The competing risk approach adapted to left truncation allows reliable estimation of abortion risk from observational studies

Reinhard Meister

TFH Berlin

Observational Studies in Embryotoxicology

TIS Teratology Information Services

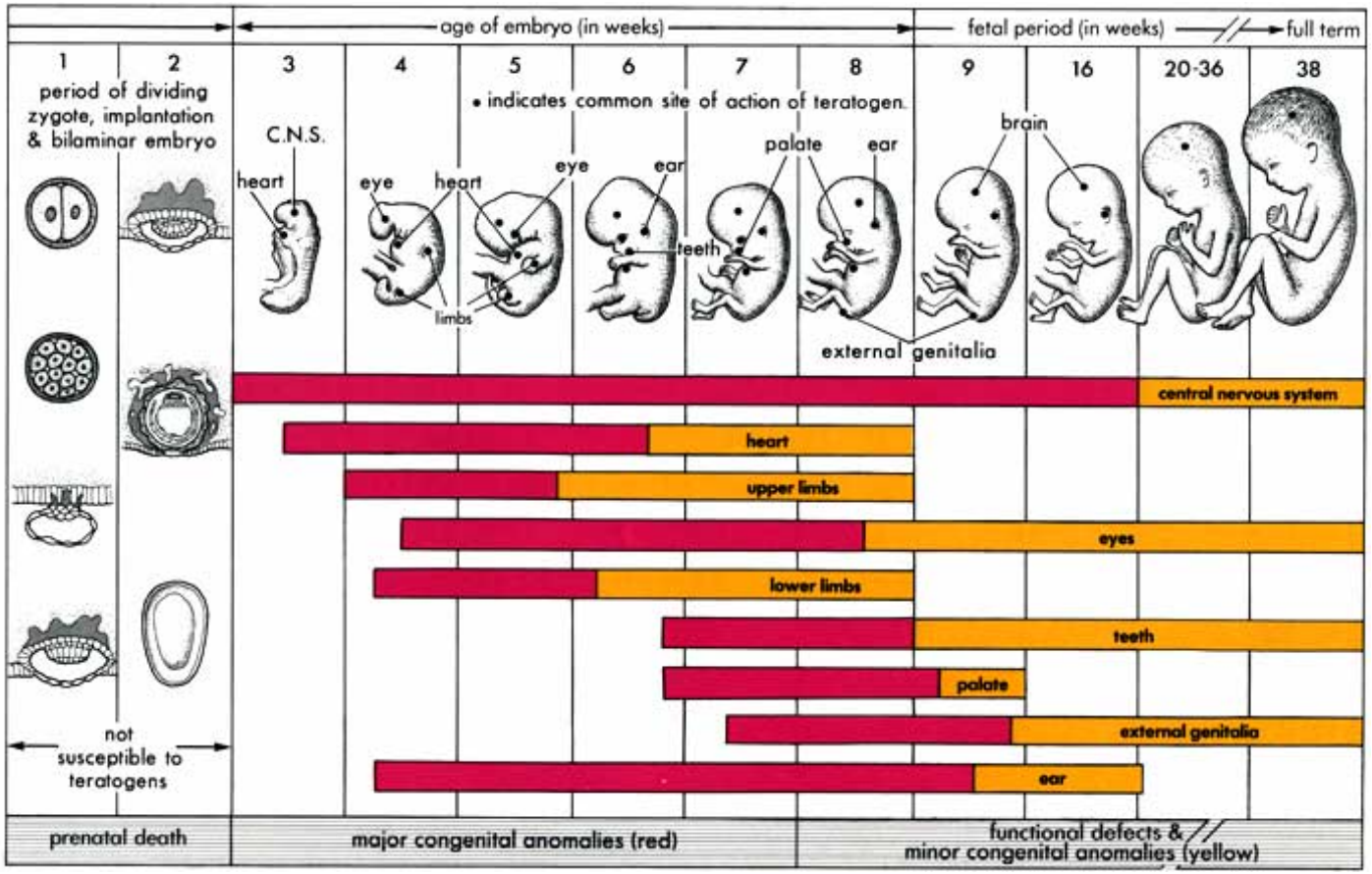
provide risk assessment and treatment recommendations to health care providers and pregnant women

Data collection as byproduct

- ascertain patient data prospectively
- data: diseases and their treatment, other exposures, vitamins, assisted conception, contraceptives during pregnancy, social drugs
- investigate the outcome of drug exposed pregnancies
- endpoints: congenital malformations if live birth; abortions

Schaefer C, Hannemann D, Meister R. Post-marketing surveillance system for drugs in pregnancy - 15 years experience of ENTIS. Reprod Toxicol 2005;20:331–43.

