

Item no	Measure	Estimate
1	R^2_{LG}	0.28
2	R^2_{Brier}	0.26
4	$c - index$	0.90



Predictive ability in logistic regression:

- Which measure to use:
- Use the $Brier\ score$ if the interest is in accuracy of the estimates of $\Pr(Y = 1)$ at individual level·
- Use $R^2{}_{LG}$ to quantify the amount of prognostic information in the "latent" variable model.
- Use the $c - index$ if you want to describe the capacity that the model has for distinguishing an individual who experience the event from a non-event subject.

Assessment of risk predictions in survival models:

- It is not straightforward to define appropriate tools because:
  - Censoring makes it more complicated
  - The underlying distribution of time is unknown in the Cox PH model
  - The Cox model has no error term.
- Several tools proposed, but still **no consensus**

Predictive ability in survival models:

MRC Hubs for Trials Methodology Research
MRC Clinical Trials Unit Hub

Item no	Group	Name	Author
1		R^2_{PM}	Kent & O'Quigley (1988)
2		R^2_{KS}	Korn & Simon (1990)
3		R^2_{OF}	O'Quigley & Flandre (1994)
4	Explained Variation (EV)	R^2_{AK}	Akazawa (1997)
5		R^2_{XO}	Xu & O'Quigley (2001)
6		R^2_{D}	Royston & Sauerbrei (2004)
7		R^2_{R}	Royston (2006)
8		ρ^2_{W}	Kent & O'Quigley (1988)
9		$\rho^2_{W,A}$	Kent & O'Quigley (1988)
10	Explained Randomness (ER)	ρ^2_{n}	Negelkerke (1991)
11		ρ^2_{XO}	Xu & O'Quigley (1999)
12		ρ^2_{K}	O'Quigley et al (2005)
13		V_1/V_2	Schemper (1990/1994)
14	Predictive Accuracy (PA)	$R^2_{BS}(T)$	Graf et al (1999)
15		$V_{SH}(T)$	Schemper & Henderson (2000)
16	Other	R^2_{SK}	Schemper & Kaider (1997)
17		R^2_{H}	Harrell (1986)

Research Article

Statistics in Medicine

Received 18 January 2010, Accepted 9 February 2011 Published online 26 April 2011 in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.4242

A simulation study of predictive ability measures in a survival model I: Explained variation measures

Babak Choodari-Oskooei*†, Patrick Royston and Mahesh K. B. Parmar

Statistics in Medicine

Research Article

Received 13 October 2010, Accepted 13 March 2012 Published online 5 July 2012 in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.5460

A simulation study of predictive ability measures in a survival model II: explained randomness and predictive accuracy

B. Choodari-Oskooei,*† P. Royston and Mahesh K. B. Parmar

Several R^2-type measures have been proposed to evaluate the predictive ability of a survival model. In Part I,

Predictive ability in survival models:

MRC Hubs for Trials Methodology Research
MRC Clinical Trials Unit Hub

Model based

Model based

Item no	Group	Name	Author
1		R^2_{PM}	Kent & O'Quigley (1988)
2		R^2_{KS}	Korn & Simon (1990)
3		R^2_{OF}	O'Quigley & Flandre (1994)
4	Explained Variation (EV)	R^2_{AK}	Akazawa (1997)
5		R^2_{XO}	Xu & O'Quigley (2001)
6		R^2_{D}	Royston & Sauerbrei (2004)
7		R^2_{R}	Royston (2006)
8		ρ^2_{W}	Kent & O'Quigley (1988)
9		$\rho^2_{W,A}$	Kent & O'Quigley (1988)
10	Explained Randomness (ER)	ρ^2_{n}	Negelkerke (1991)
11		ρ^2_{XO}	Xu & O'Quigley (1999)
12		ρ^2_{K}	O'Quigley et al (2005)
13		V_1/V_2	Schemper (1990/1994)
14	Predictive Accuracy (PA)	$R^2_{BS}(T)$	Graf et al (1999)
15		$V_{SH}(T)$	Schemper & Henderson (2000)
16	Other	R^2_{SK}	Schemper & Kaider (1997)
17		R^2_{H}	Harrell (1986)

Different survival C statistics: Quantifying discrimination of Framingham risk score

