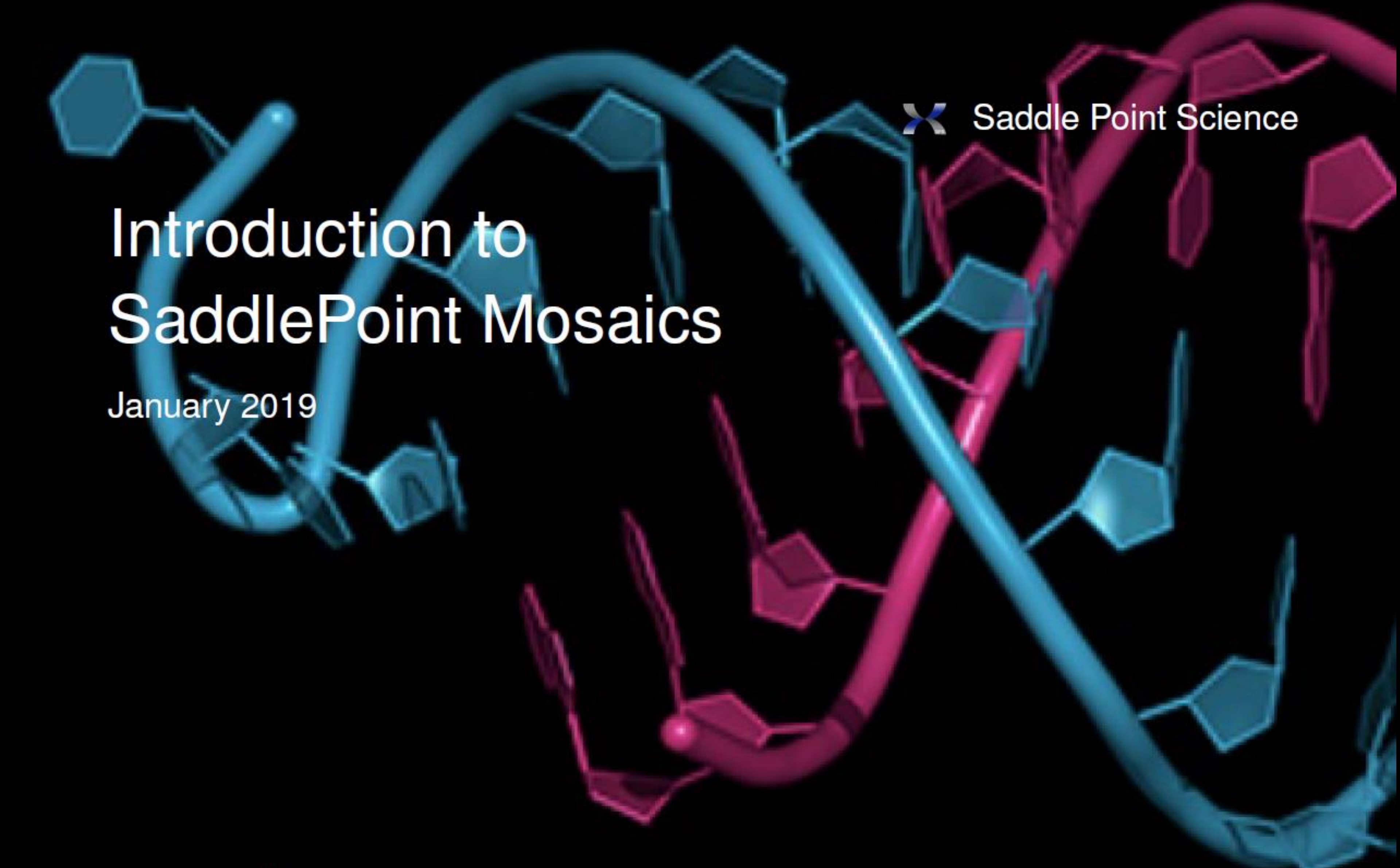


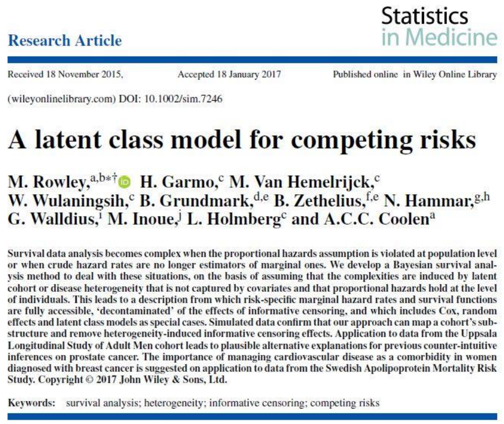
Saddle Point Science

Introduction to SaddlePoint Mosaics

January 2019

London, UK
www.saddlepointscience.com

Survival analysis for ‘complex’ cohorts


Research Article

Statistics in Medicine

Received 18 November 2015, Accepted 18 January 2017 Published online in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/sim.7246

A latent class model for competing risks

M. Rowley,[a,b*†] H. Garmo,[c] M. Van Hemelrijck,[c]
W. Wulaningsih,[c] B. Grundmark,[d,e] B. Zethelius,[f,e] N. Hammar,[g,h]
G. Walldius,[i] M. Inoue,[j] L. Holmberg[c] and A.C.C. Coolen[a]

Survival data analysis becomes complex when the proportional hazards assumption is violated at population level or when crude hazard rates are no longer estimators of marginal ones. We develop a Bayesian survival analysis method to deal with these situations, on the basis of assuming that the complexities are induced by latent cohort or disease heterogeneity that is not captured by covariates and that proportional hazards hold at the level of individuals. This leads to a description from which risk-specific marginal hazard rates and survival functions are fully accessible, ‘decontaminated’ of the effects of informative censoring, and which includes Cox, random effects and latent class models as special cases. Simulated data confirm that our approach can map a cohort’s sub-structure and remove heterogeneity-induced informative censoring effects. Application to data from the Uppsala Longitudinal Study of Adult Men cohort leads to plausible alternative explanations for previous counter-intuitive inferences on prostate cancer. The importance of managing cardiovascular disease as a comorbidity in women diagnosed with breast cancer is suggested on application to data from the Swedish Apolipoprotein Mortality Risk Study.

Keywords: survival analysis; heterogeneity; informative censoring; competing risks

Aimed at survival data where latent or disease heterogeneity and informative censoring arising from competing risks might be present

Multi-risk latent class modelling

Each class obeys the proportional hazards assumption (for each risk)

All risks modelled simultaneously for more efficient extraction of information from cohort data

Each model is defined by the following quantities:

- Number of latent classes (L)
- Personalised hazard rate complexity (M)
- Complexity of base hazard rates (K)

Model configuration

- Latent class weightings
- Frailty parameters
- Association parameters
- Base hazard rate parameters

Model parameters

Risk-specific marginal hazard rates and survival functions, “decontaminated” of the effects of informative censoring