Developmental Stages during Pregnancy



Coumarin Derivatives During Pregnancy

Coumarin Derivatives

- Vitamin K antagonists and oral anticoagulants
- prevention of clottings in case of hereditary coagulation disorders, prosthetic heart valves, lung embolism etc

Association with pregnancy outcome: Berlin Study

- cohort of 1186 pregnant women
- 173 of them were exposed therapeutically to a coumarin anticoagulant (phenprocoumon)
- 1013 not exposed to potential teratogens served as controls
- risk of spontaneous abortion possibly increased

Schaefer C, Hannemann D, Meister R, et al. Vitamin K antagonists and pregnancy outcome. A multi-centre prospective study. Thromb Haemost 2006;95:949–57.

Spontaneous Abortions

Basics

- spontaneous abortions are frequent,
  rates between 11 and 16% of (self) diagnosed pregnancies $>= 5$ weeks post LMP
  (Goldhaber 2000, Saraiya 1999, Modvig 1990)
- possibly severe ultimative response to a an embryotoxic exposure
- therefore, important for risk assessment of drugs

Spontaneous Abortions

Crude rate estimation

- Follow up of n pregnancies
- finding a spontaneous abortions
- crude rate estimate: $\widehat{p} = \frac{a}{n}$

Causes for end of pregnancy by exposure, crude rates.
(Coumarin study TIS Berlin, $n_{exposed} = 173$, $n_{control} = 1013$)

	exposed	control
live birth	0.53	0.91
induced abortion	0.22	0.02
spontaneous abortion	0.25	0.07

Spontaneous Abortions

Crude rate estimation

- Follow up of n pregnancies
- finding a spontaneous abortions
- crude rate estimate: $\widehat{p} = \frac{a}{n}$

Causes for end of pregnancy by exposure, crude rates. (Coumarin study TIS Berlin, $n_{exposed} = 173, \quad n_{control} = 1013$)

	exposed	control
live birth	0.53	0.91
induced abortion	0.22	0.02
spontaneous abortion	0.25	0.07

Crude Rate Estimation

Results are biased and unreliable

Reasons

- delayed entry of cases
  wrong denominator – number at risk is time-dependent
- competing risks e.g. spontaneous, induced abortion, live birth
  different causes cannot be assumed to be independent

How to achieve reliable rate estimates?

Modeling observation times

- data given in the form $T_{start} < T_{stop}$
- time-dependent incidence rates $r(t) := P(T_{stop} = t)$
- split into hazard $\times$ survival

$$\begin{aligned} r(t) &= P(T_{stop} = t | T_{stop} \geq t) \times P(T_{stop} \geq t) \\ &= \lambda(t) \times S(t-) \end{aligned}$$

- cumulative incidence $F(t) = \sum_{t' \leq t} r(t')$

Adaptation to Delayed Entry

Estimating hazard and survival

- determine number at risk $n_i = \#\{T_{start} < t_i \leq T_{stop}\}$
- determine number of events $d_i = \#\{t_i = T_{stop}\}$
- hazard estimate $\widehat{\lambda}(t_i) = d_i/n_i$
- survival estimate $\widehat{S}(t_i-) = \prod_{t<t_i}\{1 - d_t/n_t\}$
  nonparametric ML estimate (Tsai et al)

Critical assumption

left truncation independent of events

Tsai WY, Jewell NP, Wang MC. A note on the product-limit estimator under right censoring and left truncation. Biometrika 1987;74:883–6.

