# Survival Analysis I (CHL5209H)

Olli Saarela

Dalla Lana School of Public Health
University of Toronto

olli.saarela@utoronto.ca

January 19, 2019

# Literature

- Clayton D & Hills M (1993): *Statistical Models in Epidemiology*. Not really useful as a reference text but interesting pedagogical approach.
- Kalbfleisch JD & Prentice RL (2002): *The Statistical Analysis of Failure Time Data, Second Edition*. Introductory, serves as a reference text.
- Klein JP & Moeschberger ML (2003): *Survival Analysis - Techniques for Censored and Truncated Data, Second Edition*. Introductory, serves as a reference text.
- Aalen OO, Borgan Ø, Gjessing H (2008): *Survival and Event History Analysis - A Process Point of View*. For those looking for something more theoretical.

# Models for survival

- Survival analysis focuses on a single event per individual (say, first marriage, graduation, diagnosis of a disease, death). Analysis of multiple events would be referred to as event history analysis.
- In principle we could model survival times $T_i$ by specifying a linear model for its logarithm, such as

$$\log T_i = \alpha + \beta' X_i + \sigma\varepsilon_i,$$

  where $X_i$ are individual-level covariates, and where some error distribution is assumed for $\varepsilon_i$.
- We will see some examples of such parametric survival models later.
- The immediate problem with such models is that we cannot fit them using standard regression methods.
- This is because, due to *censoring*, we do not observe the event time for everyone.

# Models for hazard function

- An alternative approach to modeling survival is to model a different quantity, the *rate parameter*, through e.g.

$$\log \lambda_i = \alpha + \beta' X_i,$$

or the time-dependent version, the *hazard function*, through e.g.

$$\log \lambda_i(t) = \alpha(t) + \beta' X_i.$$

- Note that the regression coefficients now have a very different interpretation compared to the previous log-linear survival model.
- Survival probability is determined by the hazard function. We will discuss this connection in detail shortly.

# More about rates

- The rates can be for example mortality or incidence rates.
- Suppose for now that we do not have individual-level covariates and the rate is assumed the same for everyone: $\lambda_i = \lambda$.
- Rate parameter is the parameter of the Poisson distribution, characterizing the rate of occurrence of the events of interest.
- The expected number of events $\mu$ in a total of $Y$ years of follow-up time and $\lambda$ are connected by

$$\mu = \lambda Y.$$

- The observed number of events $D$ in $Y$ years of follow-up time is distributed as $D \sim \mathrm{Poisson}(\lambda Y)$.
- How to estimate the rate parameter $\lambda$?

# An estimator for $\lambda$

- A possible estimator is suggested by

$$\mu = \lambda Y \quad \Leftrightarrow \quad \lambda = \frac{\mu}{Y}.$$

- It would seem reasonable to replace here the expected number of events $\mu$ with the observed number of events $D$ and take

$$\hat{\lambda} = \frac{D}{Y}.$$

- This is known as the empirical rate.

# Follow-up data

- Clayton & Hills (1993, p. 41):

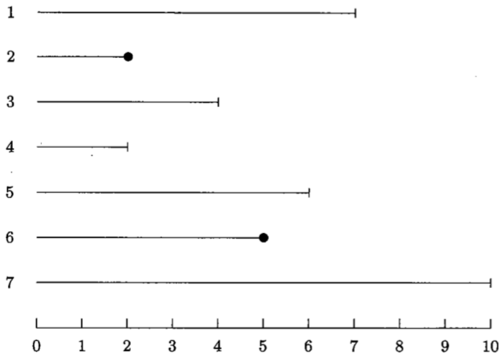


**Fig. 5.1.** The follow-up experience of 7 subjects.

# Estimate

- For the 7 subjects (individuals) there is a total of 36 time units of follow-up time/person-time, and 2 outcome events (for individuals 2 and 6).
- The follow-up of the other individuals was terminated by censoring (e.g. by events other than the outcome event of interest).
- Now
$$\hat{\lambda} = \frac{D}{Y} = \frac{2}{36} \approx 0.056.$$
- To recap:
  - Estimand/parameter/object of inference: $\lambda$
  - Estimator: $\frac{D}{Y}$
  - Estimate: 0.056.