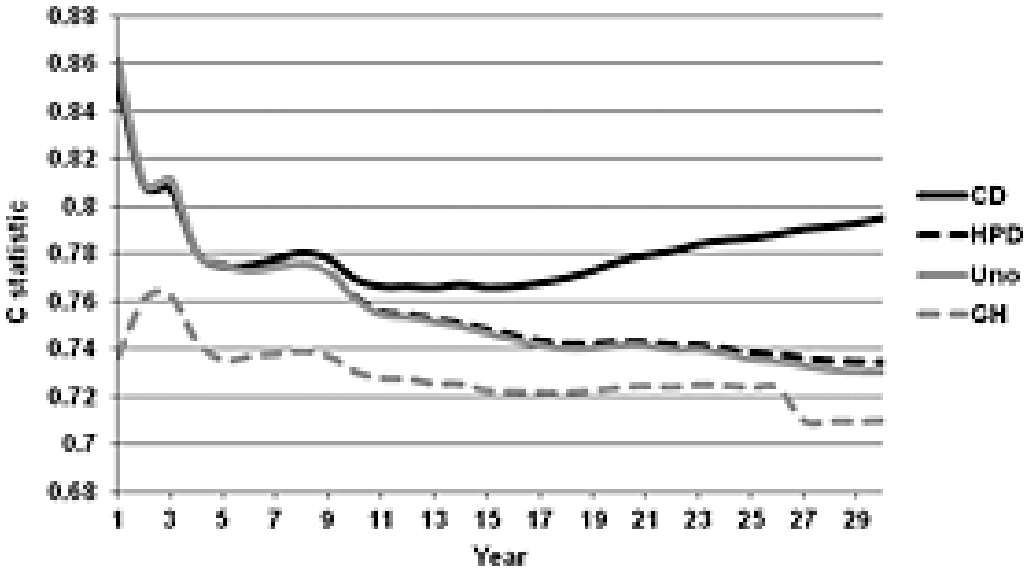
1) C_{CD}-index

2) C_{HPO}-index

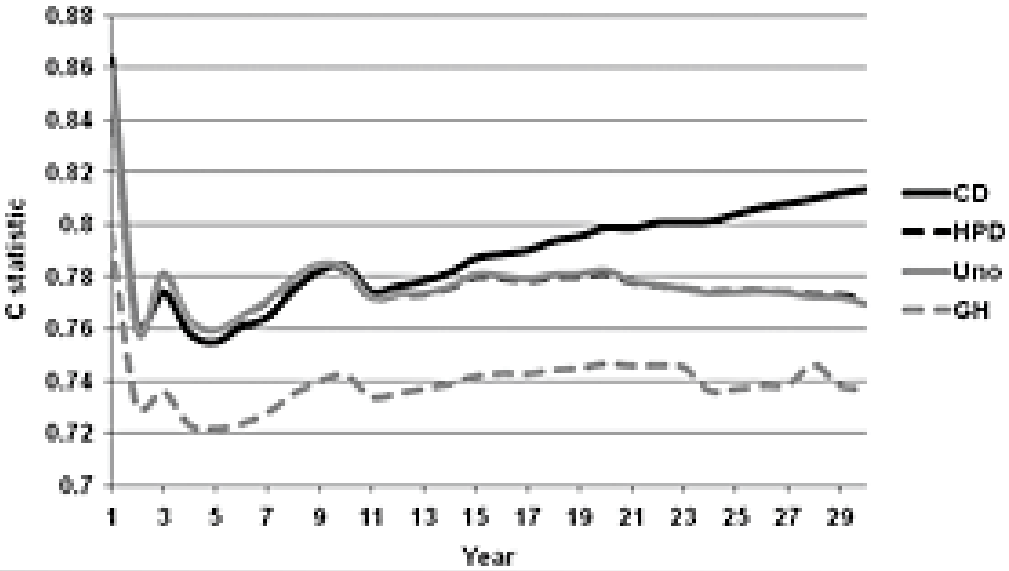
3) C_{Uno}-index

4) C_{GH}-index/k-statatistic

Men:



Women:

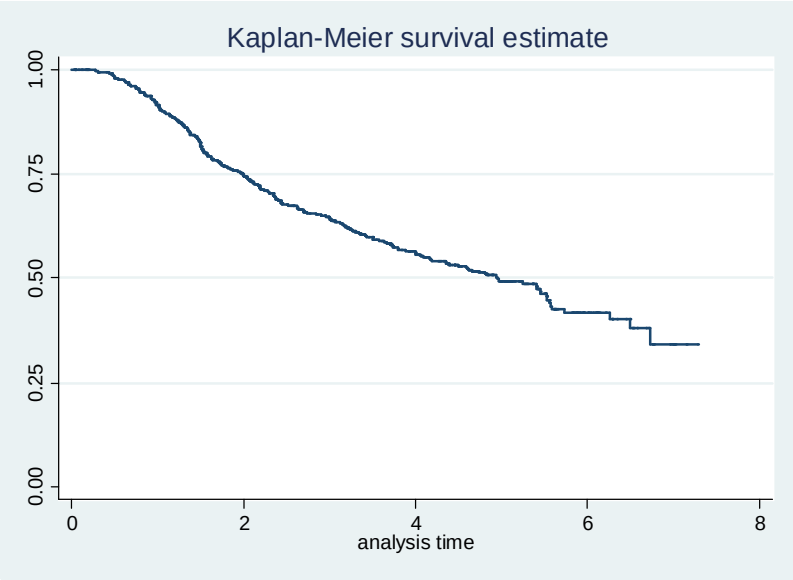


Note: In other examples the 4 estimates can differ substantially

Statistics in Medicine
Volume 31, Issue 15, pages 1543-1553, 17 FEB 2012 DOI: 10.1002/sim.4508
http://onlinelibrary.wiley.com/doi/10.1002/sim.4508/full#sim4508-fig-0002

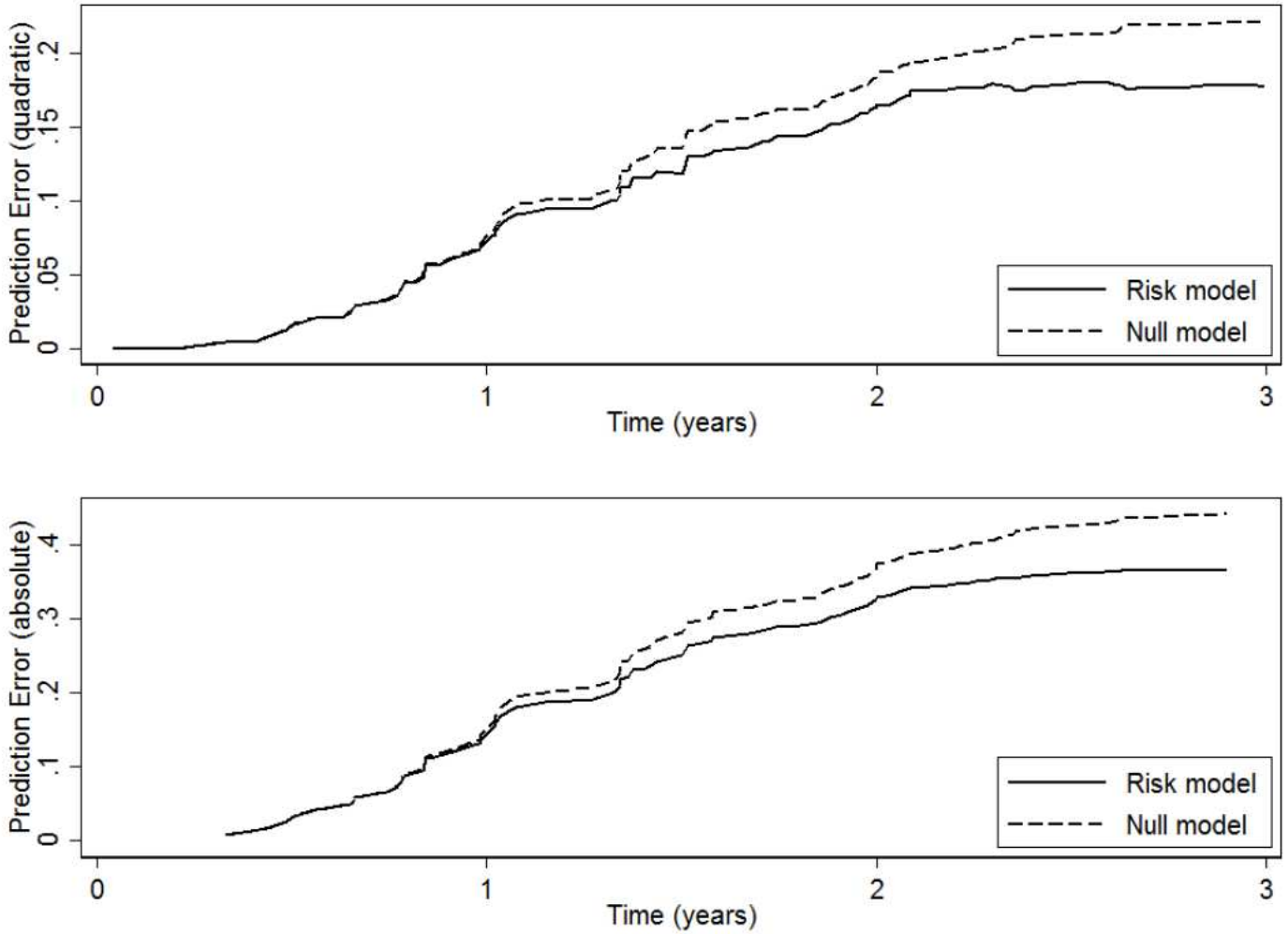
Example of a risk prediction model: breast cancer (Royston & Sau. 1999)

