Using the average hazard ratio to evaluate treatment effects with non-proportional hazards

Georg Heinze

CeMSIIS – Section for Clinical Biometrics

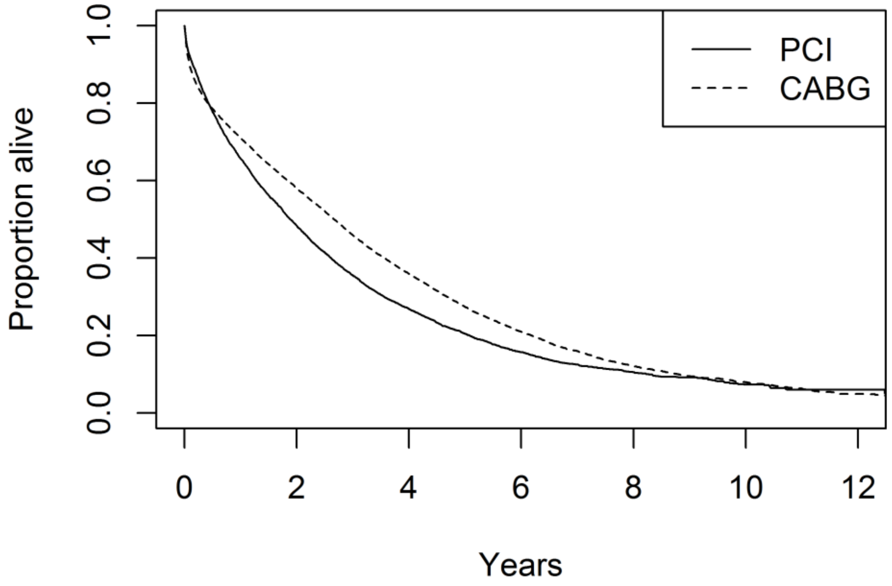
Contents

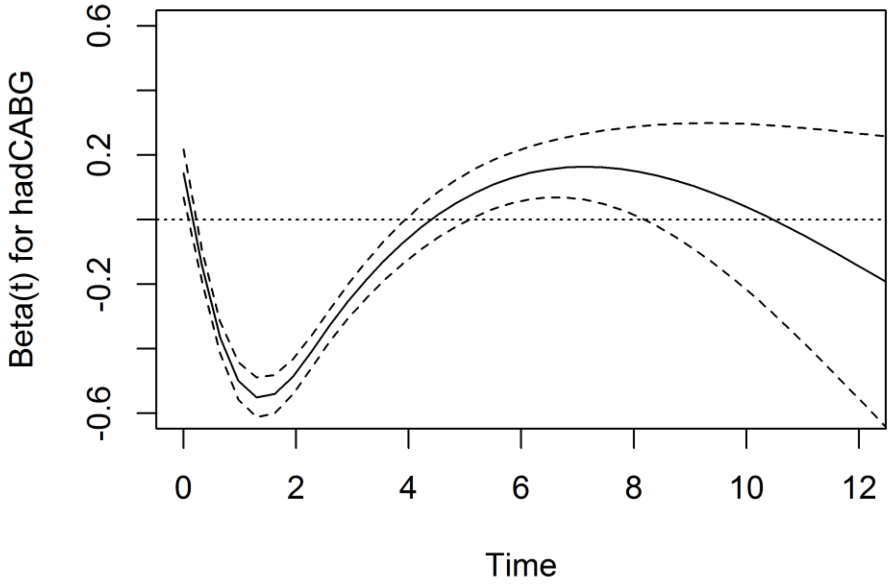
- **How to define an Average Hazard Ratio** (AHR)
  (Kalbfleisch-Prentice 1981)
- **To estimate a ratio or to test a difference?**
  (Wakounig et al 2015)
- **A pragmatic approach: weighted Cox regression**
  (Schemper et al 2009, Dunkler et al 2016)
- **Examples**
  (Chang et al 2012)

Non-proportional hazards: CABG vs. PCI in ESRD

- Chang et al, JASN 2012: Multivessel Coronary Artery Bypass Grafting Versus Percutaneous Coronary Intervention in ESRD

- N=21,981; 15,532 deaths





Definitions of Average Hazard Ratios (AHR)

- How could an AHR be defined?