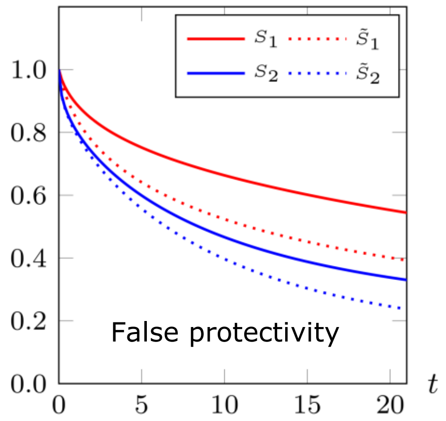
Bayesian regression and model selection

Avoid over-fitting and unnecessary model complexity

Cohort stratification

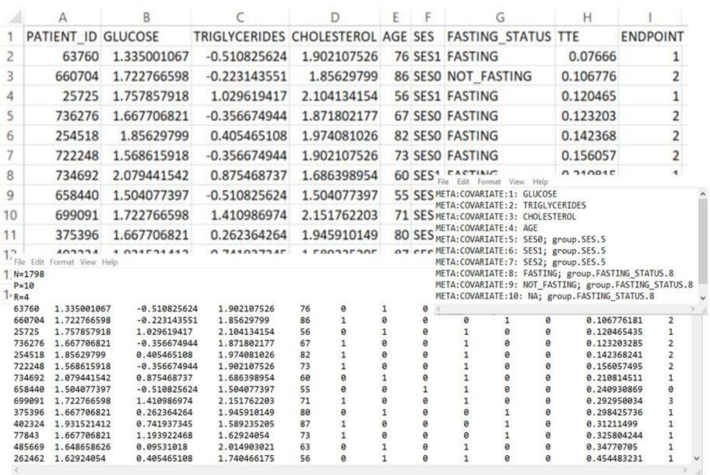
Retrospective class assignment

Provides additional insight into a cohort



Multi-risk latent class analysis / Dataset management

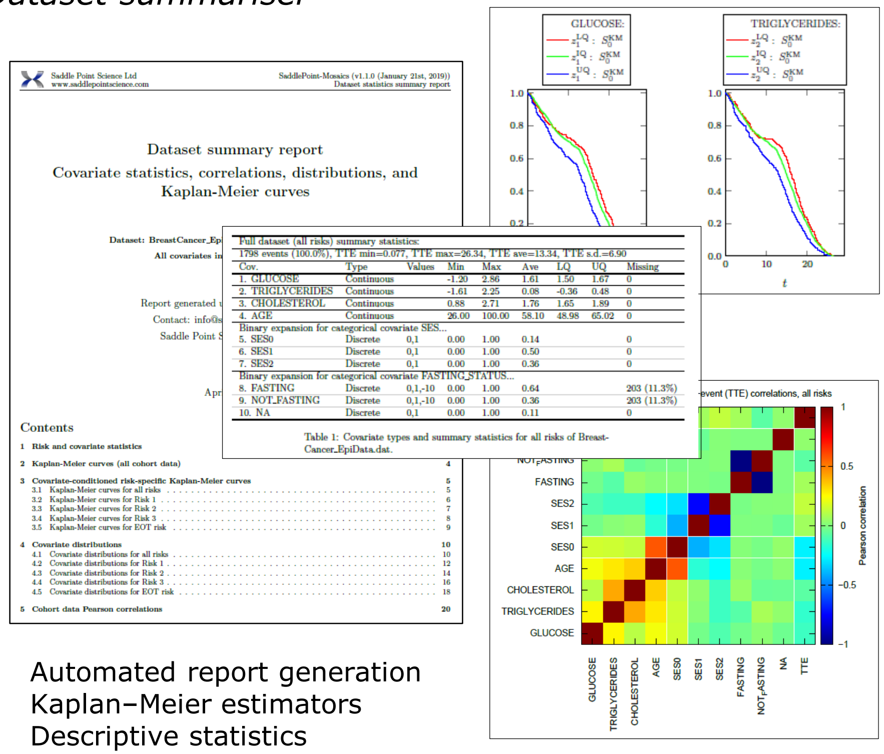
User-friendly dataset management



	A	B	C	D	E	F	G	H	I
1	PATIENT_ID	GLUCOSE	TRIGLYCERIDES	CHOLESTEROL	AGE	SES	FASTING_STATUS	TTE	ENDPOINT
2	63760	1.335001067	-0.510825624	1.902107526	76	SES1	FASTING	0.07666	1
3	660704	1.722766598	-0.223143551	1.85629799	86	SES0	NOT_FASTING	0.106776	2
4	25725	1.757857918	1.029619417	2.104134154	56	SES1	FASTING	0.120465	1
5	736276	1.667706821	-0.356674944	1.871802177	67	SES0	FASTING	0.123203	2
6	254518	1.85629799	0.405465108	1.974081026	82	SES0	FASTING	0.142368	2
7	722248	1.568615918	-0.356674944	1.902107526	73	SES0	FASTING	0.156057	2
8	734692	2.079441542	0.875468737	1.686398954	60				
9	658440	1.504077397	-0.510825624	1.504077397	55				
10	699091	1.722766598	1.410986974	2.151762203	71				
11	375396	1.667706821	0.262364264	1.945910149	80				

Automated categorical covariate expansion with user selection of baseline
Missing data regression variable generation
Automated metadata generation

