Model comparison for analysis of population surveillance data

Rosie Meng
Richard Woodman
Steven Coles
Erin Symonds

Email: rosie.meng@flinders.edu.au
Stata User Group Meeting
Canberra, 24-25 September 2015

Overview

- Research questions & background
- Data
- Models
- Stata routines

Research question

- What characteristics of colorectal adenoma diagnosed at index colonoscopy are associated with degree of neoplasia advancement at 1st surveillance colonoscopy?

Application: which patients diagnosed with adenoma at index would not be at significant risk of developing an advanced neoplasia at 1st surveillance colonoscopy - allow colonoscopy interval to be extended.

Research question background

- Uncertainty and deviation from surveillance guideline regarding surveillance interval.
- Few studies that provide evidence for duration of surveillance based on lesion features.
- Conservative approach is common:

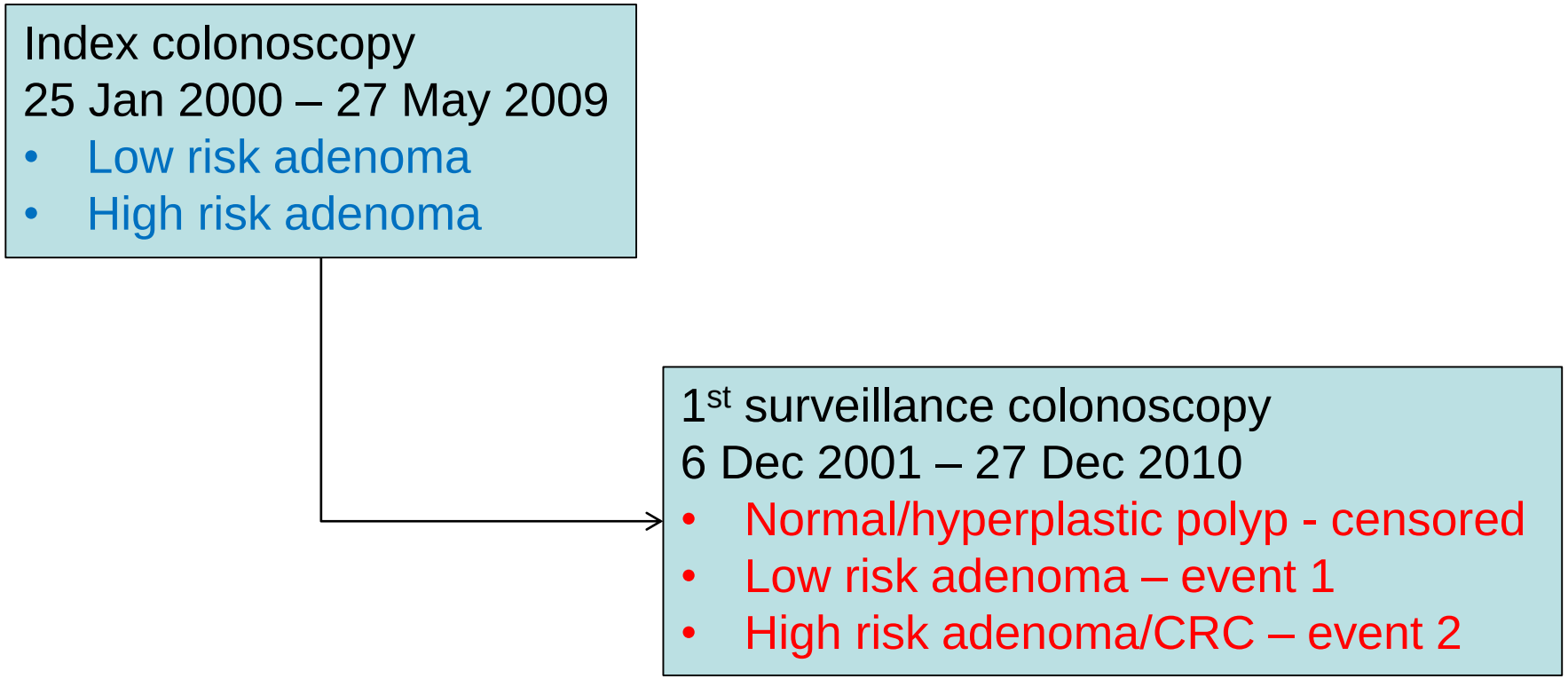
  evidence based professional guidelines