Cumulative Incidence Functions for Competing Risks

Notation

$$\text{event time } T, \text{cause} = k, \quad k = 1, \ldots, K$$

Cumulative incidences

$$F_k(t) := P(T \leq t, \text{cause} = k), \quad P(T \leq t) = \sum_{k=1}^{K} F_k(t)$$

Quantification of cause specific rates

$$1 = \sum_{k=1}^{K} \pi_k, \quad \text{where} \quad \pi_k := \lim_{t \to \infty} F_k(t), \quad k = 1, \ldots, K$$

Kalbfleisch JD, Prentice RL. The Statistical analysis of failure time data. New York: John Wiley and Sons 1980.

Estimating Cumulative Incidences

Putting things together

- calculate hazard estimate
  $\widehat{\lambda}_k(t_i) = d_{ki}/n_i$ cause specific $k = 1, \ldots, K$
- calculate survival estimate
  $\widehat{S}(t_i-) = \prod_{t<t_i}\{1 - d_t/n_t\}$ taking all causes as events
- cumulative incidence function given by
  $\widehat{F_k}(t) := \sum_{t_i \le t} \widehat{\lambda}_k(t_i) \times \widehat{S}(t_i-)$

Variance estimation

Marubini E, Valsecchi MG. Analysing survival data from clinical trials and observational studies, Wiley and Sons 1995.

$$\widehat{\mathrm{Var}}\left[\widehat{F_k}(t)\right] = \sum_{t_i \le t}\left(\left[\widehat{F_k}(t) - \widehat{F_k}(t_i)\right]^2 \frac{d_i}{n_i(n_i - d_i)} + \left[\widehat{S}(t_i-)\right]^2 \frac{(n_i - d_{ki})}{n_i}\frac{d_{ki}}{n_i^2} - 2\left[\widehat{F_k}(t) - \widehat{F_k}(t_i)\right]\widehat{S}(t_i-)\frac{d_{ki}}{n_i^2}\right)$$