Dataset summariser



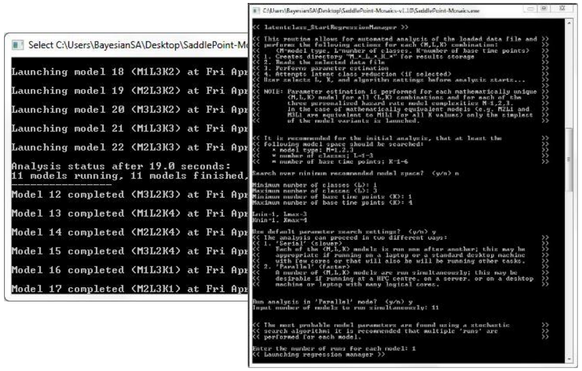
Full dataset (all risks) summary statistics:								
1798 events (100.0%), TTE min=0.077, TTE max=26.34, TTE ave=13.34, TTE s.d.=6.90								
Cov.	Type	Values	Min	Max	Ave	LQ	UQ	Missing
1. GLUCOSE	Continuous		-1.20	2.86	1.61	1.50	1.67	0
2. TRIGLYCERIDES	Continuous		-1.61	2.25	0.08	-0.36	0.48	0
3. CHOLESTEROL	Continuous		0.88	2.71	1.76	1.65	1.89	0
4. AGE	Continuous		26.00	100.00	58.10	48.98	65.02	0
Binary expansion for categorical covariate SES...								
5. SES0	Discrete	0,1	0.00	1.00	0.14			0
6. SES1	Discrete	0,1	0.00	1.00	0.50			0
7. SES2	Discrete	0,1	0.00	1.00	0.36			0
Binary expansion for categorical covariate FASTING_STATUS...								
8. FASTING	Discrete	0,1,-10	0.00	1.00	0.64			203 (11.3%)
9. NOT_FASTING	Discrete	0,1,-10	0.00	1.00	0.36			203 (11.3%)
10. NA	Discrete	0,1	0.00	1.00	0.11			0

Table 1: Covariate types and summary statistics for all risks of Breast-Cancer_EpiData.dat.

Automated report generation
Kaplan–Meier estimators
Descriptive statistics
Covariate Pearson correlations

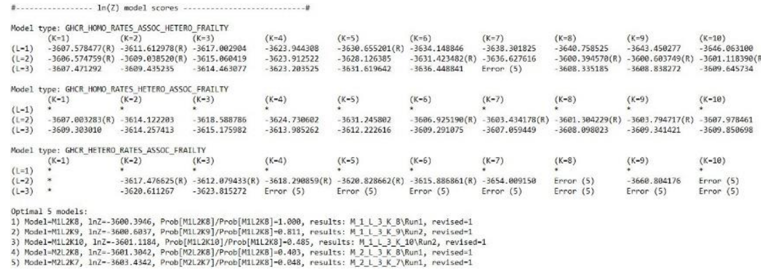
Multi-risk latent class analysis / Regression management

Automated regression management



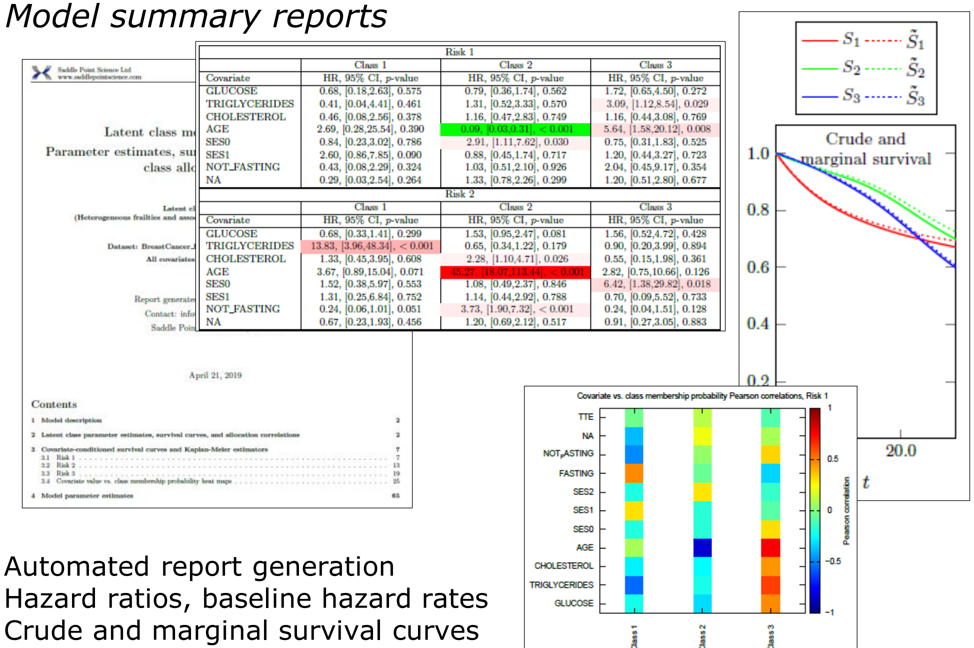
Automated regression and model ‘likelihood’ score determination for all candidate models
Baseline hazard rate(s), covariate association, and frailty estimation for each model
Bayesian priors on all parameters