  VERSUS

  specialist's preference to minimize chance of a future finding of advanced neoplasia
- Consequences: surveillance colonoscopy interval shortened, colonoscopy service overloaded, and increased risk of complication (e.g. bowel perforation).

Data

- Database: South Australian Southern Cooperative Program for the Prevention of Colorectal Cancer (SCOOP)
- Study period: 25 Jan 2000 – 21 Dec 2010 (n=379)



Data cont.

- Study cohort at index colonoscopy (379 subjects)
  - Low risk adenoma (n=187)
  - High risk adenoma (n=192)
- Outcomes at 1st surveillance colonoscopy:
  - Normal/hyperplastic polyp
  - Low risk adenoma
  - High risk adenoma/CRC
- Predictors
  - Time between two colonoscopies
  - Risk category at the index
  - Gender
  - Age at the index
  - Reason for the index colonoscopy
  - Reason for the 1st surveillance colonoscopy

Data cont.

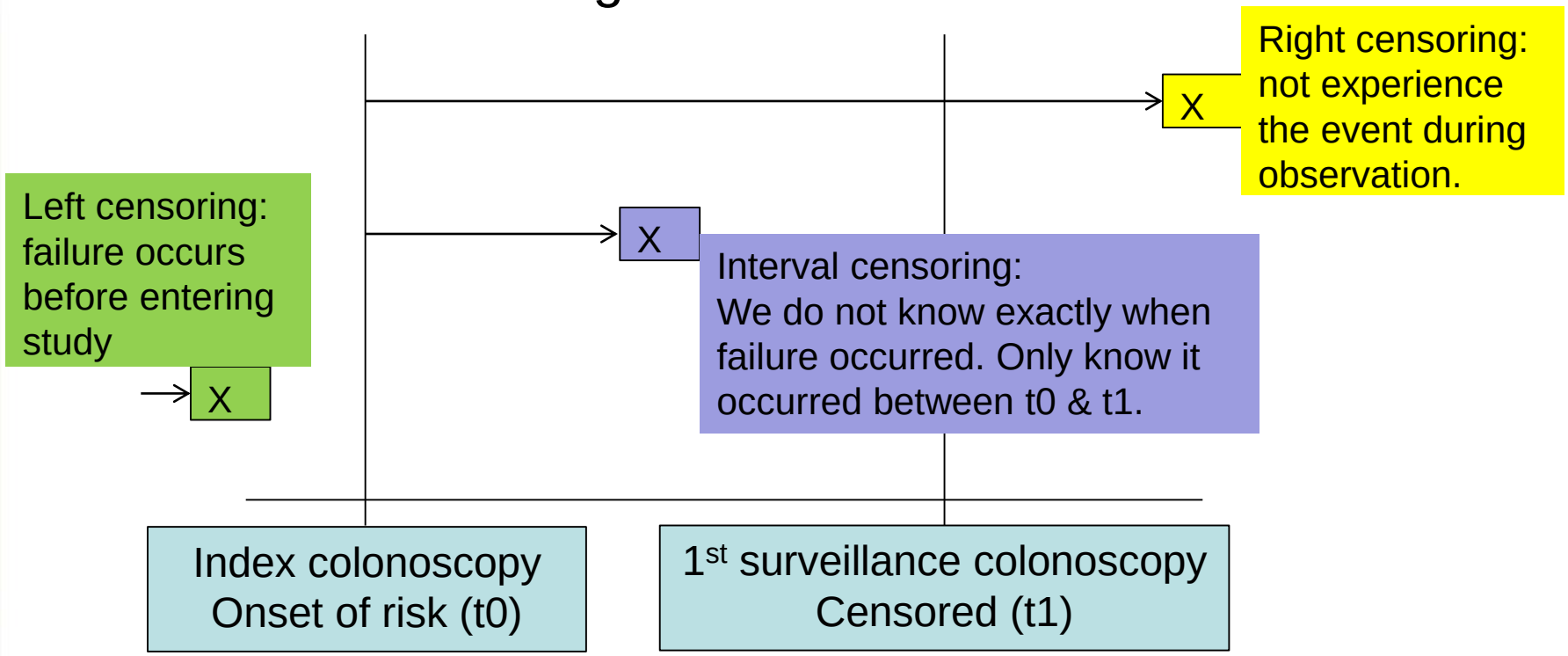
Risk grouping:

- High risk adenoma has one or more following features
  - ≥10mm size
  - High grade dysplasia
  - Villous or serrated morphology
  - ≥3 polyps
- Low risk adenoma - all patients with a diagnosis of adenoma other than high risk adenoma.

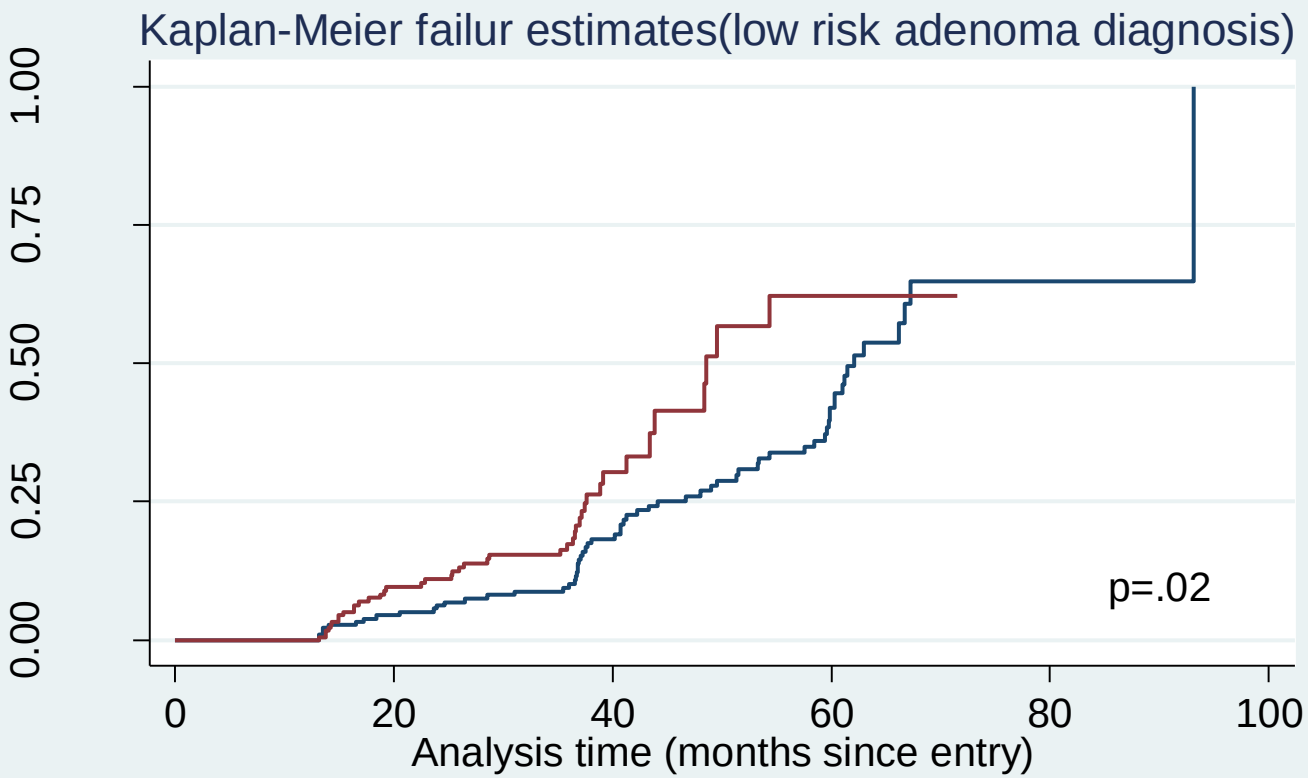
Data cont. - Censoring

- Right censoring – most common type
- Interval censoring
- Left censoring

} No difference mathematically



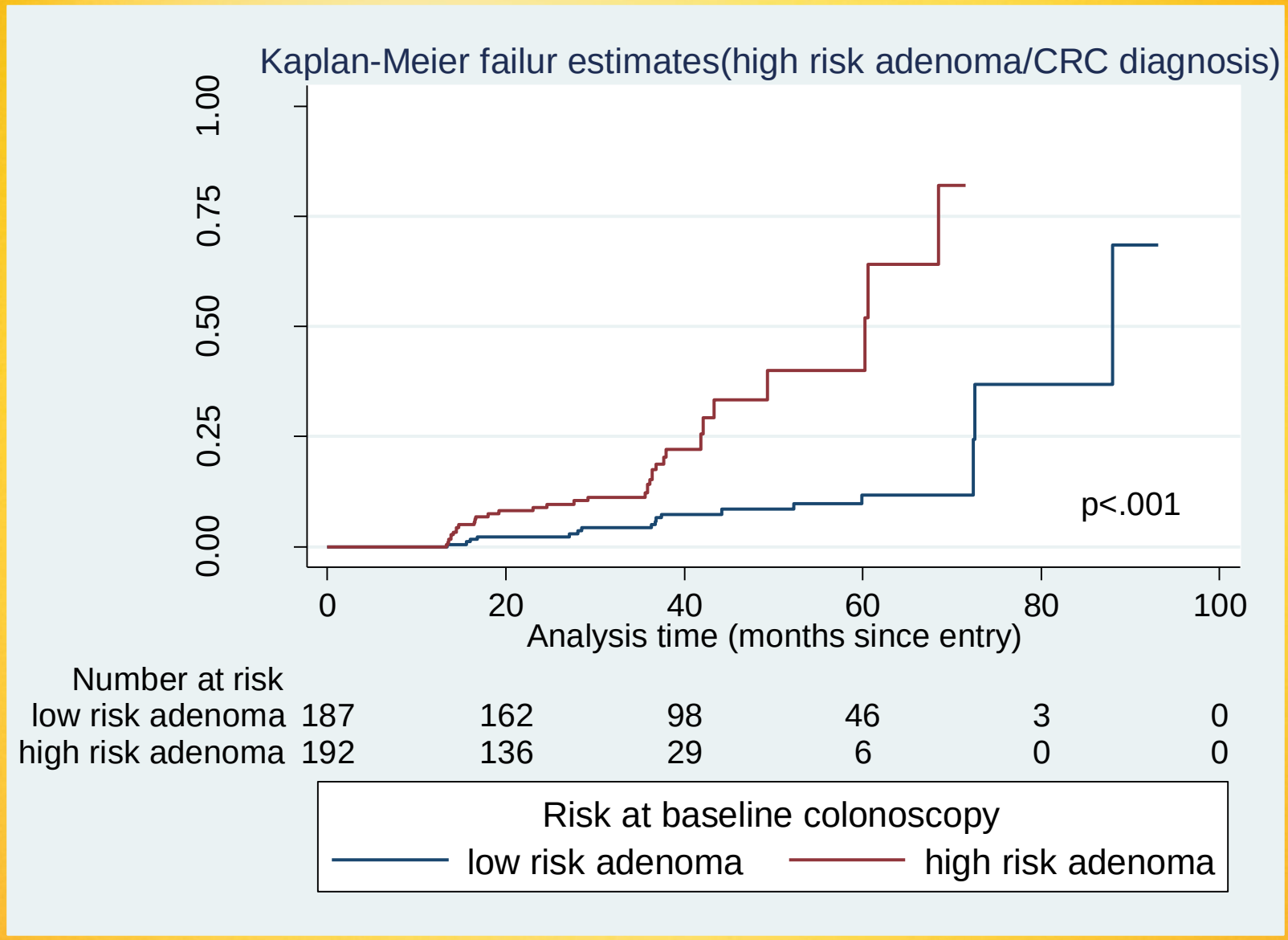
Results – risk of low risk adenoma diagnosis



