Multi-state models: Rates, Risks, and Pseudo-Values

Overview of course

I Introduction to multi-state models

II Non-parametric estimation and regression models for intensities (Cox)

III Estimation of marginal parameters using plug-in

IV Direct regression models for marginal parameters (Cox, Fine-Gray, Ghosh-Lin)

V Pseudo-values

IV: Direct regression models for marginal parameters

- Plug-in vs. direct models
- Two-state model
  - $Q_0(t \mid Z)$
  - $\varepsilon_0(t \mid Z)$
- Competing risks: the Fine-Gray model
- Recurrent events: the Ghosh-Lin (and LWYY) model
- The g-formula

Plug-in vs. direct models

Plug-in estimation of marginal parameters:

- Builds on 'correctness' of intensity models and on *likelihood inference*
- Does not provide parameters that directly describe the association between the marginal parameter and covariates

An alternative strategy uses *direct models for the marginal parameter*:

- This will typically entail fewer modeling assumptions (we need not model *all* intensities)
- This directly targets a marginal parameter of interest (the 'estimand')
- This requires setting up relevant *estimating equations* as we can no longer make likelihood inference
- This can also be obtained using *pseudo-values* – more later!

The two-state model, $Q_0(t \mid Z)$

Because of the one-to-one correspondence between the hazard function $\alpha_{01}(t)$ and the survival function $S(t) = Q_0(t)$, a Cox model for the intensity immediately provides a regression model for the survival function (at *all* time points).

Parameters (hazard ratios) are then to be interpreted on the scale of the *link function*, $g(\cdot) = \text{cloglog}(\cdot)$

$$\text{cloglog}(Q_0(t \mid Z)) = \log(-\log(1 - Q_0(t \mid Z)))$$

$$= \log(A_0(t)) + \beta^{\mathsf{T}} Z.$$

The two-state model, $\varepsilon_0(t \mid Z)$

Tian et al. (2014, *Biostatistics*) studied estimating equations for the RMST $\varepsilon(t_0 \mid Z) = E(\min(T, t_0) \mid Z)$. This uses the idea of *inverse probability of censoring weighting*, IPCW, and the *generalized estimating equations*, GEE, are:

$$\sum_i \frac{I(\min(T_i, t_0) < C_i)}{\widehat{G}(\min(T_i, t_0))} Z_i(\min(T_i, t_0) - g^{-1}(\beta^{\mathsf{T}} Z_i)) = 0,$$

(g is the link function, often taken to be the identity or $\log$).

This uses only the completely observed restricted survival times, $\min(T_i, t_0)$, whose contributions are *up-weighted* with the inverse probability of being uncensored. If censoring depends on covariates then a regression model for the censoring distribution $G(t \mid Z)$ is required. Variance of $\widehat{\beta}$ is estimated using the GEE-sandwich estimator (with an extra term arising from the need to estimate G).

Competing risks, the Fine-Gray model

The *Fine-Gray* model is a direct model for the cumulative incidence:

$$Q_h(t \mid Z) = F_h(t \mid Z) = P(T \leq t, D = h \mid Z), h = 1, \ldots, k$$

(at all time points). Recall from a Cox model for all-cause mortality that:

$$\log(-\log(1 - Q_0(t \mid Z))) = \log(A_0(t)) + \beta^{\mathsf{T}} Z.$$

Fine & Gray (1999, *JASA*) studied the similar model for a cumulative incidence:

$$\log(-\log(1 - Q_h(t \mid Z)) = \log(\widetilde{A}_{0h}(t)) + \widetilde{\beta}_j^{\mathsf{T}} Z.$$

The Fine-Gray model

This is a model for the hazard for the improper random variable

$$T_h^* = T \cdot I(D = h) + \infty \cdot I(D \neq h) = \inf_t \{V(t) = h\},$$

i.e., for

$$\widetilde{\alpha}_h(t) = -\frac{\mathrm{d}}{\mathrm{d}t} \log(1 - Q_h(t)).$$

That is, the transformation which for all-cause mortality takes us from cumulative risk to hazard is used for a cumulative incidence in a competing risks model.

The resulting $\widetilde{\alpha}_h(t)$ is denoted the *sub-distribution hazard* and the Fine-Gray model is thus a proportional sub-distribution hazards model.