Bayesian model selection



Automated Bayesian model likelihood determination and model ranking
Aikake and Bayesian Information Criterion scores also available

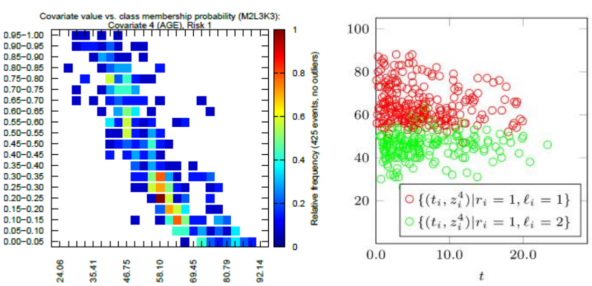
Model summary reports



Risk 1			
	Class 1	Class 2	Class 3
Covariate	HR, 95% CI, p-value	HR, 95% CI, p-value	HR, 95% CI, p-value
GLUCOSE	0.68, [0.18,2.63], 0.575	0.79, [0.36,1.74], 0.562	1.72, [0.65,4.50], 0.272
TRIGLYCERIDES	0.41, [0.04,4.41], 0.461	1.31, [0.52,3.33], 0.570	3.09, [1.12,8.54], 0.029
CHOLESTEROL	0.46, [0.08,2.56], 0.378	1.16, [0.47,2.83], 0.749	1.16, [0.44,3.08], 0.769
AGE	2.69, [0.28,25.54], 0.390	0.09, [0.03,0.31], < 0.001	5.64, [1.58,20.12], 0.008
SES0	0.84, [0.23,3.02], 0.786	2.91, [1.11,7.62], 0.030	0.75, [0.31,1.83], 0.525
SES1	2.60, [0.86,7.85], 0.090	0.88, [0.45,1.74], 0.717	1.20, [0.44,3.27], 0.723
NOT_FASTING	0.43, [0.08,2.29], 0.324	1.03, [0.51,2.10], 0.926	2.04, [0.45,9.17], 0.354
NA	0.29, [0.03,2.54], 0.264	1.33, [0.78,2.26], 0.299	1.20, [0.51,2.80], 0.677
Risk 2			
	Class 1	Class 2	Class 3
Covariate	HR, 95% CI, p-value	HR, 95% CI, p-value	HR, 95% CI, p-value
GLUCOSE	0.68, [0.33,1.41], 0.299	1.53, [0.95,2.47], 0.081	1.56, [0.52,4.72], 0.428
TRIGLYCERIDES	13.83, [3.96,48.34], < 0.001	0.65, [0.34,1.22], 0.179	0.90, [0.20,3.99], 0.894
CHOLESTEROL	1.33, [0.45,3.95], 0.608	2.28, [1.10,4.71], 0.026	0.55, [0.15,1.98], 0.361
AGE	3.67, [0.89,15.04], 0.071	45.27, [18.07,113.44], < 0.001	2.82, [0.75,10.66], 0.126
SES0	1.52, [0.38,5.97], 0.553	1.08, [0.49,2.37], 0.846	6.42, [1.38,29.82], 0.018
SES1	1.31, [0.25,6.84], 0.752	1.14, [0.44,2.92], 0.788	0.70, [0.09,5.52], 0.733
NOT_FASTING	0.24, [0.06,1.01], 0.051	3.73, [1.90,7.32], < 0.001	0.24, [0.04,1.51], 0.128
NA	0.67, [0.23,1.93], 0.456	1.20, [0.69,2.12], 0.517	0.91, [0.27,3.05], 0.883

Automated report generation
Hazard ratios, baseline hazard rates
Crude and marginal survival curves
Covariate Pearson correlations