Number at risk						
low risk adenoma	187	162	98	46	3	0
high risk adenoma	192	136	29	6	0	0

Risk at baseline colonoscopy
low risk adenoma
high risk adenoma

Results – risk of high risk adenoma/CRC diagnosis

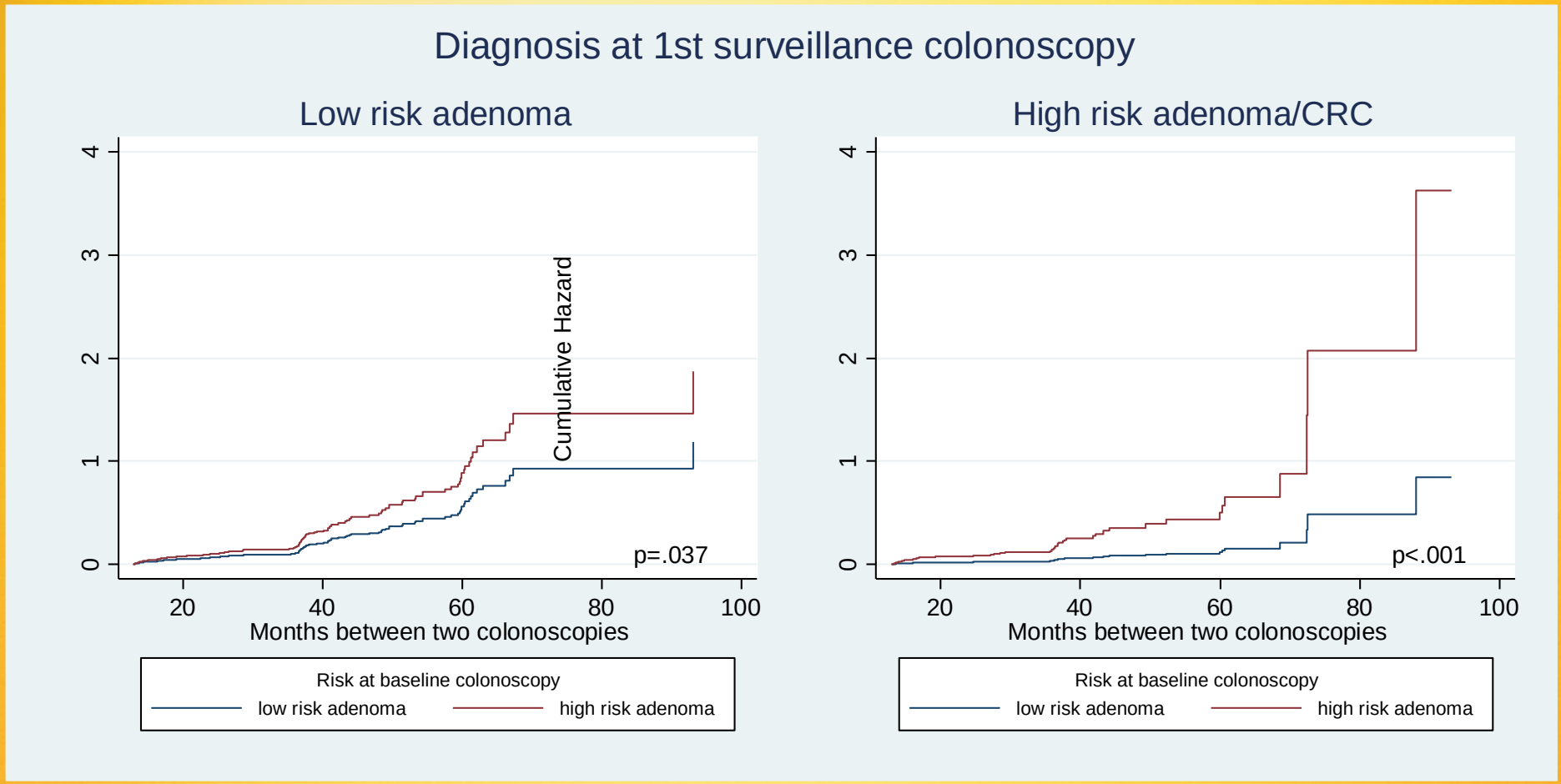


Results – model comparison

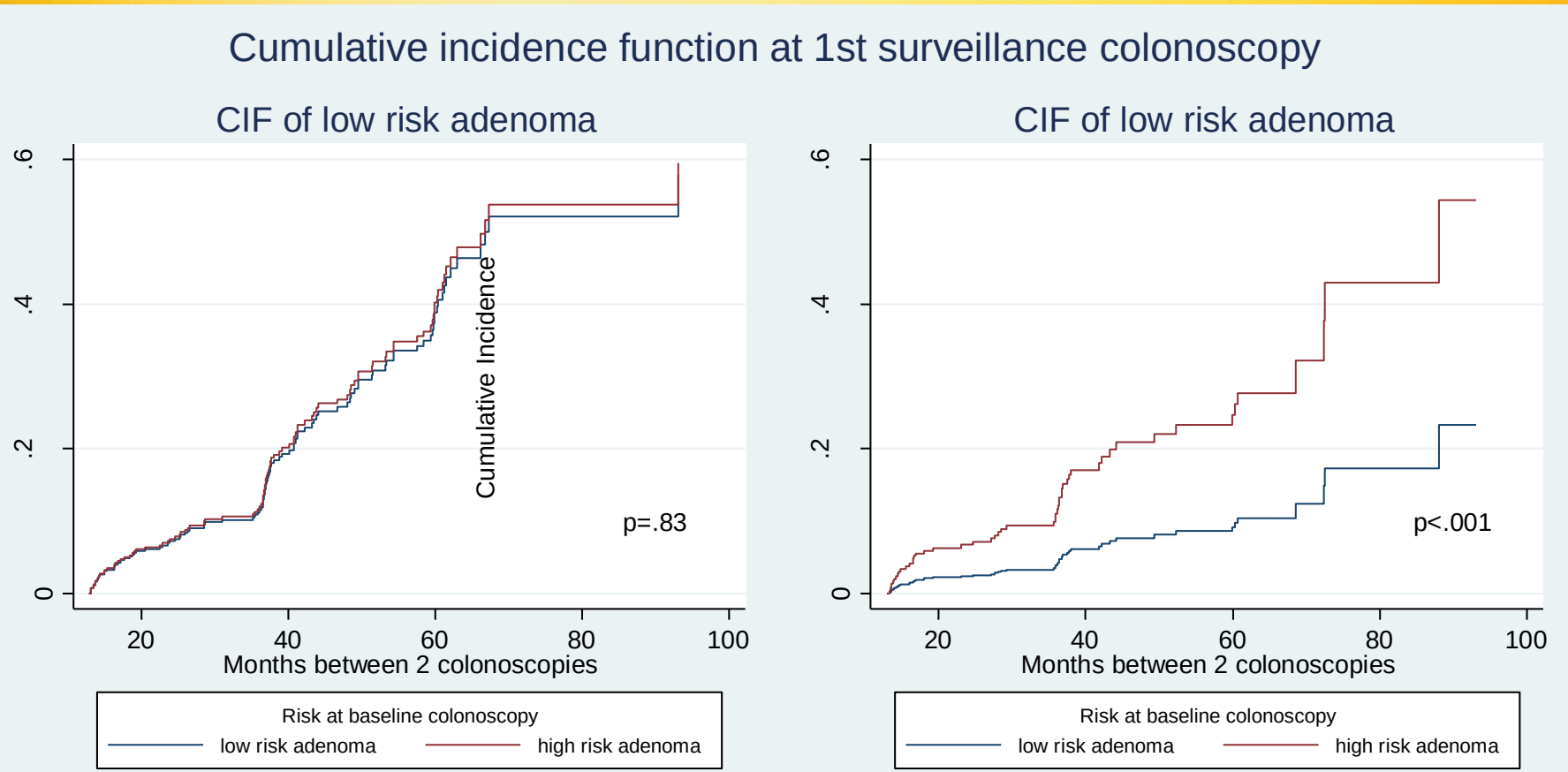
	Semi-parametric Cox model (stcox)	Parametric model (streg – Weibull)	Competing-risks survival model (stcrreg)	Stratified Cox model (stcox,…strata())	Multinominal logistic model mlogit
	HR [95% CI]	HR [95% CI]	SHR [95% CI]	HR [95% CI]	IRR [95% CI]
Low risk adenoma					
Risk category at index					
Low risk adenoma	1.00	1.00	1.00	**1.00**	1.00
High risk adenoma	**1.58 [1.03,2.42]***	**1.58 [1.04,2.42]***	**1.05 [0.69,1.58]**	**2.78 [1.98,3.89]*****	**0.49 [0.28,0.84]****
Gender					
Female	1.00	1.00	1.00	**1.00**	