- Average the hazard ratio:

$$-\int \frac{h_0(t)}{h_1(t)} dS(t)$$

- What if groups are swapped…?

Definitions of Average Hazard Ratios (AHR)

- How could an AHR be defined?

- Ratio of average hazards:

$$\frac{-\int h_0(t)dS(t)}{-\int h_1(t)dS(t)}$$

- Not straightforward to estimate…

Definitions of Average Hazard Ratios (AHR)

- How could an AHR be defined?

- Ratio of average *partial* hazards:

$$\frac{-\int [h_0(t)/h(t)]\, dS(t)}{-\int [h_1(t)/h(t)]\, dS(t)}$$

- As we will see, this definition can be estimated by a weighted Mantel–Haenszel estimator and connects to other measures of interest

AHR of Kalbfleisch and Prentice (1981)

- The weight function $-dS(t)$ reflects the relative importance of the hazard ratios in different time periods
- Kalbfleisch and Prentice considered the class $S_\alpha(t) = S_0^\alpha(t)S_1^\alpha(t)$

- For $\alpha = 1/2$:
  - weighting function is common density $f(t) = -d\{[S_0(t)S_1(t)]^{\frac{1}{2}}\}/dt$,
  - Average hazard ratio is AHR_{MH} (MH is for Mantel-Haenszel)

AHR of Kalbfleisch and Prentice (1981)

- The weight function $-dS(t)$ reflects the relative importance of the hazard ratios in different time periods
- Kalbfleisch and Prentice considered the class $S_\alpha(t) = S_0^\alpha(t) S_1^\alpha(t)$
- For $\alpha = 1$:
  - weighting function is $-dS(t)/d(t) = S_0(t)S_1(t)h(t)$ as $-dS(t) = f(t) = h(t)S(t)$
  - Average hazard ratio is AHR_{OC}, (OC is for odds of concordance)

$$\mathrm{AHR_{OC}} = \frac{-\int [h_0(t)/h(t)]\, dS(t)}{-\int [h_1(t)/h(t)]\, dS(t)} = \cdots = \frac{\mathrm{P(T_0 < T_1)}}{1 - \mathrm{P(T_0 < T_1)}} = \mathrm{OC}$$

Wakounig et al, SMMR 2015

What does it mean?

Average hazard ratio = odds of concordance

AHR/(1+AHR) = concordance probability

- This is true, even for non-proportional hazards, if using $\alpha = 1$
- The time-dependent HR might be difficult to explain
- However, the concordance probability $P(T_0 < T_1)$ is a clearcut measure of treatment effect

Under proportional hazards…

- Under proportional hazards, HR/(1+HR) = concordance probability

- Then, $\theta(t) = \frac{h_1(t)}{h_0(t)} = \theta$, and it follows $S_1(t) = S_0(t)^\theta$
- Define $c = P(T_1 < T_0) = \int f_1(t) S_0(t)$.
- Then, $\frac{c}{1-c} = \frac{P(T_1<T_0)}{P(T_0<T_1)} = \frac{\int f_1(t)S_0(t)}{\int f_0(t)S_1(t)} = \frac{\int \theta h_0(t)S_0(t)^{\theta+1}}{\int h_0(t)S_0(t)^{\theta+1}} = \theta$

Dunkler et al, Bioinformatics 2010

To estimate a ratio or to test a difference – the same thing?

- Summarize data at each of m distinct failure time points t_j:

Group	Died	Survived	Total
G_0	d_{0j}	$n_{0j} - d_{0j}$	n_{0j}
G_1	d_{1j}	$n_{1j} - d_{1j}$	n_{1j}
Total	d_j	$n_j - d_j$	n_j

- Define the weighted log-rank statistic T for no group effect on survival as

$$T = \sum_j w_j (d_{0j} - \frac{n_{0j} d_j}{n_j}) / \sqrt{V}$$

To estimate a ratio or to test a difference – the same thing?

Interestingly, T can be rewritten as:

$$T = \sum_j w_j \frac{d_{0j}(n_{1j} - d_{1j})}{n_j} - \sum_j w_j \frac{d_{1j}(n_{0j} - d_{0j})}{n_j} = A - B$$

For $w_j = 1$, this is Mantel‘s log-rank test statistic T_M, and the Mantel-Hanszel estimate of the relative risk is

$$RR_{MH} = A/B$$

To estimate a ratio or to test a difference – the same thing?