Cause Specific Hazard and Left Truncation

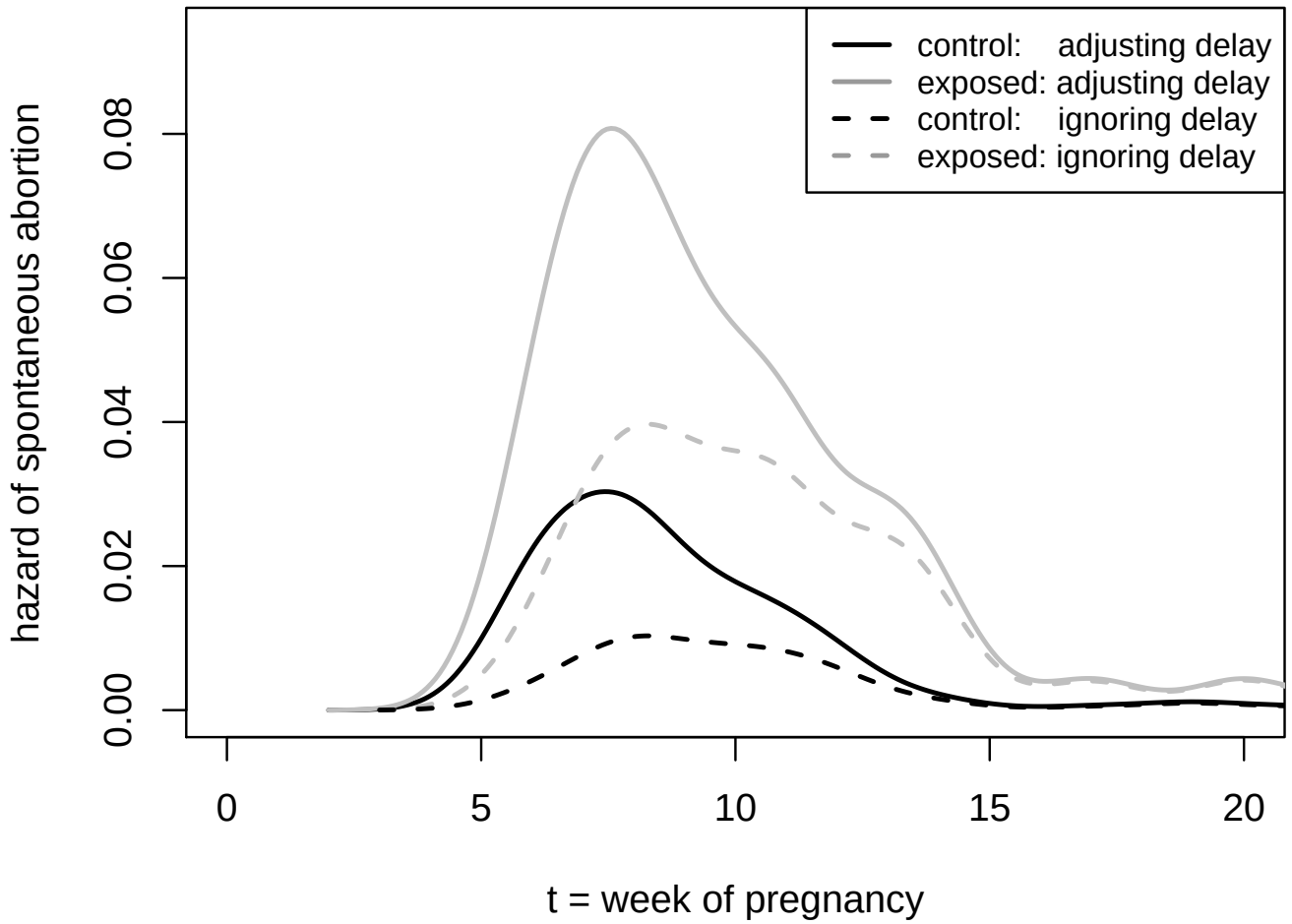
Comparing crude and adapted estimates

- estimate corrected for left truncation
  $\widehat{\lambda}_k(t_i) = d_{ki}/n_i$
- estimate pretending $T_{start} = 0$
  $\widetilde{\lambda}_k(t_i) = d_{ki}/\widetilde{n}_i$ where $\widetilde{n}_i$ is taken as $n - \sum_{k,j<i} d_{kj}$
- Kernel smoothed estimate

$$\widehat{\lambda}_k^S(t) = \frac{1}{b}\sum_i K\left(\frac{t - t_i}{b}\right)\frac{d_{ki}}{n_i}, \quad \text{with Kernel } K \text{ and bandwidth } b$$

Ramlau-Hansen H. Smoothing counting process intensities by means of kernel functions. Ann Statist 1983;11:453–66

Coumarin Study Hazard Spontaneous Abortion



Computations Using **R**

Hazard function and Kernel smoother

```
s.con <- summary(survfit(Surv(start,stop,
                 cause=="spontaneous abortion"),data=control))
hazard <- function(fit)
   {
    time<-fit$time
    h   <-fit$n.event/fit$n.risk
    return(data.frame(time=time,h=h))
    }
h.con <- hazard(s.con)
...
lines(density(h.con$time,weights=h.con$h,bw=1))
```