The Fine-Gray model

A problem is that, while the hazard function $\alpha(t) = \alpha_{01}(t)$ has the useful 'rate' interpretation:

$$\alpha(t) \approx P(\text{die before } t + dt \mid \text{ alive } t)/dt, \quad dt > 0 \text{ small},$$

and so has the cause-specific hazard $\alpha_h(t) = \alpha_{0h}(t)$:

$$\alpha_h(t) \approx P(\text{die from cause } h \text{ before } t+dt \mid \text{ alive } t)/dt, \quad dt > 0 \text{ small},$$

the sub-distribution hazard has *not*. Thus, for $dt > 0$ (small),

$$\widetilde{\alpha}_h(t) \approx P(\text{die from cause } h \text{ before } t + dt \mid$$
$$either \text{ alive at } t \ \ or \text{ dead from a competing cause by } t)/dt.$$

The Fine-Gray model

The model for the sub-distribution hazard is:

$$\widetilde{\alpha}_h(t \mid Z) = \widetilde{\alpha}_{0h}(t) \exp(\widetilde{\beta}_1 Z_1 + ... + \widetilde{\beta}_p Z_p),$$

but, while a 'sub-distribution hazard' sounds like a hazard, it is not! Therefore, the resulting parameters $\exp(\widetilde{\beta})$ in the Fine-Gray model have a rather indirect interpretation as 'sub-distribution hazard ratios'.

Anyway, the model is being used quite a bit and it is, indeed, useful by giving parameters that directly link the cumulative incidence to covariates (on the cloglog cumulative incidence scale).

The model also provides (covariate adjusted) *significance tests* for cumulative incidences.

Estimation

With no censoring, Fine and Gray defined the cause h 'risk set'

$$\widetilde{R}_h(t) = \{i : (T_i \geq t) \text{ or } (T_i \leq t, D_i \neq h)\}$$

and $\widetilde{\beta}_h$ is estimated by the partial likelihood score equation

$$U_h(\widetilde{\beta}_h) = \sum_i \int_0^\infty \Big(Z_i - \frac{\sum_{j \in \widetilde{R}_h(t)} Z_j \exp(\widetilde{\beta}_h^\mathsf{T} Z_j)}{\sum_{j \in \widetilde{R}_h(t)} \exp(\widetilde{\beta}_h^\mathsf{T} Z_j)}\Big) dN_{hi}(t) = 0$$

corresponding to replacing times of failures from causes other than h by $+\infty$.

With known (e.g., 'administrative') censoring (at C_i), the cause h risk set is replaced by

$$\widetilde{R}_h(t) = \{i : (\min(T_i, C_i) \geq t) \text{ or } (T_i \leq t, D_i \neq h, C_i \geq t)\}.$$

Estimation (ctd.)

- To identify this 'risk set', we need to know the times C of censoring for subjects who failed
- With general censoring, an Inverse Probability of Censoring Weighted (IPCW) score equation is used and to use this, a model for censoring is needed
- In the simplest case, one uses the 'Kaplan-Meier for censoring', that is, estimating $P(C > t)$. (In this analysis 'failures are censorings')
- If censoring depends on covariates then a model for $P(C > t \mid Z)$ is needed for the weights, e.g., a Cox model
- A Breslow-type estimator for the cumulative baseline sub-distribution hazard and asymptotic theory were also discussed by Fine and Gray (1999)

The Fine-Gray model

The Fine-Gray model provides parameters describing the relationship between the covariates and the cause h risk. For example, for a binary covariate Z_1 with an estimated regression coefficient $\widetilde{\beta}_1 > 0$ it follows that for all values, Z_2^0, for the other covariates in the model we have that

$$\widehat{Q}_h(t \mid Z_1 = 1, Z_2^0) > \widehat{Q}_h(t \mid Z_1 = 0, Z_2^0).$$

The positive regression coefficient has the *qualitative* meaning that individuals with $Z_1 = 1$ have a uniformly increased cause h cumulative incidence compared to those with $Z_1 = 0$.

However, the resulting estimates $\exp(\widetilde{\beta}_h)$ are sub-distribution hazard ratios, so the *quantitative* meaning of the regression coefficient is not simple.

Example: the PBC-3 trial

PBC-3 trial in liver cirrhosis: Estimated coefficients (and SD) from Fine-Gray models for death without transplantation and transplantation.