- The close relationship of $RR_{MH} = A/B$ and the Mantel log-rank statistic $T_M = A - B$ have fallen in oblivion
- For other, weighted log-rank statistics this has not even received attention:
  - Consider $w_j = \hat{S}(t_j)$ or $w_j = n_j$:

    $T = A - B$ is then the Prentice or Breslow test statistic,

    $RR = A/B$ is a corresponding average relative risk,

    weighted by ‚importance‘

    it estimates the $AHR_{OC} = OC$

Censoring

- Censoring can easily be accommodated in this scheme by inverse probability of follow-up weights:
- $\hat{G}(t)$ is the estimated probability to be still under follow-up at t,
- can be estimated by Kaplan-Meier (with reverse status indicator) or (directly) by

$$G^{-1}(t) = \frac{N(0)S(t)}{N(t)} = \frac{\text{expected N at risk}}{\text{observed N at risk}}$$

with $N(t)$ the number at risk at t

- The proposed weight is therefore $w(t) = S(t)G^{-1}(t)$

The pragmatic approach of Schemper et al 2009

- Schemper et al (StatMed 2009) proposed ‚Weighted Cox regression' to estimate AHR_{OC}
- They suggested to use Cox regression, where the contributions at each failure time are weighted by $\hat{S}(t)\hat{G}^{-1}(t)$
- In the simple two-group comparison, they use the Kaplan-Meier estimate $\hat{S}(t)$ instead of the (proper)

$$[S_0(t)f_1(t) + S_1(t)f_0(t)]/[f_0(t) + f_1(t)]$$

- The approximation error is negligible unless AHR_{OC} is very large (AHR_{OC}>4 and considerably non-proportional hazards)

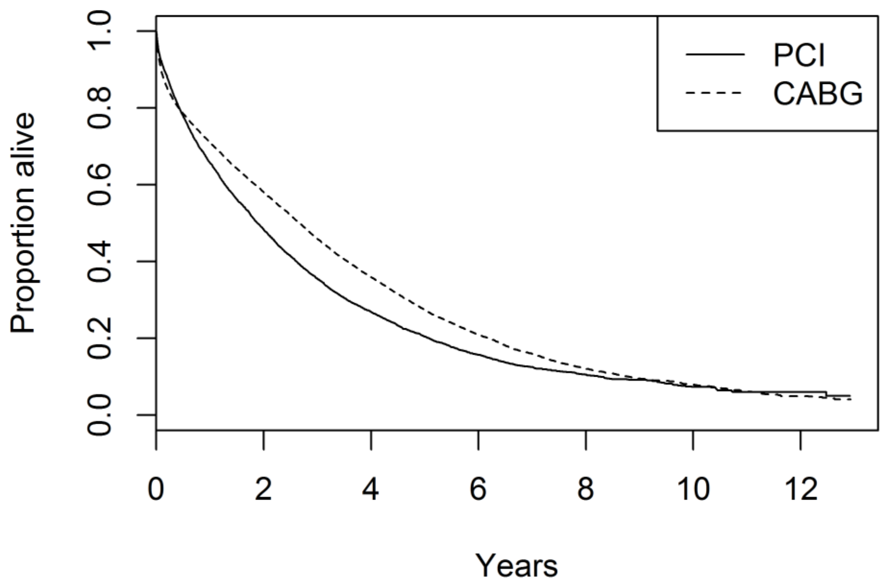
Two folklores – Mantel-Haenszel and Cox

- Mantel–Haenszel:
  - Sum up hazard in group 1
  - Sum up hazard in group 0
  - Take the ratio
- Using appropriate weights, AHR_{OC} can be estimated
- Dealt with in
  Wakounig et al, *SMMR* 2015

- Cox:
  - Partial log likelihood is a sum over risk sets
  - 'Find log HR that best fits all risk sets'
- Using appropriate weights for risk sets, AHR_{OC} can be estimated
- Dealt with in
  Schemper et al, *StatMed* 2009
  Dunkler et al, *Bioinformatics* 2010
  Dunkler et al, *JStatSoft* 2016