Retrospective class allocation

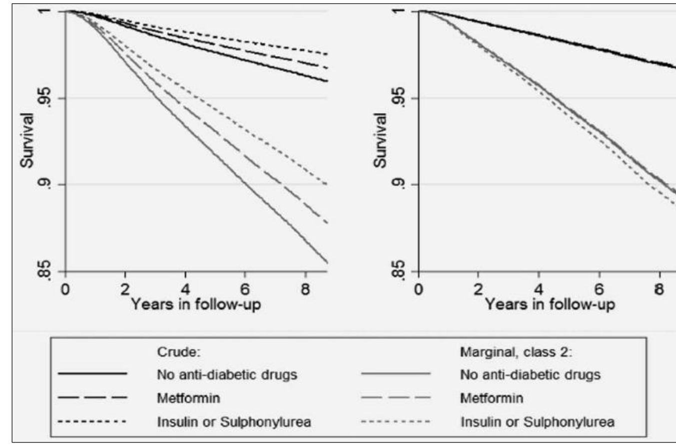


Allocate patients to most probable latent class
Data-driven cohort stratification
Covariate and class membership correlations
Aids search for new informative biomarkers

Some application examples

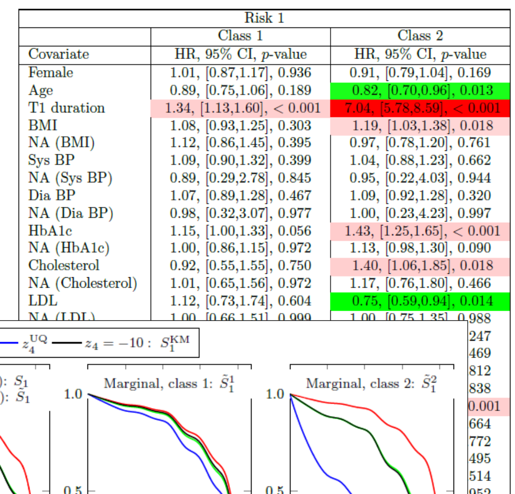
Prostate cancer:
Heterogeneity and competing risks effects found

Haggstrom et al., "Heterogeneity in risk of prostate cancer: A Swedish population-based cohort study of competing risks and Type 2 diabetes mellitus", IJC, 2018



Colorectal cancer:
Re-analysis of COIN trail (NEJM, 2011)

PR Barber et al., under review, 2019



	Risk 1	
	Class 1	Class 2
Covariate	HR, 95% CI, *p*-value	HR, 95% CI, *p*-value
Female	1.01, [0.87,1.17], 0.936	0.91, [0.79,1.04], 0.169
Age	0.89, [0.75,1.06], 0.189	0.82, [0.70,0.96], 0.013
T1 duration	1.34, [1.13,1.60], < 0.001	7.04, [5.78,8.59], < 0.001
BMI	1.08, [0.93,1.25], 0.303	1.19, [1.03,1.38], 0.018
NA (BMI)	1.12, [0.86,1.45], 0.395	0.97, [0.78,1.20], 0.761
Sys BP	1.09, [0.90,1.32], 0.399	1.04, [0.88,1.23], 0.662
NA (Sys BP)	0.89, [0.29,2.78], 0.845	0.95, [0.22,4.03], 0.944
Dia BP	1.07, [0.89,1.28], 0.467	1.09, [0.92,1.28], 0.320
NA (Dia BP)	0.98, [0.32,3.07], 0.977	1.00, [0.23,4.23], 0.997
HbA1c	1.15, [1.00,1.33], 0.056	1.43, [1.25,1.65], < 0.001
NA (HbA1c)	1.00, [0.86,1.15], 0.972	1.13, [0.98,1.30], 0.090
Cholesterol	0.92, [0.55,1.55], 0.750	1.40, [1.06,1.85], 0.018
NA (Cholesterol)	1.01, [0.65,1.56], 0.972	1.17, [0.76,1.80], 0.466
LDL	1.12, [0.73,1.74], 0.604	0.75, [0.59,0.94], 0.014
NA (LDL)	1.00, [0.66,1.51], 0.999	1.00, [0.75,1.35], 0.988

Diabetic retinopathy:
Heterogeneity in disease severity found for patients with Type 1 diabetes mellitus

Larsen et al., under review, 2019