# Maximum likelihood criterion

- The empirical rate $\hat{\lambda} = \frac{D}{Y}$ is in fact a maximum likelihood estimator.
- Maximum likelihood estimate is the value that maximizes the probability of observing the data.
- The probability of the observed data is given by the statistical model, which is now

$$D \sim \text{Poisson}(\lambda Y).$$

- Probabilities under the Poisson distribution are given by

$$P(D; \lambda) = \frac{(\lambda Y)^D}{D!} e^{-\lambda Y}.$$

- We consider this probability as a function of $\lambda$, and call it the likelihood of $\lambda$.
- Which value of $\lambda$ maximizes the likelihood?

# Maximizing the likelihood

- We may ignore any multiplicative terms not depending on the parameter, and instead maximize the expression
$$L(\lambda) = \lambda^D e^{-\lambda Y}.$$
- Or, for mathematical convenience, its logarithm
$$l(\lambda) = D \log \lambda - \lambda Y.$$
- How to find the argument value which maximizes a function?
- Set the first derivative to zero and solve w.r.t. $\lambda$:
$$l'(\lambda) = \frac{D}{\lambda} - Y = 0 \Leftrightarrow \lambda = \frac{D}{Y}.$$
- Check that the second derivative is negative:
$$l''(\lambda) = -\frac{D}{\lambda^2} < 0.$$
- It is, so we take $\hat{\lambda} = \frac{D}{Y}$ to be the maximum likelihood estimator.

# Approximate likelihoods

- With $D = 7$ outcome events observed in $Y = 500$ person-years of follow-up, $\hat{\lambda} = 7/500 = 0.014$, and the log-likelihood function would look like (Clayton & Hills 1993, p. 81)

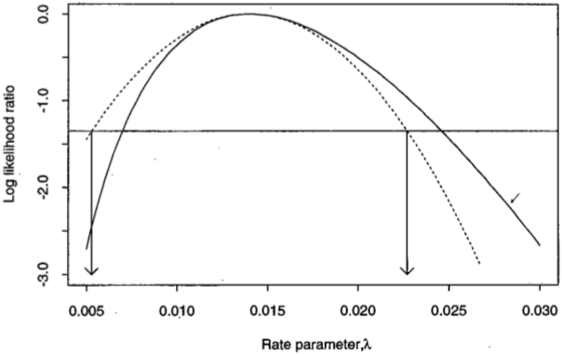


**Fig. 9.2.** True and approximate Poisson log likelihoods.

# Approximate likelihoods (2)

- The dotted line is a quadratic curve centered at $\hat{\lambda}$.
- The logarithm of normal density w.r.t. to the mean parameter is a quadratic curve, with the second derivative being equivalent to negative inverse of the variance.
- This implies that the inverse of negative second derivative of the log-likelihood has something to do with the variance of $\hat{\lambda}$. (Why?)
- The normal approximation means that we take $\hat{\lambda}$ to approximately normally distributed with variance $\frac{\lambda^2}{D} \approx \frac{(D/Y)^2}{D} = D/Y^2$.
- Thus, the *standard error* of $\hat{\lambda}$ is $\sqrt{D}/Y$.
- Unfortunately, because $\lambda$ is non-negative, this approximation may not be very good.
- The log-likelihood for $\log \lambda$ should be more symmetric (Clayton & Hills 1993, p. 82):