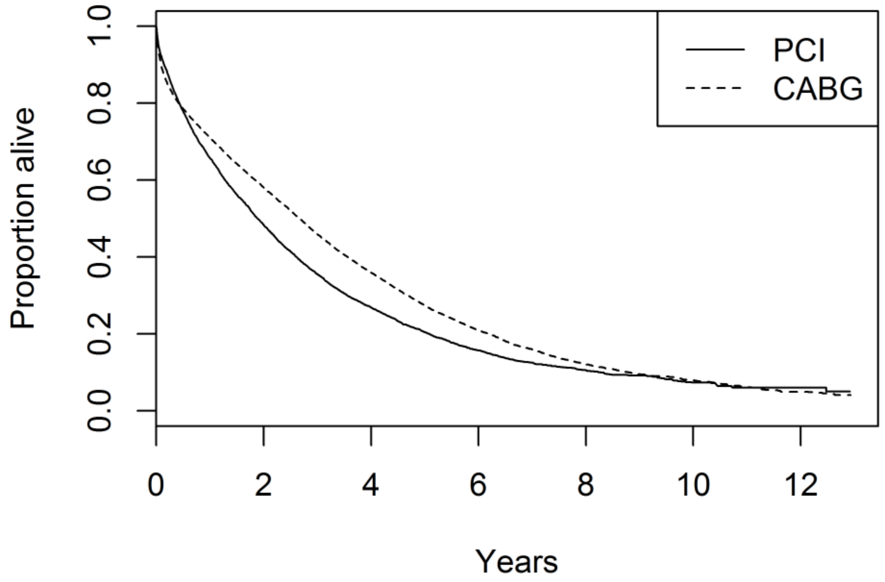
CABG vs. PCI revisited

Estimate	HR	95%CI
Cox	0.82	0.80–0.85
Weighted Cox	0.81	0.78–0.84
(30% censoring)		



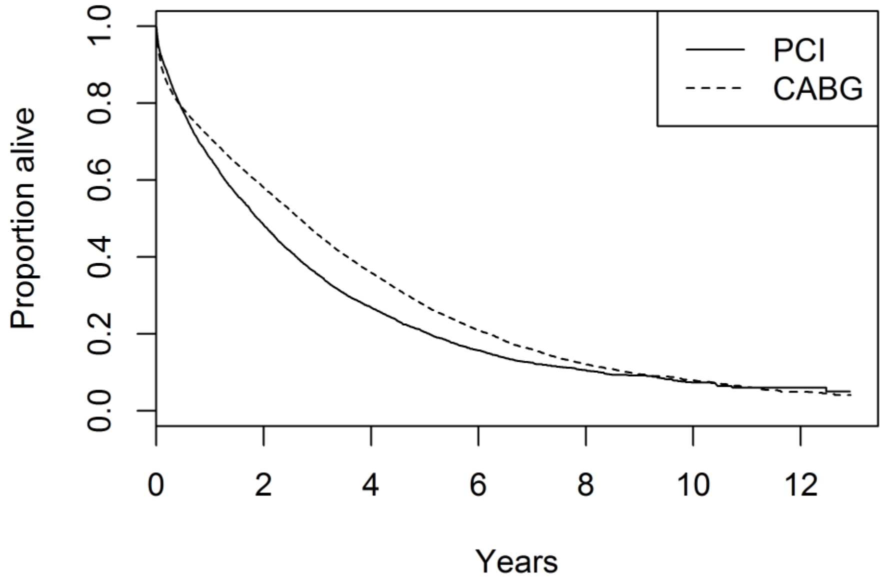
CABG vs. PCI revisited

Estimate	HR	95%CI
Cox	0.82	0.80–0.95
Weighted Cox	0.81	0.78–0.84
Cox stopped at 2 years	0.79	0.75–0.82
Weighted Cox stopped at 2 years	0.81	0.78–0.84
(58% censoring)		