- Outcome: PFS
- Patients: 686
- Events: 299
- Prognostic (Cox PH) model:
  - Age; tumour grade; positive lymph nodes; progesterone receptor; hormone therapy



Measure	Estimate (95% CI)	
R^2_{PM}	0.28	(0.21 to 0.35)
R^2_{D}	0.27	(0.21 to 0.34)
$R^2_{BS}(3)$	0.19	(0.09 to 0.31)
C_H	0.70	(0.66 to 0.77)
C_U	0.70	(0.64 to 0.75)
C_{GH}	0.69	(0.62 to 0.70)
D-statistic	1.26	(0.67 to 1.32)

Prediction error in breast cancer example:



Performance in validation setting*:

MRC Hubs for Trials Methodology Research
MRC Clinical Trials Unit Hub

- Aim of the study:
  1. Investigate the performance of a developed risk model
  2. Examine the performance of the tools, e.g. censoring impact

- Internal validation: Split sample, cross validation, bootstrapping
  - 2/3 development data
  - 1/3 validation or test data

- External validation: validation data is from a different a more homogenous population
  1. Low risk profile – majority are long-term survivors
  2. High risk profile – majority are short-term survivors

*)Ambler G, Rahman MS, Choodari-Oskooei B, Omar R (2015) Performance measures for validating risk models for survival data. Submitted to the International Journal of Epidemiology,

Validation of a risk prediction model: Results on internal validation - reproducibility

MRC Hubs for Trials Methodology Research
MRC Clinical Trials Unit Hub

Censoring (%)	R^2_{PM} (SD) (0.28)	R^2_{D} (SD) (0.28)	R^2_{BS}(3) (SD) (0.19)
0	0.28 (0.04)	0.28 (0.04)	0.18 (0.04)
20	0.28 (0.04)	0.28 (0.04)	0.18 (0.04)
50	0.28 (0.05)	0.28 (0.05)	0.18 (0.05)
80	0.28 (0.07)	0.29 (0.07)	0.18 (0.08)

Censoring (%)	C_H (SD) (0.69)	C_U (SD) (0.69)	C_{GH} (SD) (0.69)	D (SD) (1.26)	CS
0	0.69 (0.02)	0.69 (0.02)	0.69 (0.01)	1.27 (0.11)	0.98 (0.10)
20	0.69 (0.02)	0.69 (0.02)	0.69 (0.01)	1.28 (0.12)	0.98 (0.11)
50	0.70 (0.02)	0.69 (0.02)	0.69 (0.02)	1.29 (0.15)	0.98 (0.13)
80	0.71 (0.04)	0.70 (0.06)	0.69 (0.02)	1.32 (0.23)	0.99 (0.18)

CS: calibration slope - the slope of the regression of the observed survival outcomes on the predicted prognostic index.

Validation of a risk prediction model: Results on external validation - transportability

MRC Hubs for Trials Methodology Research
MRC Clinical Trials Unit Hub

Risk Profile	Cens. (%)	R^2_{PM} (SD) (0.28)	R^2_{D} (SD) (0.28)	R^2_{BS}(3) (SD) (0.19)
Low	0	0.23 (0.03)	0.23 (0.03)	0.13 (0.04)
Low	20	0.23 (0.04)	0.23 (0.04)	0.13 (0.04)
Low	50	0.23 (0.05)	0.24 (0.05)	0.13 (0.04)
Low	80	0.24 (0.07)	0.26 (0.08)	0.13 (0.06)
High	0	0.25 (0.04)	0.24 (0.03)	0.16 (0.04)
High	20	0.25 (0.04)	0.24 (0.04)	0.16 (0.04)
High	50	0.25 (0.05)	0.24 (0.05)	0.16 (0.05)
High	80	0.25 (0.07)	0.25 (0.07)	0.16 (0.11)

