

Method-specific reporting for reproducible competing-risk prediction in Stata

Our thoughts on what can be reported by authors publishing competing risks prognostic models, to ensure smooth external validation?

Aisha Vayani, MSc | Professor Mark Rutherford, PhD

Biostatistics Research Group, University of Leicester
Contact email address: av248@leicester.ac.uk

UK Stata Conference 2026 | London | 3–4 September



A simple Cox prediction

Under proportional hazards

Table 3: Variables Most Significantly Associated with Survival – Cox Proportional-hazards Model

95 % Confidence Interval	Odds Ratio	p-value	Variable
1.46–3.07	2.12	<0.001	Male group
1.32–8.47	3.35	0.008	Age*
1.52–4.09	2.49	<0.001	New-onset atrial fibrillation
1.10–5.42	2.45	0.041	Female group
0.18–1.09	0.45	0.083	Age*
			Diabetes mellitus
			Treatment with statins

* For each 10 years increment.

Miriam Shteinshnaider, Dorit Almozino-Sarafian, Sylvia Berman, Irma Tzur, Natan Cohen, Oleg Gorelik, Gender-related Differences in Clinical Variables Associated with Survival of Patients with Atrial Fibrillation, *European Cardiology* 2012;8(3):152–7
<https://doi.org/10.15420/ecr.2012.8.3.152>

Worked 5-year example

Covariate profile: Female, aged 70, diabetes mellitus = yes, treatment with statins = no.

$$\text{linear predictor} = (2.49 \times ?) + (2.45 \times ?) + (0.45 \times ?)$$

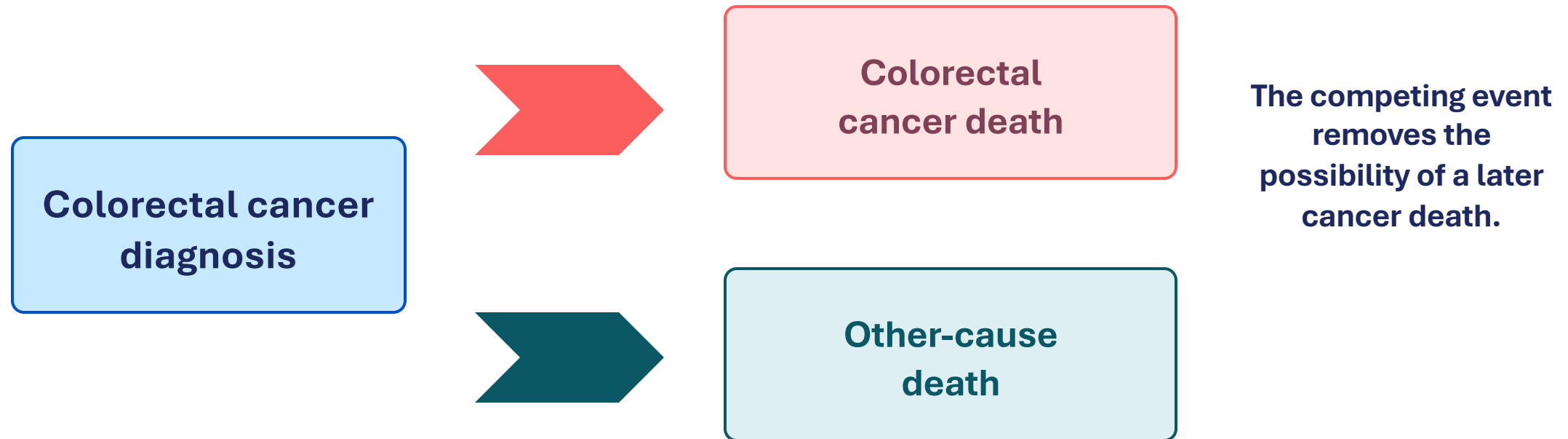
$$\text{Survival}(5 | x) = \text{Baseline}(5) \times \exp(\text{linear predictor})$$

$$\text{Baseline}(5) = ???,$$

$$\text{Absolute risk}(5 | x) = 1 - \text{Survival}(t|x) = ??$$

We need the baseline function, full set of coefficients and the covariate coded values.