Male	**2.14 [1.37,3.34]*****	**2.10 [1.35,3.28]*****	**2.19 [1.40,3.40]*****	**1.13 [0.81,1.58]**	**3.29 [1.93,5.59]*****
Age at index (years)	**1.02 [1.00,1.04]***	**1.02 [1.00,1.04]***	**1.01 [1.00,1.03]**	**1.01 [0.99,1.02]**	**1.03 [1.01,1.05]***
Reason for 1st surveillance					
Scheduled surveillance	1.00	1.00	1.00	**1.00**	
FOBT positive	**2.08 [1.23,3.52]****	**2.01 [1.20,3.37]****	**1.48 [0.84,2.62]**	**1.10 [0.77,1.57]**	1.33 [0.69,2.58]
Time between two colonoscopy	NA	NA	NA	NA	0.99 [0.98,1.01]
High risk adenoma/CRC					
Risk category at index					
Low risk adenoma	1.00	1.00	1.00		1.00
High risk adenoma	**4.31 [2.19,8.45]*****	**4.25 [2.26,7.98]*****	**2.95 [1.65,5.28]*****		**1.55 [0.76,3.15]**
Gender					
Female	1.00	1.00	1.00		
Male	0.91 [0.52,1.59]	0.95 [0.54,1.65]	0.81 [0.46,1.43]		1.31 [0.70,2.47]
Age at index (years)	**1.03 [1.00,1.06]***	**1.03 [1.00,1.06]***	**1.03 [1.00,1.06]***		**1.04 [1.01,1.07]****
Reason for 1st surveillance					
Scheduled surveillance	1.00	1.00	1.00		
FOBT positive	**2.95 [1.46,5.97]****	**3.58 [1.83,6.99]*****	**2.46 [1.25,4.83]****		2.06 [0.94,4.53]
Time between two colonoscopy	NA	NA	NA		1.00 [0.97,1.02]

* p<0.5; ** p<.01; *** p<.001

Results - adjusted cumulative hazard function of diagnoses at surveillance colonoscopy by index risk groups (*stcurve, cumhaz.....*after *stcox* model)



Results – CIF from competing risks model (*stcrreg*)



Summary of the differences between models

- Logistic regression vs. survival analysis

  High risk adenoma at index had reduced risk of advancing to low risk adenoma, and no difference in risk of advancing to high risk adenoma compared to low risk adenoma cohort.