Risk Profile	Cens. (%)	C_H (SD) (0.69)	C_U (SD) (0.69)	C_{GH} (SD) (0.69)	D (SD) (1.26)	CS
Low	0	0.67 (0.02)	0.67 (0.02)	0.67 (0.01)	1.10 (0.11)	0.98 (0.11)
Low	20	0.67 (0.02)	0.67 (0.02)	0.67 (0.01)	1.11 (0.12)	0.98 (0.12)
Low	50	0.68 (0.02)	0.67 (0.02)	0.67 (0.02)	1.14 (0.15)	0.99 (0.14)
Low	80	0.69 (0.04)	0.67 (0.06)	0.67 (0.02)	1.20 (0.24)	0.99 (0.19)
High	0	0.68 (0.02)	0.68 (0.02)	0.68 (0.01)	1.16 (0.11)	0.98 (0.11)
High	20	0.68 (0.02)	0.68 (0.02)	0.68 (0.01)	1.16 (0.12)	0.98 (0.12)
High	50	0.68 (0.02)	0.68 (0.02)	0.68 (0.02)	1.16 (0.15)	0.98 (0.14)
High	80	0.69 (0.04)	0.68 (0.06)	0.68 (0.03)	1.19 (0.23)	0.99 (0.20)

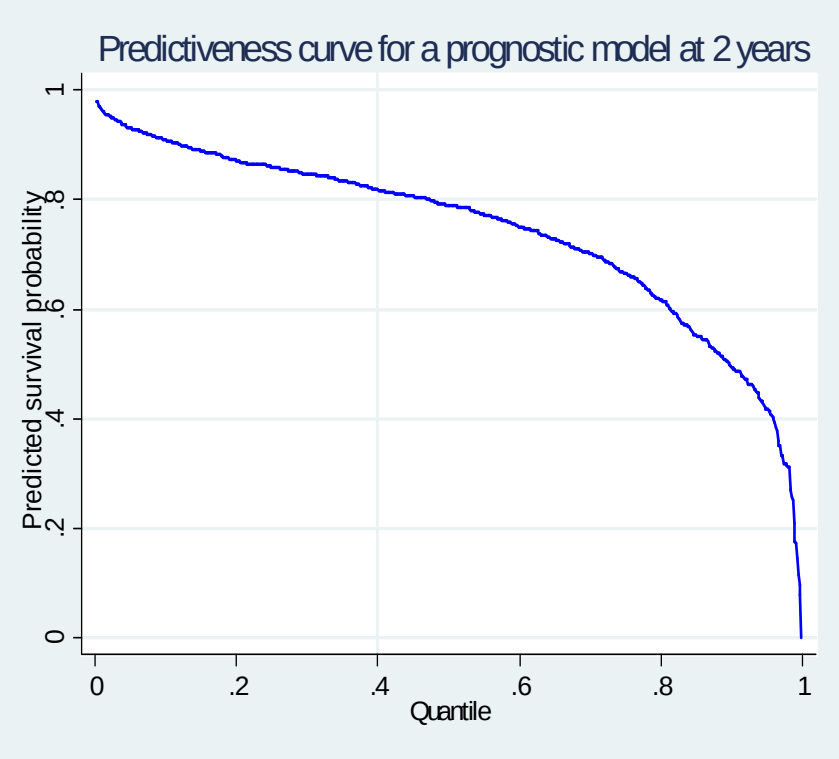
Total gain (TG) measure:

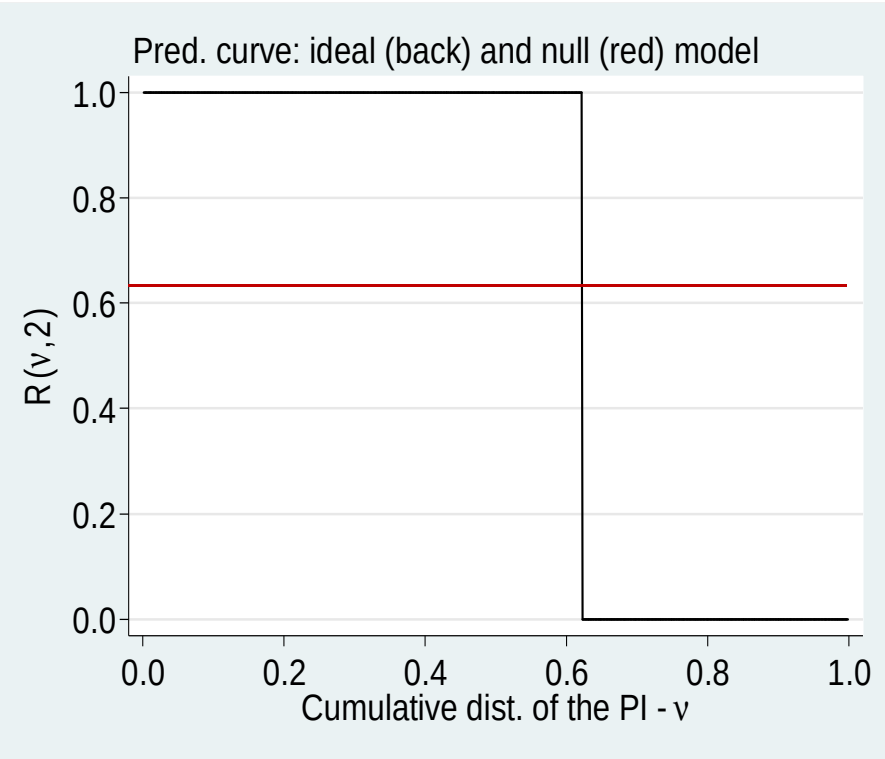
- Most existing measures of predictive ability only do not handle the case where time-dependent covariates (i.e. non-PH assumption) exist

- The existing explained variation measures only provide an estimate for the whole follow-up period

Total gain (TG) measure:
TG is based on the predictiveness curve

- Predictiveness curve is the distribution function of the predicted survival probabilities at time T.
- This gives the graph a useful interpretation
  - For example, 40% of the individuals in the data have predicted survival probabilities of more than 0.82