		Death without transplantation		Transplantation	
Covariate		$\widehat{\beta}$	SD	$\widehat{\beta}$	SD
Treatment	CyA vs. placebo	-0.353	0.260	-0.409	0.368
Albumin	per 1 g/L	-0.061	0.031	-0.070	0.033
$\log_2$(Bilirubin)	per doubling	0.616	0.089	0.619	0.101
Sex	male vs. female	-0.415	0.317	-0.092	0.580
Age	per year	0.087	0.016	-0.075	0.017

R code: Fine-Gray model, survival package

```
## The finegray function creates a new data set for a
## cause and fitting a certain Cox model to it
## gives the Fine-Gray model.
library(survival)
dat1<-finegray(Surv(years,factor(status))~ .,
data=pbc3, etype=1)
cox1<-coxph(Surv(fgstart,fgstop,fgstatus) ~
tment + alb + log2(bili) + sex + age, data=dat1, weight=fgwt)
summary(cox1)

dat2<-finegray(Surv(years,factor(status))~ .,
data=pbc3, etype=2)
cox2<-coxph(Surv(fgstart,fgstop,fgstatus)~
tment + alb + log2(bili) + sex + age, data=dat2, weight=fgwt)
summary(cox2)
```

R code: Fine-Gray (and Aalen-Johansen), mets package

```
library(mets)
cif1 <-
cif(Event(years,status)~strata(tment), data = pbc3, cause=1)
plot(cif1)
cif2 <-
cif(Event(years,status)~strata(tment), data = pbc3, cause=2)
lines(cif2)

pbc3small<-subset(pbc3,!is.na(alb)) # Missing values!
fg1 <- cifreg(Event(years,status) ~ tment + alb + log2(bili)
+ sex + age, data=pbc3small, cause=1, propodds=NULL)
summary(fg1)
fg2 <- cifreg(Event(years,status) ~ tment + alb + log2(bili)
+ sex + age, data=pbc3small, cause=2, propodds=NULL)
summary(fg2)
```

Recurrent events, LWYY model

The main marginal parameter is $\mu(t) = E(N(t))$, and a multiplicative regression model for this is (Eq. (4.12)):

$$\mu(t \mid Z) = \mu_0(t) \exp(\beta^{\mathsf{T}} Z).$$

Without competing risks, this is the Lin-Wei-Yang-Ying (2000) model ('LWYY'), and β-parameters may be estimated by solving exactly the Cox partial likelihood 'score' equations, which are still unbiased, though no longer likelihood-based.

For this reason, *robust* ('sandwich') variance estimators must be used, rather than those based on the second order derivatives of a log likelihood.

However, situations without competing risks are rare (if an event recurs, it is not all cause mortality and, hence, mortality will 'always' be a competing risk).

Recurrent events, the Ghosh-Lin model

In this situation, the model $\mu(t \mid Z) = \mu_0(t)\exp(\beta^{\mathsf{T}} Z)$ is the Ghosh-Lin (2002) regression model.

Estimation follows to a large extent what we saw for the Fine-Gray model in the sense that Ghosh and Lin first studied the case where censoring times C_i are known for everybody - even for those who died and, next, an IPCW method was derived. For the first case the estimating equation is:

$$U(\beta) = \sum_i \int_0^\infty \big(Z_i - \bar{Z}^U(t)\big) I(C_i \geq t) dN_i(t) = 0$$

where $\bar{Z}^U(t)$ is the average

$$\bar{Z}^U(t) = \frac{\sum_j I(C_j \geq t) Z_j \exp(\beta^{\mathsf{T}} Z_j)}{\sum_j I(C_j \geq t) \exp(\beta^{\mathsf{T}} Z_j)}.$$

The Ghosh-Lin model

In the more common case where censoring times are not known for all, a model for $P(C > t \mid Z)$ (possibly without Z) is needed. Let $\widehat{G}(t)$ be the estimator for the censoring distribution, e.g., Kaplan-Meier if C is independent of Z or a Cox-model-based estimator otherwise.

Ghosh and Lin introduced the same weights

$$w_i(t) = I(C_i \geq \min(T_i, t))\widehat{G}(t)/\widehat{G}(\min(\widetilde{T}_i, t))$$

as those used by Fine and Gray and showed that:

$$E(w_i(t)) \approx G(t)$$

which is the expectation of $I(C_i \geq t)$ and, therefore, this indicator in the previous estimating equation can be replaced by $w_i(t)$.