Competing risks



Target: a cause-specific cumulative incidence function (CIF)

$$\text{CIF}_k(t|x) = P(T \leq t, J = k | x)$$

Probability of cause k by time t, knowing that another cause could occur first.

Reporting needs differ by competing-risks model

CIF based models

Fine-Gray models

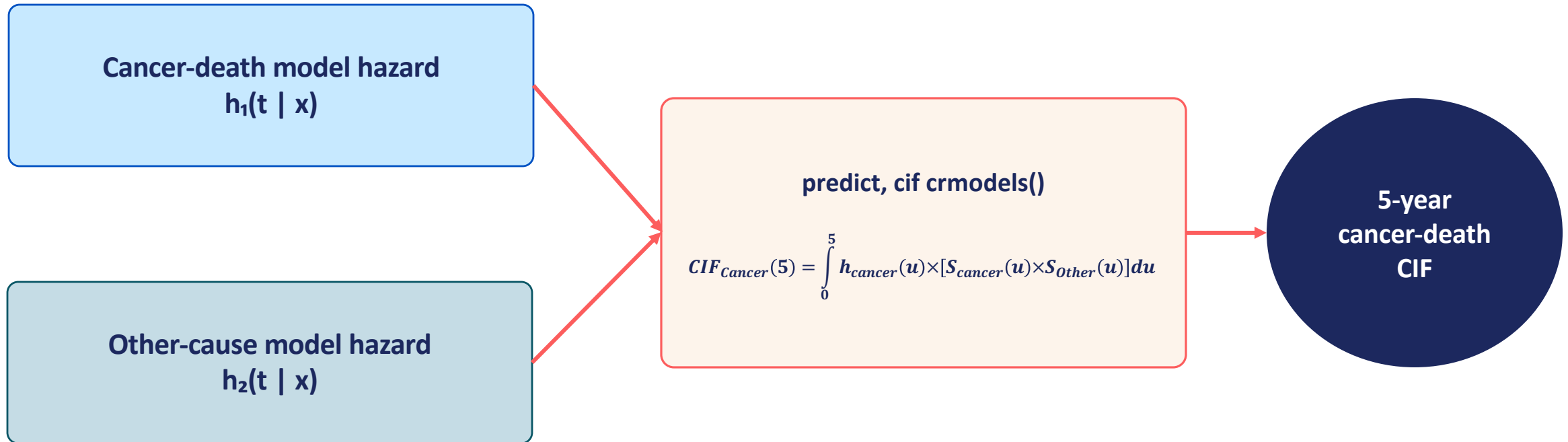
- Models the CIF directly
- Under proportional subhazards: coefficients + baseline CIF
- Separate cause models don't always form a coherent joint risk
- Time dependent effects makes the reporting more complex

Cause-specific hazard models

Flexible parametric models (FPMs)

- Fit one cause-specific hazard model for every cause
- Report the fitted components for every model
- State how the models are combined to obtain the CIF

Two cause-specific models produce one cancer-death CIF



Fit and save one stpm3 model for each cause

Cause 1: colorectal cancer death in 1975-84 dataset

```
use "internal.dta", clear

stset surv_yy, failure(status==1) id(id) exit(time 5)

stpm3 i.stage i.sex ///
@rcs(age, df(3) winsor(2.5 97.5)) ///
i.stage#i.sex, ///
df(3) ///
tvc(i.stage @rcs(age, df(3) winsor(2.5 97.5))) dftvc(2) ///
scale(lncumhazard) initvaluesloop

estimates store M_time_cancer
estimates save "M_time_cancer.ster", replace
```

Repeat for other-cause death

Use the same model form.

Change only:

failure(status==1)

↓

failure(status==2)

Then save M_time_other.ster

Same model structure for simplicity, with different failure definitions.