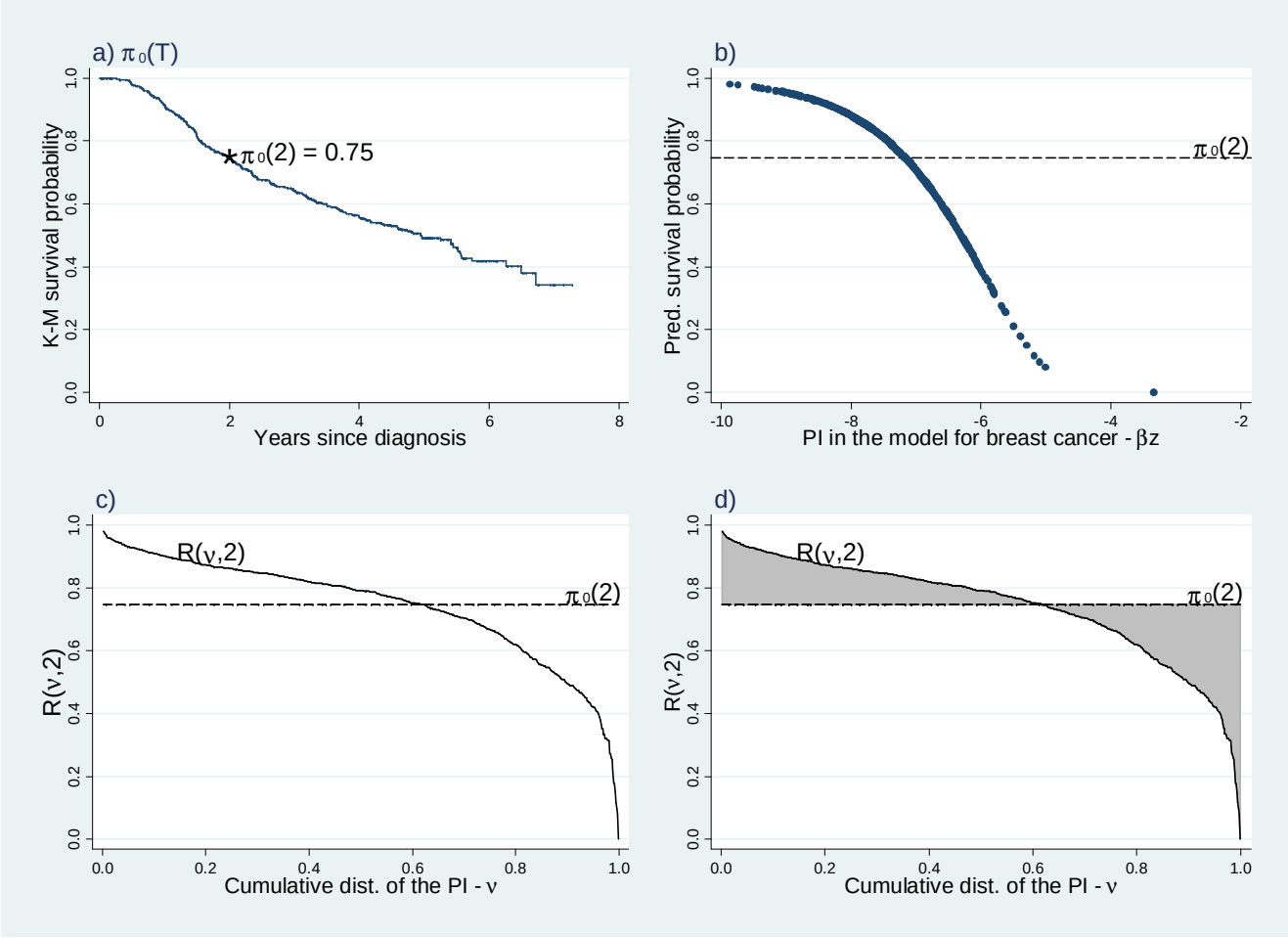




Total gain (TG) measure:
TG is based on the predictiveness curve

MRC Hubs for Trials Methodology Research
MRC Clinical Trials Unit Hub

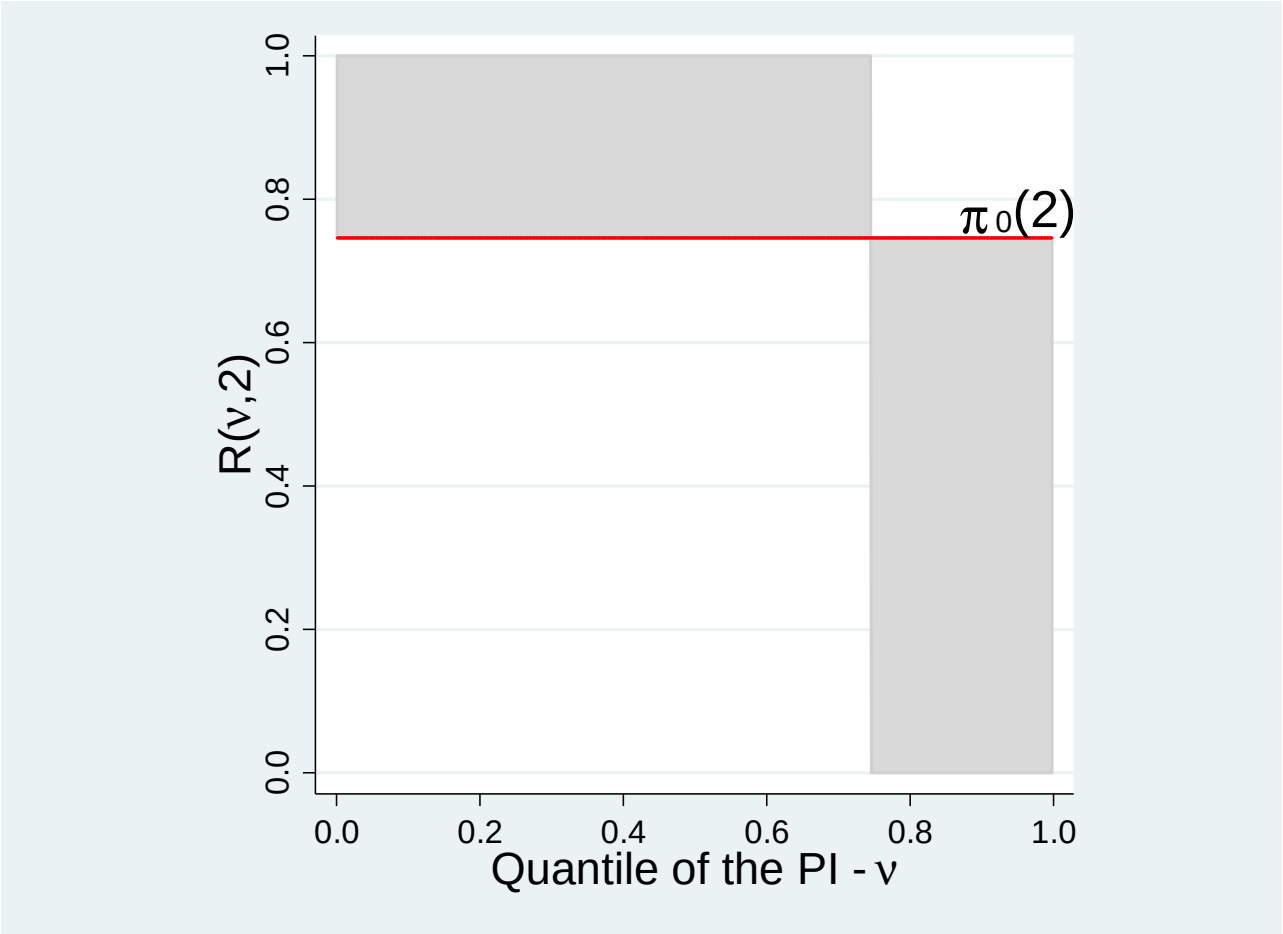
- Steps to be taken to estimate the (standardised) TG