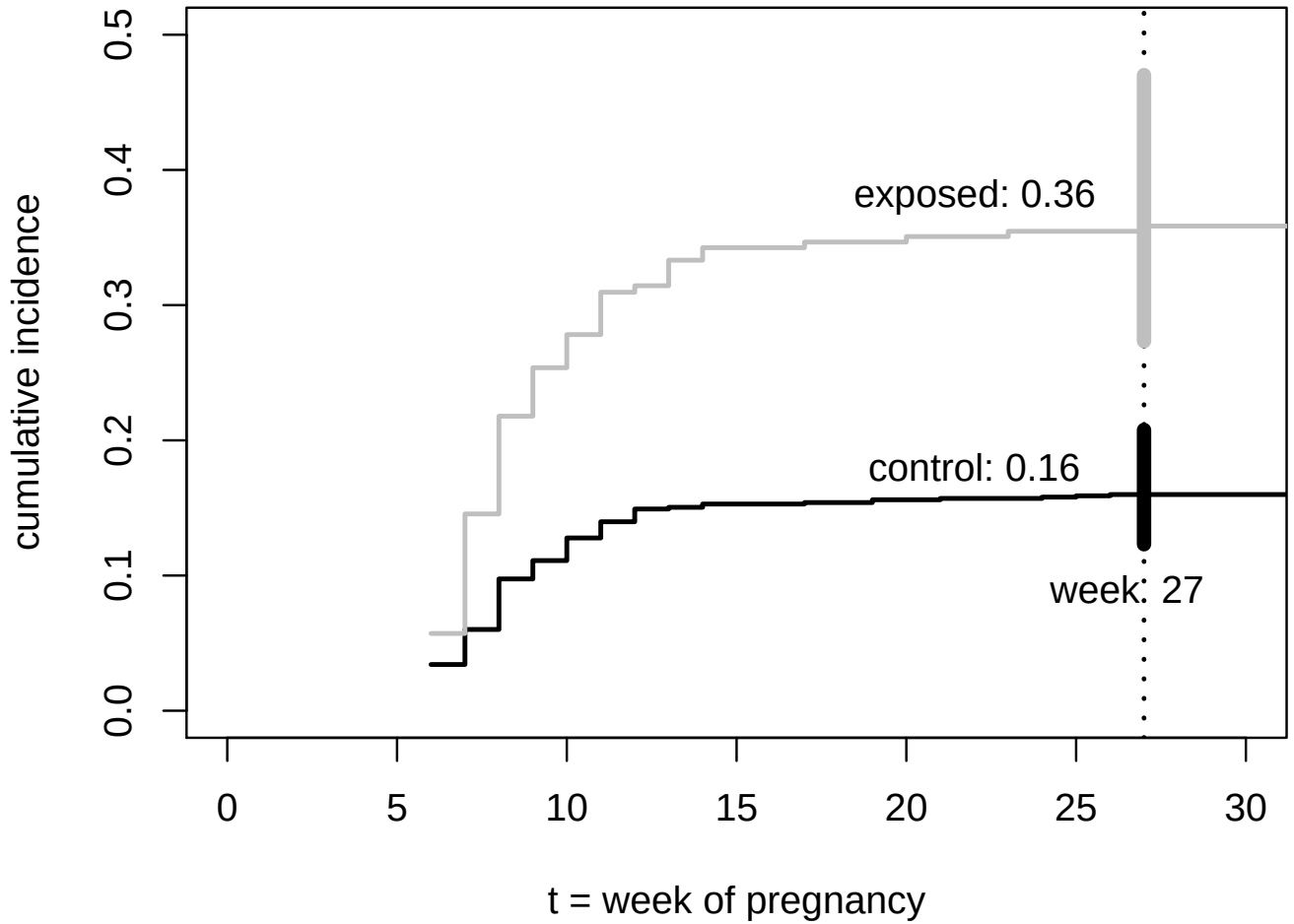
Computations Using **R**

Cumulative incidence and variance

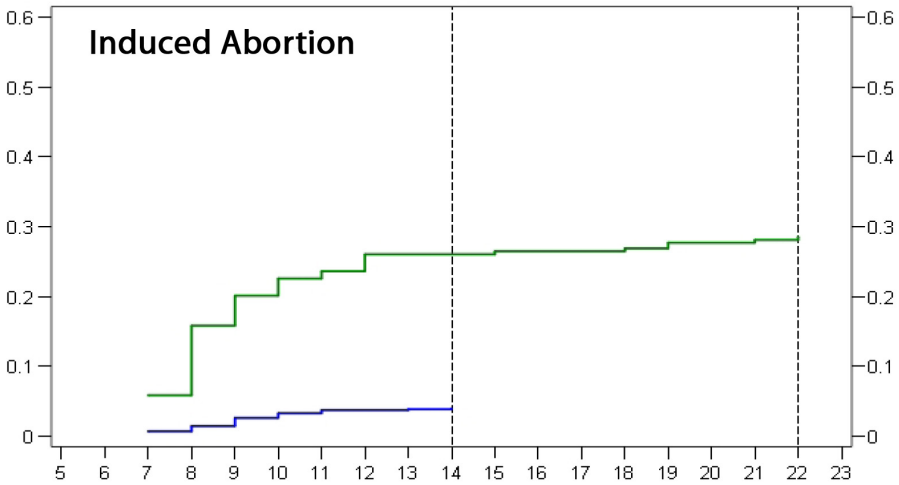
```
F.hat.var<-function(t.start,t.stop,cause,dataset)
    {
    ... compute "at risk": n, "events", "all": d.all,
        "specific": d "Survival": S
   f <- cumsum(S*d/n) # cumulative incidence rate
    ...
     del.f <- lower.triangle(outer(f,f,FUN="-"))
    v<-(del.f^2) %*% ( d.all/(n*(n-d.all)) )
       + cumsum(S^2*d*(n-d)/(n^3)) -2*del.f %*% (S*d/(n^2))
    ...
    return(data.frame(...))
    }
```