The Ghosh-Lin model

The resulting estimating equation, Eq. (5.30), is

$$U(\beta) = \sum_i \int_0^\infty \big(Z_i - \bar{Z}^G(t)\big) w_i(t) dN_i(t) = 0$$

with

$$\bar{Z}^G(t) = \frac{\sum_j w_j(t) Z_j \exp(\beta^\mathsf{T} Z_j)}{\sum_j w_j(t) \exp(\beta^\mathsf{T} Z_j)}.$$

Finally, (sandwich-type) variances and the baseline mean function can be estimated, the latter by a Breslow-type estimator:

$$\widehat{\mu}_0(t) = \sum_i \int_0^t \frac{w_i(u) dN_i(u)}{\sum_j w_j(u) \exp(\widehat{\beta}^\mathsf{T} Z_j)}.$$

Example: Repeated episodes in affective disorder

Recurrent episodes in affective disorders: Estimated ratios between mean numbers of psychiatric episodes between patients with bipolar vs. unipolar diagnosis (c.i.: confidence interval).

Model	Mortality treated as	$\exp(\widehat{\beta})$	95% c.i.
LWYY model	Censoring	1.52	(1.07, 2.17)
Ghosh-Lin model	Competing risk	1.96	(1.34, 2.87)

Estimated hazard ratio between bipolar and unipolar patients in a Cox model for the marginal mortality rate is 0.410 with 95% confidence limits from 0.204 to 0.825.

R code: mets package

```
library(mets)
# Ghosh-Lin model
fit2 <- recreg(Event(prev, stop, status) ~ bip + cluster(id),
data = angst, cause = 1, cens.code = 0, death.code = 2)
summary(fit2)

# LWYY model (censor for death)
library(survival)
fit1<-coxph(Surv(prev,stop,status==1) ~ bip + cluster(id),
data = angst)
summary(fit1)
# Cox model for death
fit3<-coxph(Surv(prev,stop,status==2) ~ bip, data = angst)
summary(fit3)
```

The g-formula

Using either plug-in or direct modeling, it is possible to estimate a marginal parameter *for given covariates*, Z.

To obtain an estimate for the population, the 'g-formula' may be used.

As an example, an average survival curve for each treatment group ($z_0 \in \{0, 1\}$) in the PBC-3 study can be estimated by:

$$\widehat{S}(t \mid Z_0 = z_0) = \frac{1}{n} \sum_i \widehat{S}(t \mid Z_0 = z_0, Z_i),$$