- The shaded area is the total gain (TG) statistic.

Total gain (TG) measure:
TG is based on the predictiveness curve

- Predictiveness curve for an "ideal" prognostic survival model



Properties of $TG_{STD}(T)$:

BMC Medical Research Methodology

IMPACT FACTOR 2.27

Research article

Highly accessed Open Access

The extension of total gain (TG) statistic in survival models: properties and applications

Babak Choodari-Oskooei*, **Patrick Royston** and **Mahesh K.B. Parmar**

* Corresponding author: Babak Choodari-Oskooei b.choodari-oskooei@ucl.ac.uk

Author Affiliations

MRC Clinical Trials Unit at UCL, Aviation House, 125 Kingsway, London WC2B 6NH, UK

For all author emails, please log on.

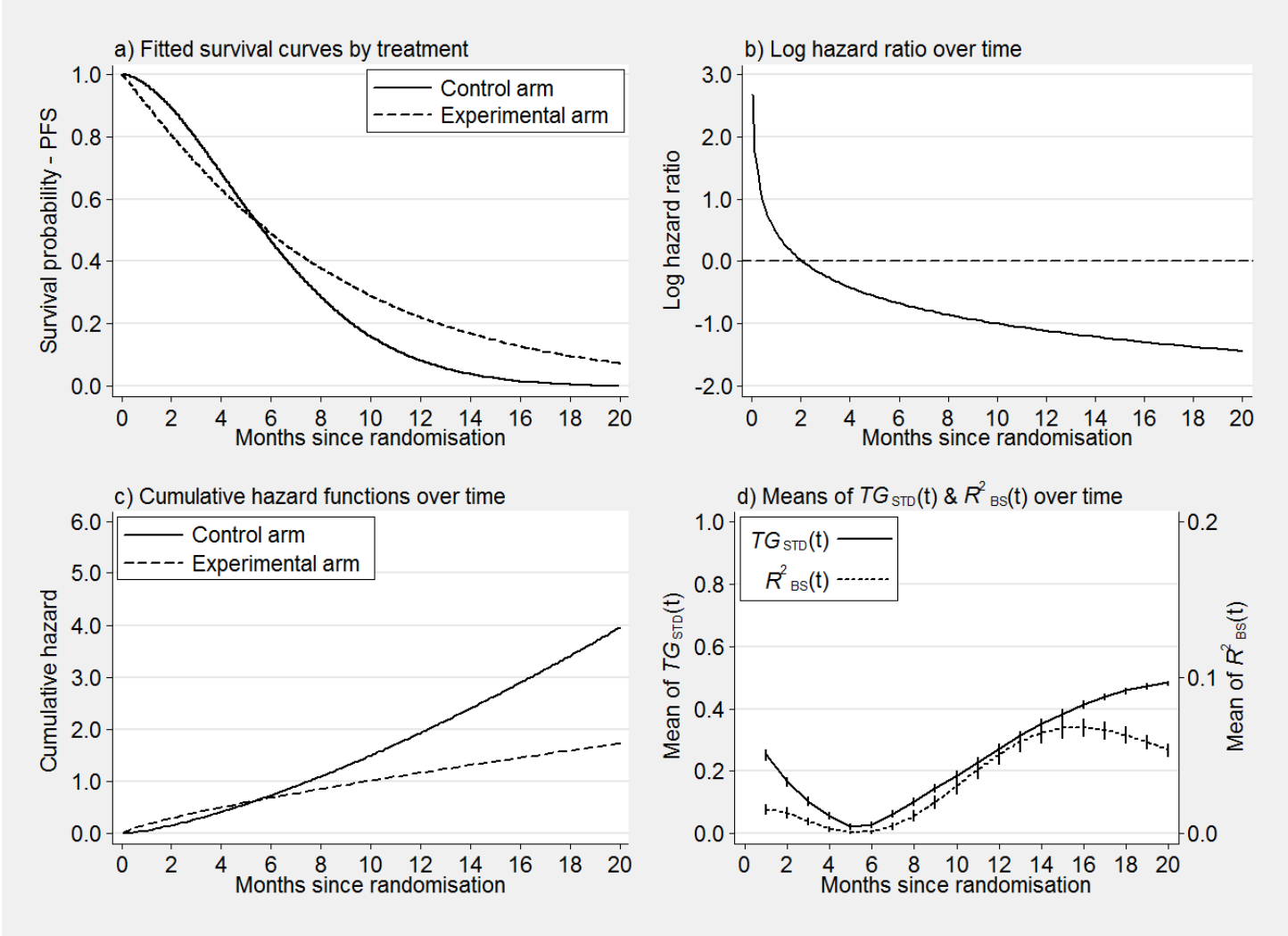
BMC Medical Research Methodology 2015, **15**:50 doi:10.1186/s12874-015-0042-x