  -Contradicted to Kaplan-Meier results (reason?)

- Stratified Cox model

  No estimates for the stratified variable – but the variable is our interest

- Cox model (semi-parametric) vs. parametric survival model

  Estimates are similar

- Cause-specific parametric survival model vs. competing risks model

  HR attenuated in competing risks model

Discussion

Why not nonparametric survival analysis?

Kaplan-Meier (*sts graph)*; with log-rank test (sts test)

Demerit:

- can not take into account of the effect of covariates.

Merits:

- good preliminary assessment for individual risk factors.
- Visualization for proportional hazard assumption.

Discussion cont.

Why not multinomial logistic regression (*mlogit*)?

Demerits:

- Cannot assess the relationship between predictors and survival time – time is a predictor in logistic regression.
- Cannot take into account of censoring
- Can misinterpret the effect of time – a bit complicated

Merits:

- Easy to perform the analysis
- Easy to interpret – although results could be misleading

Discussion cont.

Why not stratified Cox model (*stcox…, strata(type of events)*)?

Demerits & merits:

- Single estimate and easy to interpret – but only if we are not interested to know the difference between different type of events.

stset time, failure(event)
stcox i.index_risk i.sex age_index…, strata(surveillance)

Competing risks model example:
*primary interest - low risk adenoma
stset time, failure(surveillance==1)
stcrreg i.index_risk i.sex age_index, compete(surveillance==2)
*primary interest - high risk adenoma/CRC
stset time, failure(surveillance==2)
stcrreg i.index_risk i.sex age_index, compete(surveillance==1)

Discussion cont.

Why not parametric survival mode (*streg*)?

Demerits:

- Have the assumptions on the shape of hazard
- Whatever the hazard shape is, it is the same for everybody

Merits:

- When the assumption on shape of hazard for intervening is correct, parametric estimates are more efficient

Discussion cont.

Why not competing risks survival analysis (*stcrreg*)?

Merits:

- Incidence-rate curve represent the observed data in the presence of competing failure events – more close to real life scenario.
- Describe covariates effect is more straightforward.

Demerits:

- Competing events assumptions

  For this particular data, the events are not actually mutually excluded. Classification was based on the highest pathology rating.
- More difficult to interpret subdistribution hazard ratio (SHR).

Discussion cont.

Why Cox cause-specific proportional hazard model (*stcox*)?

- No assumption need to be made for the shape of the hazard over time – can be any shape
- Whatever the hazard shape is, it is the same for everybody
- Effect of covariates and HR are easy to interpret

Discussion cont.

So…

- Test the PH assumption and see if it is met.
- If PH assumption is met, then stick with survival analysis models, such as Cox, competing risks, stratification or multiple events analysis, depending on research questions and primary interest.
- Multinomial logistic is clearly inappropriate for such data.

More:

If have time varying predictor(s), try *“stpm2”* (flexible parametric survival model).

References

1. Saini SD, Kim HM, Schoenfeld P. Incidence of advanced adenomas at surveillance colonoscopy in patients with a personal history of colon adenomas: a meta-analysis and systematic review. Gastrointestinal endoscopy 2006;**64**(4):614-26.
2. 5. Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. Journal of the American Statistical Association 1999;**94**:496-509.
3. 6. Dignam JJ, Zhang Q, Kocherginsky M. The use and interpretation of competing risks regression models. Clinical cancer research : an official journal of the American Association for Cancer Research 2012;**18**(8):2301-8.
4. 7. Andersen PK, Geskus RB, de Witte T, et al. Competing risks in epidemiology: possibilities and pitfalls. Int J Epidemiol 2012;**41**(3):861-70.
5. 8. Cleves M, Gould W, Gutierrez RG, et al. Chapter 1. The problem of survival analysis. An introduction to survival analysis using Stata. Third ed. Texas: Stata Press, 2010.

Thank you!