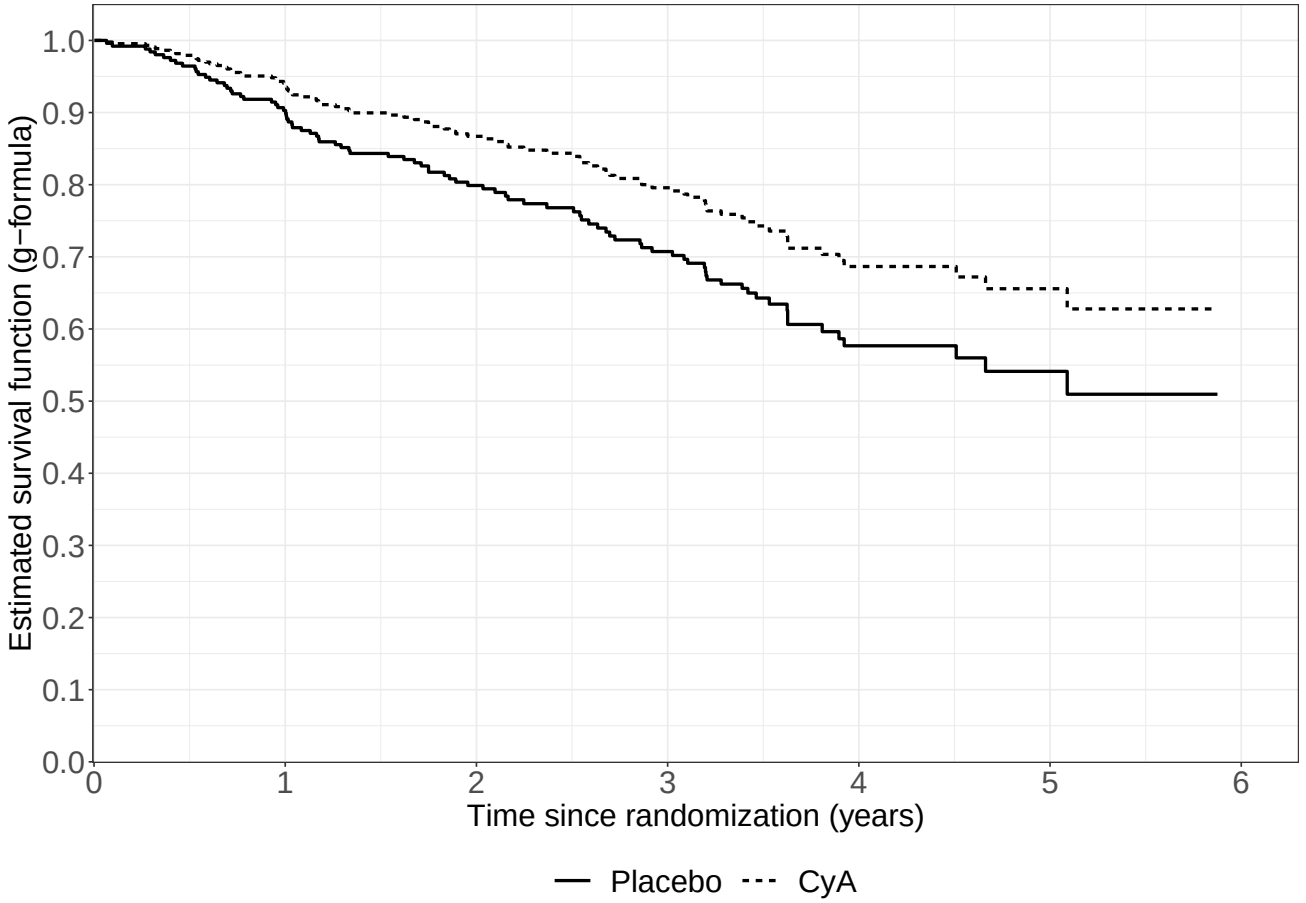
i.e., treatment is set to z_0 for every one but other individual covariate values, Z_i are kept as observed and interpretation of individual covariate effects is less important (Fine-Gray model!).

This is a way of illustrating the adjusted treatment effect on the survival probability scale.

The SD of the estimator may be obtained using bootstrap.

Average survival curves, PBC-3, g-formula

Estimates are based on plugging-in Cox models including treatment, albumin, and $\log_2$(bilirubin).



R code: g-formula, mets package

```
library(mets)
pbc3small<-subset(pbc3,!is.na(alb)) # missing values!

# G-formula based on Cox model

surv<-phreg(Surv(years,status>0) ~ tment + alb + log2(bili),
data = pbc3small)
summary(surv)

survivalG(surv, data=pbc3small, time=1)
plot(survivalGtime(surv,data=pbc3small))
```

R code: g-formula, mets package

```
# G-formula based on Fine-Gray models

death<-cifreg(Event(years,status)~tment+alb+log2(bili)+sex+age,
data = pbc3small, cause=2, cox.prep=TRUE, propodds=NULL)
summary(death)
survivalG(death, data = pbc3small, time = 1)
plot(survivalGtime(death,  data = pbc3small))

trans<-cifreg(Event(years,status)~tment+alb+log2(bili)+sex+age,
data = pbc3small, cause=1, cox.prep=TRUE, propodds=NULL)
summary(trans)
survivalG(trans, data = pbc3small, time = 1)
plot(survivalGtime(trans, data = pbc3small))
```

R exercises - Direct marginal

Use the PBC3 data and `mets` package. NB: Use data set `pbc3small` as in lecture.

Two-state model

1. Fit a multiple Cox model with `tment + alb + log2(bili)` using `phreg`.
2. Estimate adjusted risk difference using g-formula at year 3 based on above Cox model – compare to marginal risk difference estimate from plug-in exercises.
3. Plot g-formula estimates.

Competing risks

4. Fit a multiple Fine-Gray model for each transition with `tment + alb + log2(bili)` using `mets` function `cifreg`.
5. Estimate adjusted risk difference for each cause using g-formula at year 3 based on above Fine-Gray models – compare to marginal risk difference (cumulative incidence difference) estimates from plug-in exercises.
6. Plot corresponding g-formula estimates for each cause.

Recurrent events

Use data set `rr.csv`.

7. Run a Ghosh-Lin model with treatment as covariate.