Properties of $TG_{STD}(T)$:

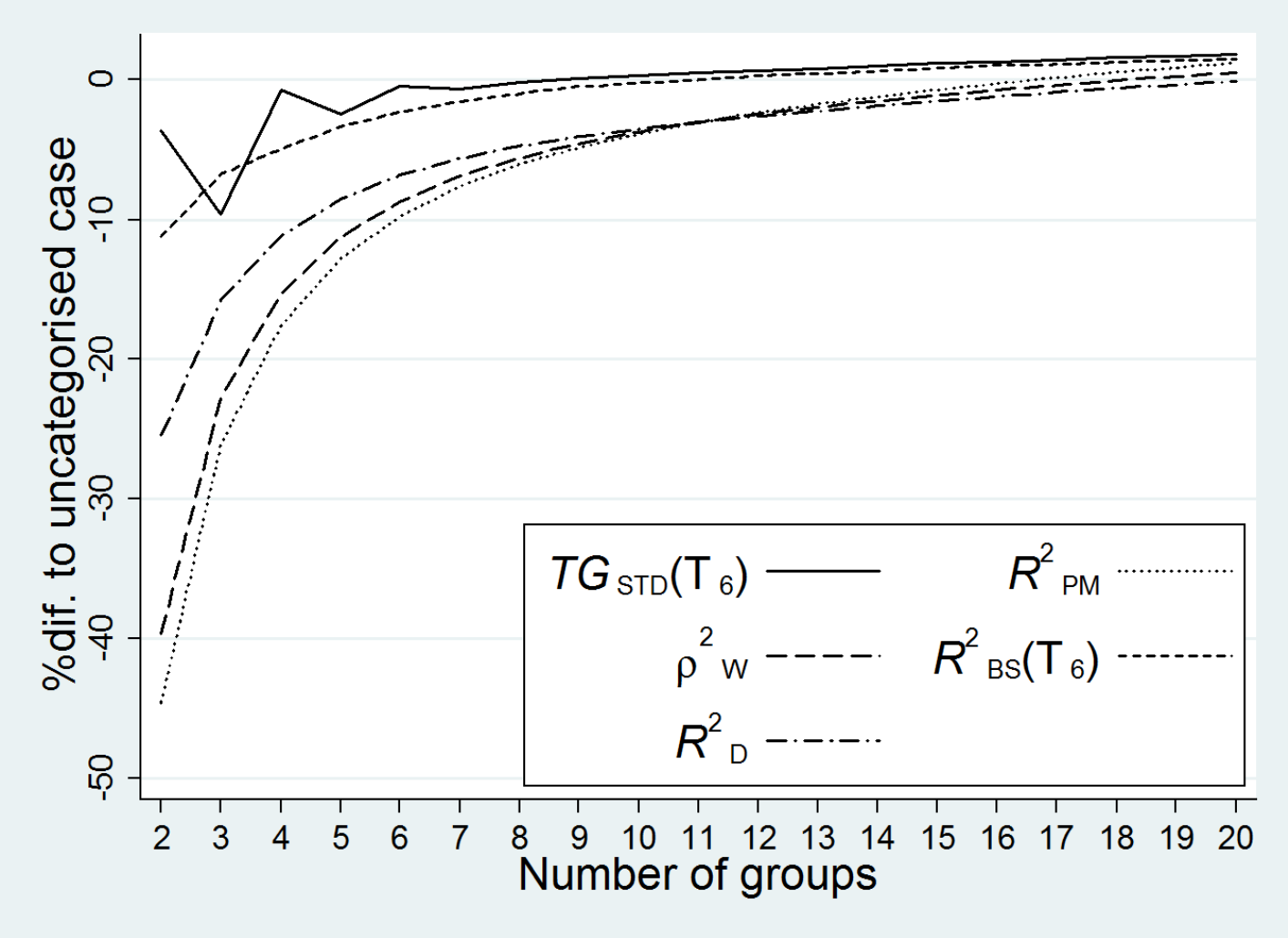
$TG_{STD}(T)$ is:

- $TG_{STD}(T) \in [0,1]$,
  - 0 means no predictive ability;
  - 1 means perfect predictive ability;
- A function of time: can deal with time-dependant covariates,
- Is not affected by random censoring,
- Is normally distributed,
- Can be extended to other survival models,