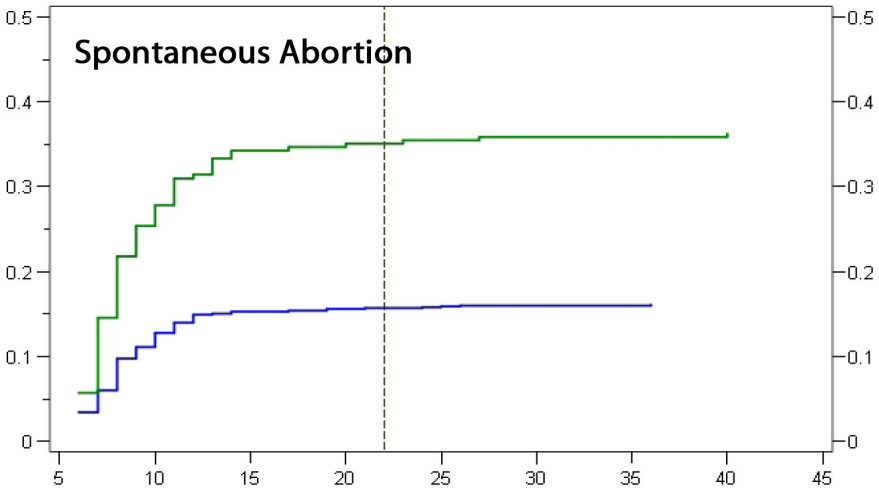
$$\widehat{\mathrm{Var}}\left[\widehat{F}_k(t)\right] = \sum_{t_i \le t} \left(\left[\widehat{F}_k(t) - \widehat{F}_k(t_i)\right]^2 \frac{d_i}{n_i(n_i - d_i)} + \left[\widehat{S}(t_i-)\right]^2 \frac{(n_i - d_{ki})}{n_i} \frac{d_{ki}}{n_i^2} - 2\left[\widehat{F}_k(t) - \widehat{F}_k(t_i)\right] \widehat{S}(t_i-) \frac{d_{ki}}{n_i^2} \right)$$

Coumarin Study Cumulative Incidence Spontaneous Abortion

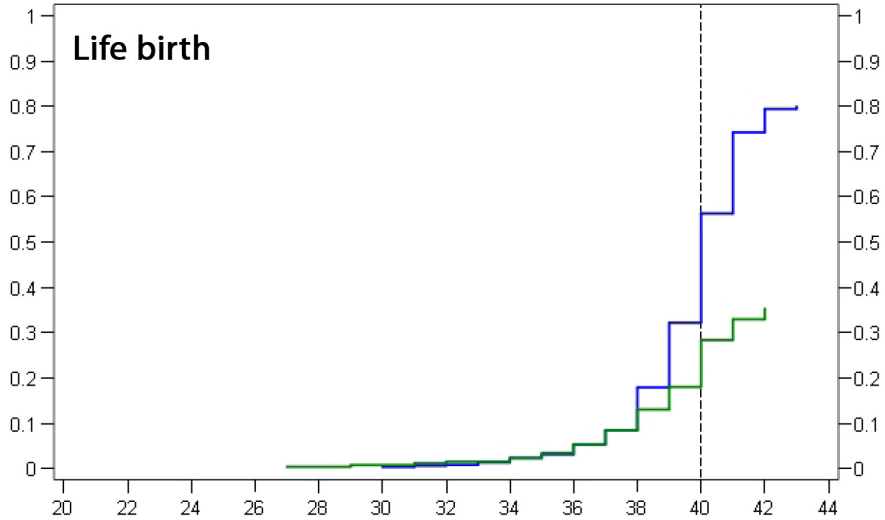


Coumarin Study Cumulative Incidence All Causes





control, exposed



Coumarin Study Cumulative Incidence All Causes

Estimated rates and standard errors

Proposed estimates	p exposed	SE	p control	SE
live birth	0.35	0.01	0.80	0.001
induced abortions	0.29	0.04	0.04	0.01
spontaneous abortions	0.36	0.05	0.16	0.02
Crude rates				
live birth	0.53		0.91	
induced abortion	0.22		0.02	
spontaneous abortion	0.25		0.07	