CABG vs. PCI revisited

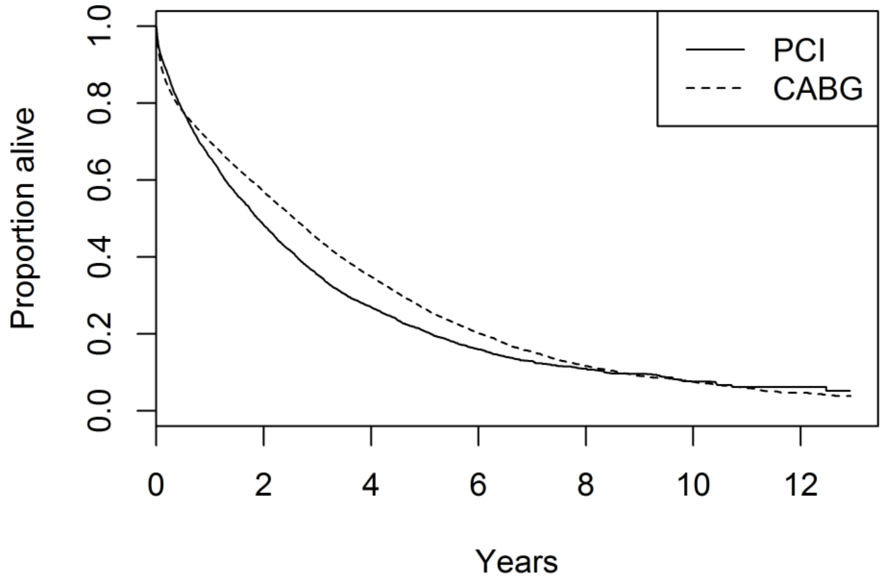
Estimate	HR	95%CI
Cox	0.82	0.80–0.95
Weighted Cox	0.81	0.78–0.84
Cox stopped at 2 years	0.79	0.75–0.82
Weighted Cox stopped at 2 years	0.81	0.78–0.84
Cox stopped at 2 years, exponential censoring	0.92	0.87–0.97
Weighted Cox stopped at 2 years, exponential censoring	0.77	0.70–0.85
(75% censoring)		



CABG vs. PCI revisited, including IPTW weights

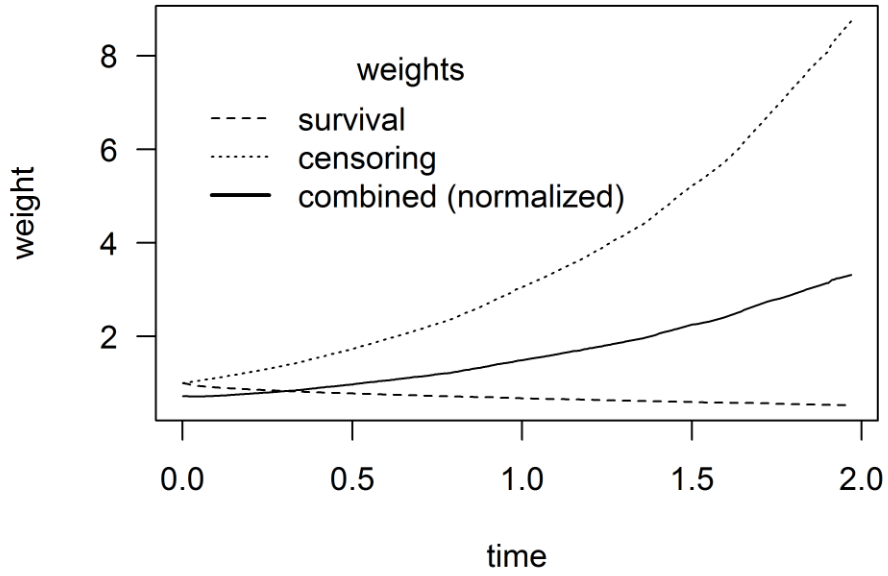
To account for confounding, analysis was weighted by inverse probability of received treatment weights

Estimate	HR	95%CI
Cox	0.85	0.83–0.87
Weighted Cox	0.86	0.83–0.89
Cox stopped at 2 years	0.81	0.79–0.84
Weighted Cox stopped at 2 years	0.84	0.80–0.88
Cox stopped at 2 years, exponential censoring	0.96	0.92–0.99
Weighted Cox stopped at 2 years, exponential censoring	0.81	0.73–0.91



Weight functions for the case ‚stopped at 2 years, exponential censoring‘

- Cox uses constant weights of 1
- For Weighted Cox, the decreasing $S(t)$ and sharply increasing $G^{-1}(t)$ weights combine into an increasing total weight



CABG vs. PCI revisited, including IPTW weights

$\Pr(T_{CABG} < T_{PCI})$ can easily be obtained by transforming the AHR:
$$\Pr(T_{CABG} < T_{PCI}) = AHR/(1 + AHR)$$