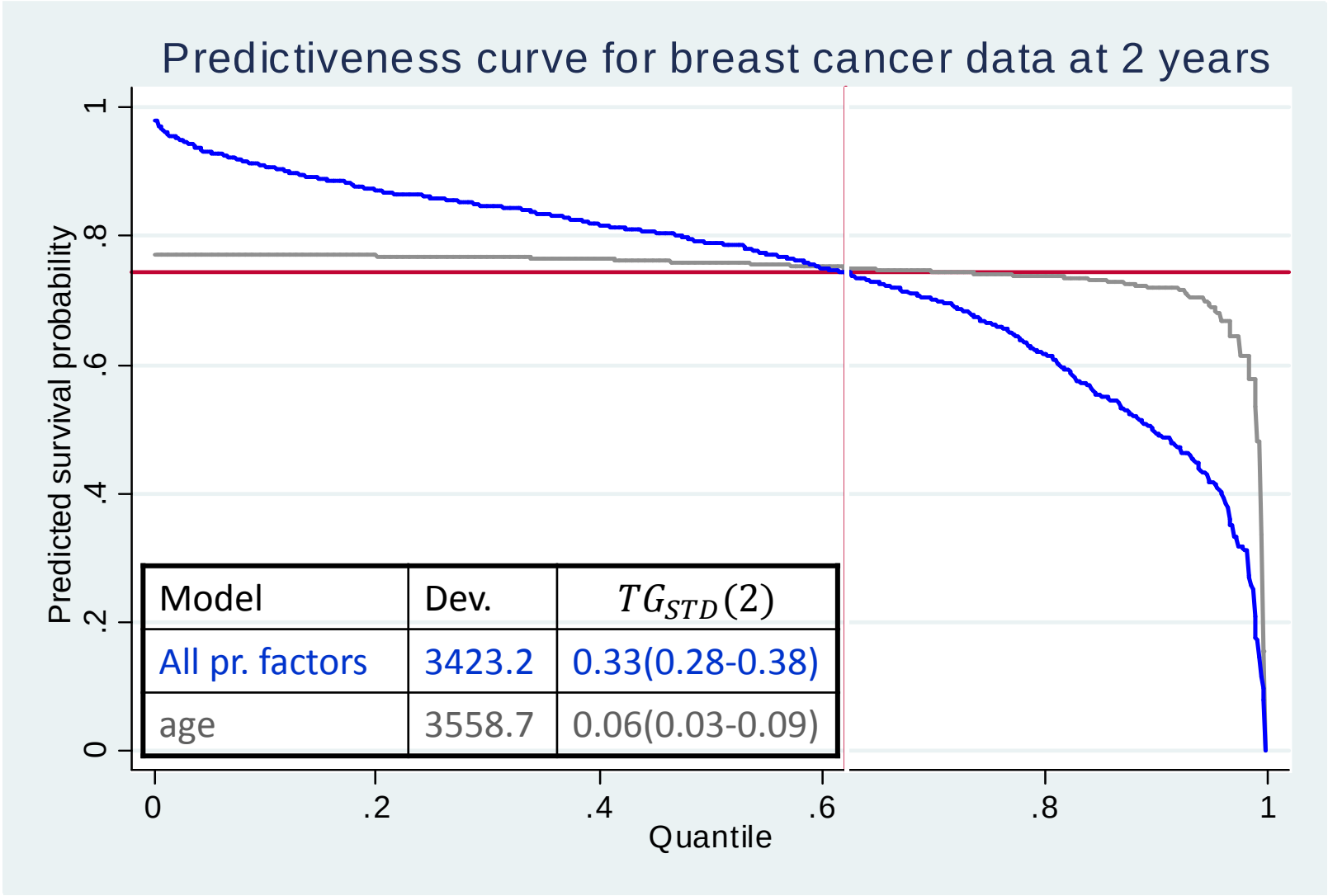
Properties of $TG_{STD}(T)$:



Properties of $TG_{STD}(T)$:



Some examples I: Breast cancer:



Some examples II: Other diseases

MRC Hubs for Trials Methodology Research
MRC Clinical Trials Unit Hub

Study	$TG_{STD}(T_2)$	R^2_{PM}	R^2_D	$R^2_{BS}(T_2)$	C_{Uno}-index
Breast cancer	0.33	0.27	0.28	0.16	0.69
	(0.28-0.38)	(0.21-0.35)	(0.21-0.35)	(0.10-0.21)	
Lymphoma	0.21	0.10	0.09	0.11	0.62
	(0.07-0.36)	(0.02-0.28)	(0.02-0.30)	(0.01-0.18)	
Lymphoma + Gene factor	0.31	0.23	0.23	0.22	0.70
	(0.18-0.44)	(0.11-0.42)	(0.11-0.40)	(0.05-0.34)	
PBC – liver disease	0.62	0.56	0.65	0.47	0.80
	(0.54-0.70)	(0.48-0.65)	(0.55-0.74)	(0.38-0.58)	
Renal cancer	0.37	0.27	0.26	0.27	0.71
	(0.31-0.42)	(0.21-0.36)	(0.20-0.33)	(0.21-0.34)	
Prostate cancer	0.24	0.13	0.13	0.11	0.63
	(0.19-0.29)	(0.09-0.20)	(0.09-0.21)	(0.06-0.15)	

Conclusions

- In most diseases, there still remains a large uncertainty regarding risk predictions at the individual level
- The existing web-tools and risk calculators should be more transparent
- They should provide more information regarding the uncertainty associated with their predicted risk
- Long-term risk predictions are less accurate than short-term
- Applying a risk prediction model to a different population will affect its predictive ability, but might not change its discrimination
- Discrimination is only part of the story. It provides little or no information on the accuracy of risk predictions
- $TG_{STD}(T)$ can be used in survival model

Future research

MRC Hubs for Trials Methodology Research
MRC Clinical Trials Unit Hub

- Design of a risk prediction study/model
  - E.g. sample size issue
  - What are the design parameters?
  - Define the "error rates" that need to be controlled?
- Repositories for risk prediction models in different diseases
  - Currently, the available information is widely dispersed!
- Comprehensive assessment of risk prediction models across different disease areas to compare the available prognostic information provided by clinical, biological, and genetic factors
- Dissemination and knowledge transfer of the available guidelines for prognostic studies in different disease areas

A relevant quote:

- "If you can measure that of which you speak, and can express it by a number, you know something of your subject, but if you cannot measure it, your knowledge is meagre and unsatisfactory."

William Thomson,
Lord Kelvin, engineer, mathematician, and physicist (1824–1907)

References

- Choodari-Oskooei B, Royston P, Parmar MKB (2015). The extension of total gain (TG) statistic in survival models: Properties and applications. *BMC Medical Research Methodology*. doi:10.1186/s12874-015-0042-x
- Choodari-Oskooei B, Royston P, Parmar MKB (2012). A simulation study of predictive ability measures in a survival model I: Explained variation measures. *Statistics in Medicine, 31 (23)*, 2627-2643. doi:10.1002/sim.4242
- Choodari-Oskooei B, Royston P, Parmar MKB (2012). A simulation study of predictive ability measures in a survival model II: explained randomness and predictive accuracy. *Statistics in Medicine, 31 (23)*, 2644-2659. doi:10.1002/sim.5460
- Ambler G, Rahman MS, Choodari-Oskooei B, Omar R (2015) Performance measures for validating risk models for survival data. Submitted to the International Journal of Epidemiology,