Restore both models and predict the 5-year CIF

Restore fitted models in 1985-94 colorectal dataset

```
use "external.dta", clear

stset surv_yy, failure(status==1) id(id) exit(time 5)

estimates use "M_time_other.ster"
estimates store M_time_other

estimates use "M_time_cancer.ster"
estimates store M_time_cancer
```

Predict the 5-year cancer-death CIF

```
gen prediction_time = 5

predict cif5, cif ///
crmodels(M_time_cancer M_time_other) ///
crnames(cancer other) ///
timevar(prediction_time) merge

rename cif5_cancer predicted_cif5
rename cif5_other predicted_other_cif5
```

Predict performs the integration. Matching variable names, covariate coding and compatible Stata versions are assumed.

Manual reconstruction requires both cause-specific models

Each fitted model

- Time scale, units and outcome definition
- Covariate definitions and coding
- Model scale and baseline spline basis and knots
- All coefficients, including intercept and baseline terms
- Non-linear and TDE predictor basis and knot locations

×2

Combining step

- Formulas to calculate the overall survival and hazards for each cause
- Integration formula and time grid used to calculate the CIF

Calibration with calplotcr

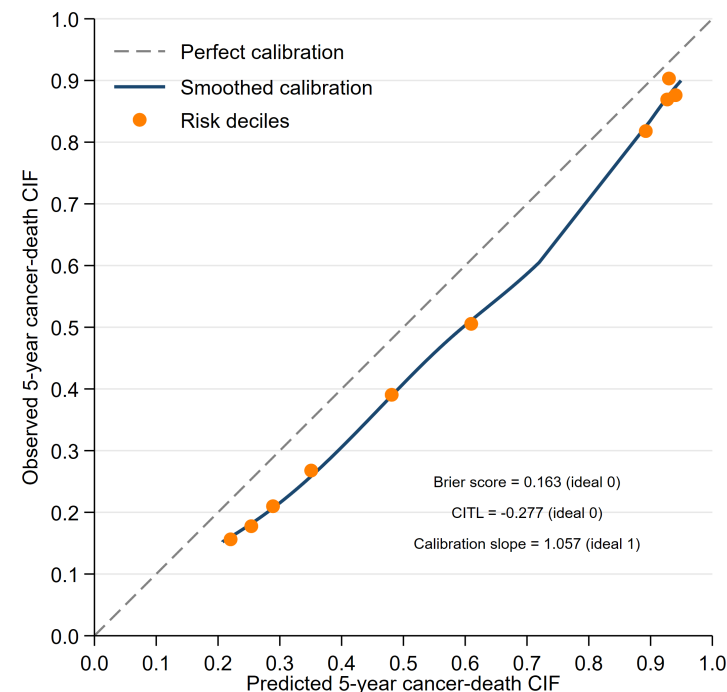
Assess calibration in 1985-94 colorectal dataset

```
* Predicted-risk groups
xtile rg_cancer = predicted_cif5, nq(10)
xtile rg_other = predicted_other_cif5, nq(10)
gen validation_data = 1

stset surv_yy, failure(status==1 2) id(id) exit(time 5)

calplotcr predicted_cif5 predicted_other_cif5, aj ///
  models(M_time_cancer M_time_other) ///
  eventvar(status) events(1 2) timepoint(5) ///
  riskgroups(rg_cancer rg_other) ///
  datavar(validation_data 1) bin(10) ///
  rcsrmsknots(5)
```

Predicted versus observed 5-year CIF



Ideal calibration follows the 45° line. Here the curve is mostly below it: the older model overpredicts risk in the later period.

A .ster file is one practical Stata-to-Stata route

.ster + prediction code

- Preserves fitted components
- Avoids manual transcription of fitted components
- Easier route for validation within Stata

Still report clearly

- Covariate coding
- Event definition, prediction time points and units
- Software and command versions

The object reduces transcription. It does not replace transparent reporting.