Estimate	HR	95%CI	$Pr(T_{CABG} < T_{PCI})$	95%CI
Cox	0.85	0.83–0.87		
Weighted Cox	0.86	0.83–0.89	0.462	0.453–0.471
Cox stopped at 2 years	0.81	0.79–0.84		
Weighted Cox stopped at 2 years	0.84	0.80–0.88	0.456	0.445–0.467
Cox stopped at 2 years, exponential censoring	0.96	0.92–0.99		
Weighted Cox stopped at 2 years, exponential censoring	0.81	0.73–0.91	0.448	0.422–0.475

Limitation

- Multivariable models and weighted Cox regression:
  - WCR still ‚assumes' proportional hazards
  - Misspecification of one variable can lead to biased estimates of other variables
  - Workaround provided in Dunkler et al, 2016

Conclusions (1)

- Weighted Cox regression and weighted RR can estimate AHR_{OC}
- This can then be translated in a concordance probability $\Pr(T_0 < T_1)$
- Under non-PH, the Cox HR can be very sensitive to the censoring pattern
- Often AHR by weighted Cox and HR by standard Cox coincide
  - Cox can then have efficiency advantages
  - Weighted Cox AHR estimate can be seen as ‚proof of robustness to non-PH'

Conclusions (2)

- As demonstrated, method works well for comparing treatments, either in randomized or in non-randomized studies (with IPTW weighting)
- Another extension is ‚*concordance regression*' as proposed by Dunkler et al 2010, implemented in concreg R package
  - It estimates $\Pr(T_0 < T_1)$, or its generalization $\Pr(T_{x+1} < T_x)$
  - It can also be used to estimate non-parametric c-indices for fitted survival models $\Pr(T_{x_1} < T_{x_2} | x_1 > x_2)$ -> equivalent to Uno's (2011) method

Credits to: Samo Wakounig, Michael Schemper

Heinze, Wakounig, Schemper (2007): Exploring the forest in the Lower Austrian Waldviertel under considerable hazards

Further credits

- Tara Chang, Maria E. Montez-Rath (Stanford) and Wolfgang C. Winkelmayer (Baylor College) for providing the CABG-PCI data set
- Daniela Dunkler, Samo Wakounig and Michael Schemper for having done all the work
- References:

  - Schemper M, Wakounig S, Heinze G. The estimation of average hazard ratios by weighted Cox regression. *Statistics in Medicine* 28, 2473-2489, 2009.
  - Dunkler D, Schemper M, Heinze G. Gene selection in microarray survival studies under possibly non-proportional hazards. *Bioinformatics* 26, 784-790, 2010.
  - Wakounig S, Heinze G, Schemper M. Non-parametric estimation of relative risk in survival and associated tests. *Statistical Methods in Medical Research* 24, 856-870, 2015.
  - Dunkler D, Ploner M, Heinze G. Weighted Cox regression using the R package coxphw. *Journal of Statistical Software*, 2016, to appear.
  - Chang TI, Shilane D, Kazi DS, Montez-Rath ME, Hlatky MA, Winkelmayer WC. Multivessel Coronary Artery Bypass Grafting Versus Percutaneous Coronary Intervention in ESRD. *Journal of the American Society of Nephrology* 23, 2042-2049, 2012.

www.cenisbs2017.org

Central European Network of the International Biometric Society
+ Austro-Swiss Region (ROeS)
+ German Region (IBS-DR)
+ Polish Region

International Society for Biopharmaceutical Statistics

MEDICAL UNIVERSITY OF VIENNA

CEN ISBS Vienna 2017 is a unique joint conference integrating multiple perspectives on current biostatistical research. Get in touch with your colleagues from academia, industry and regulators. Join us in Vienna!

Venue: Medical University of Vienna, Vienna, Austria
LOC-Chair: Georg Heinze, ROeS
SPC Co-Chairs: Martin Posch, ROeS | Tim Friede, IBS-DR | Frank Bretz, ISBS | Tomasz Burzykowski, Polish Region

Important Dates:
Call for topic-contributed session proposals: **October 1, 2016**
Deadline for topic-contributed session proposals: **January 15, 2017**
Registration & abstract submission opens: **January 15, 2017**
Deadline for abstract submission: **March 31, 2017**