Modeling Effects under Competing Risks

Constant hazard

- hazard
  $\lambda_k, k = 1, \ldots K$ constant per cause
- covariates
  $\lambda_k(x) = \lambda_k \exp(x\beta_k)$
  cause-specific proportional hazards
- cumulative incidence

$$F_k(t, x) = \frac{\lambda_k \exp(x\beta_k)}{\sum_j \lambda_j \exp(x\beta_j)}(1 - \exp\{-t \sum_j \lambda_j \exp(x\beta_j)\})$$

- demonstration in *Mathematica*

Modeling Effects under Competing Risks

Basics

- time

  partition $0 = t_0 < t_1 < \cdots < t_m$

- gaps

  $\boldsymbol{\Delta} = (t_1, t_2 - t_1, \ldots, t_m - t_{m-1})$

- baseline

  cause specific hazards, piecewise constant

  $$\mathbf{L} = \begin{pmatrix} \lambda_{11} & \ldots & \lambda_{1m} \\ \vdots & \ddots & \vdots \\ \lambda_{K1} & \ldots & \lambda_{Km} \end{pmatrix}$$ combined: $\boldsymbol{\gamma} = \mathbf{1}'\mathbf{L}$

Modeling Effects under Competing Risks

Cause specific Cox models for piecewise constant hazard

- effects

  cause specific proportional hazards: $\boldsymbol{\lambda}_k(x) = \rho_k(x)\boldsymbol{\lambda}_k(0)$

  $$\mathbf{L}^x = \begin{pmatrix} \rho_1(x) & 0 \dots & 0 \\ 0 & \ddots & \vdots \\ 0 & & \rho_K(x) \end{pmatrix} \mathbf{L}, \quad \text{combined: } \boldsymbol{\gamma}^x = \mathbf{1}'\mathbf{L}^x$$

- cumulative hazard $\boldsymbol{\Gamma} = (\boldsymbol{\gamma}\boldsymbol{\Delta}) \begin{pmatrix} 1 & 1 & \dots & 1 \\ 0 & 1 & \dots & 1 \\ 0 & 0 & \ddots & \vdots \\ 0 & \dots & 0 & 1 \end{pmatrix}$

  covariate effects: $\boldsymbol{\Gamma}^x = \text{cumsum}(\boldsymbol{\gamma}^x\boldsymbol{\Delta})$

Explicit Formula for Cumulative Incidence Functions

Cause specific Cox models for piecewise constant hazard

- overall survival

  $\mathbf{S} = \exp(-\mathbf{\Gamma})$ with covariate effects $\mathbf{S}^x = \exp(-\mathbf{\Gamma}^x)$

  no proportional hazards model in general

- cumulative incidence

  $\mathbf{F}_k^x = \frac{\boldsymbol{\lambda}_k^x}{\gamma^x}(1 - \exp\{-\gamma^x \mathbf{\Delta}\})SC_-^x$, where

$$SC_-^x = \begin{pmatrix} 1 & 1 & \dots & 1 \\ 0 & S_1^x & \dots & S_1^x \\ 0 & 0 & \ddots & \vdots \\ 0 & \dots & 0 & S_{m-1}^x \end{pmatrix}$$

Discussion

- abortion rates estimated from observational studies give important information on prenatal risk
- using event history methods, prospective observational studies allow reliable assessment of abortion rates
- ignoring delayed entry results in a negative bias, ignoring competing risks tends to positive bias
- crude abortion rates should not be used
- coumarin-study TIS Berlin: reasonable baseline values; significant association of spontaneous abortions with exposure
- assumption of constant cause specific hazard yields simple explicit expression for cumulative incidence
- concept can be generalized to piecewise constant hazards – useful for generalization of Cox regression to competing risk setting

Thank you for your attention