An Excel workbook alternative

	A	B	C	D	E	F	G	H
1	Fine-Gray cumulative incidence calculator							
2	Prediction tool for cancer death in the presence of other-cause death							
3								
4	Model specification and variable coding							
5	Prediction target	Cumulative incidence (absolute risk) of cancer death; other-cause death is treated as a competing event.						
6	Model command	stcrreg c.age10##i.stage i.sex##i.stage i.race, compete(cod == 2) [Stata 19.0]						
7	Model assumption	Proportional subdistribution hazards. This is not valid for a Fine-Gray model containing time-dependent effects.						
8	Prediction formula	CIF(t X) = 1 - [1 - F0(t)] ^ exp(X beta) ; F0(t) corresponds to the CIF(t X) when all covariates are set to 0						
9	Linear predictor	X beta is the sum of each beta coefficient multiplied by the corresponding coded patient value.						
10	Baseline profile	Age 60 years, Localised stage, Male and White. All model covariates equal zero, so CIF(t X) equals F0(t).						
11	Outcome coding	cod = 1 cancer death; cod = 2 other-cause death; cod = 0 censored.						
12	Age coding	age10 = (age - 60) / 10						
13	Stage coding	Localised is the reference; Regional and Distant are indicator variables.						
14	Sex and race coding	Male and White are the reference categories; Female and Black are indicator variables.						
15								
16	Coefficient matrix							
17	Parameter	Model component	Beta coefficient	Coded value	Beta coefficient * Coded value			
18	age10	Age centred at 60, per 10 years	0.287	0.00	0.000			
19	2.stage	Regional stage	1.259	0	0.000			
20	3.stage	Distant stage	2.906	1	2.906			
21	2.stage#c.age10	Regional stage x age10	-0.109	0.00	0.000			
22	3.stage#c.age10	Distant stage x age10	-0.185	0.00	0.000			
23	2.sex	Female	-0.267	1	-0.267			
24	2.sex#2.stage	Female x Regional stage	0.144	0	0.000			
25	2.sex#3.stage	Female x Distant stage	0.323	1	0.323			
26	2.race	Black race	0.042	1	0.042			
27	Linear predictor (X beta)				3.004			
28								

Baseline CIF values used by the calculator	
Time point	Baseline CIF F0(t)
1	0.029
3	0.056
5	0.073
10	0.100

Patient profile and predicted risk	
Edit the beige cells below	
Age (years)	60
Stage	Distant ▼
Sex	Female ▼
Race	Black ▼
Prediction time (years)	10 ▼
Linear predictor (X beta)	3.004
exp(X beta)	20.171
Baseline CIF, F0(t)	0.100
Predicted CIF	0.881
Predicted risk	88.07%

Visible coefficients, covariate coding, baseline CIFs and the prediction formula

This is what have we learned so far

1

Competing risks reporting are method-specific

Cause-specific models with splines and time-dependent effects increase the reporting requirement.

2

Model objects are a practical reporting aid

Useful for validating models in Stata. Doesn't replace transparent reporting. Limited to users who know and have a Stata licence.

3

Software-neutral tools can be an alternative

Possible fallback for situations such as changes to the format of an .ster file.

This is work in progress. We welcome any feedback

References

Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. *JASA* 1999;94:496–509.

Andersen PK, Klein JP, Rosthøj S. Generalised linear models for correlated pseudo-observations. *Biometrika* 2003;90:15–27.

Royston P, Parmar MKB. Flexible parametric proportional-hazards and proportional-odds models. *Stat Med* 2002;21:2175–2197.

Syriopoulou E et al. Regression standardisation with competing events using *standsurv*. *BMC Med Res Methodol* 2022;22:226.

van Geloven N et al. Validation of prediction models in the presence of competing risks. *BMJ* 2022;377:e069249.

Collins GS et al. TRIPOD+AI statement. *BMJ* 2024;385:e078378.

Royston P, Altman DG. External validation of a Cox prognostic model. *BMC Med Res Methodol* 2013;13:33.

Booth S et al. Temporal recalibration in competing-risk prediction models. *Stat Med* 2023;42:5007–5024.

Mozumder SI et al. Calibration of cause-specific absolute risk for external validation. *Diagn Progn Res* 2025;9:23.

StataCorp. Stata 19 documentation: estimates *save/use*, *stpm3* and *postestimation*.

Lambert PC. *stpm3* resources and Stata conference materials. pclambert.net/software/stpm3/.

Thank you

Questions and discussion

Aisha Vayani

Biostatistics Research Group, University of Leicester
av248@leicester.ac.uk

Connect on LinkedIn