# Approximate likelihoods (3)

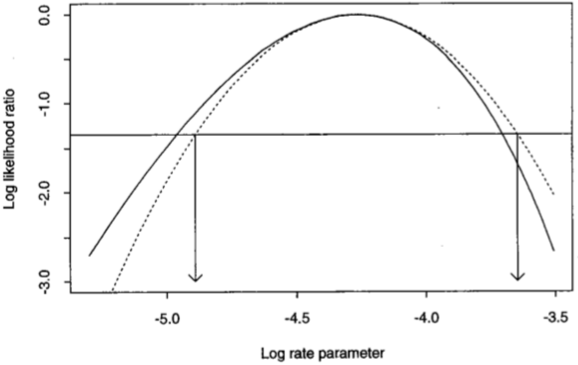


**Fig. 9.3.** Approximating the log likelihood for $\log(\lambda)$.

If we denote $\alpha = \log \lambda$, the first derivative of the log-likelihood $l(\alpha) = D\alpha - e^{\alpha} Y$ is $l'(\alpha) = D - e^{\alpha} Y$, and the second derivative is $l''(\alpha) = -e^{\alpha} Y \approx -e^{\log(D/Y)} Y = -D$, giving the familiar standard error $\sqrt{1/D}$ for $\log \hat{\lambda}$.

# Interpretation of the rate parameter

- Unlike the *risk parameter*, the probability of an event occurring within a specific time period, the rate parameter does not correspond to a follow-up period of a fixed length.
- Rather, it characterizes the instantaneous occurrence of the outcome event at any given time.
- The rate parameter is not a probability, but it can be characterized in terms of the risk parameter when the follow-up period is very short.

# Time unit

- Suppose that each of the $N = 36$ time bins here is of length $h = 0.05$ years:

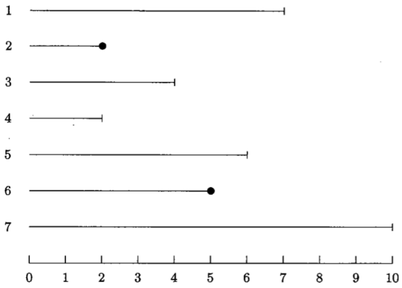


**Fig. 5.1.** The follow-up experience of 7 subjects.

- In total there is $Y = Nh = 36 \times 0.05 = 1.8$ years of follow-up.

# From risk to rate

- The empirical rate is given by $\hat{\lambda} = \frac{2}{1.8} = 1.11$ per person-year, or, say, 1110 per 1000 person-years.
- Per person-year, the empirical rate would be the same, had we instead split the person-time into 180 bins of length 0.01 years.
- Suppose that we have made the time bins short enough so that at most one event can occur in each bin.
- Whether an event occurred in a particular bin of length $h$ is now a Bernoulli-distributed variable, with the expected number of events equal to the risk $\pi$.
- Thus, because rate is the expected count divided by person-time, when $h$ is small, we have

$$\lambda = \frac{\pi}{h} \quad \Leftrightarrow \quad \pi = \lambda h.$$

- This connection is important in understanding how rate is related to survival probability.

# Survival probability

- One of the particular properties of the natural logarithm and its inverse is that when $x$ is close to zero, $e^x \approx 1 + x$, and conversely, $\log(1 + x) \approx x$.
- Suppose that we are interested in the probability of surviving $T$ years. By splitting the timescale so that $N = \frac{T}{h}$, $T = Nh$.
- The probability of surviving through a single time bin of length $h$, conditional on surviving until the start of this interval, is $1 - \pi = 1 - \lambda h$.
- By the multiplicative rule, the $T$ year survival probability is thus
$$(1 - \lambda h)^N.$$
- This motivates the well-known *Kaplan-Meier estimator*, to be encountered later.
- In turn, the logarithm of this is
$$N \log(1 - \lambda h) \approx -N\lambda h = -\lambda T.$$

# Survival and cumulative hazard

- The quantity $\lambda T$ is known as the *cumulative hazard*.
- We have (approximately, without calculus) obtained a fundamental relationship of survival analysis, namely that the $T$ year survival probability is

$$(1 - \lambda h)^N \approx e^{-\lambda T}.$$

- Let us test whether this approximation actually works. Now $\hat{\lambda} = 1.11$.
- If $T = 1$ and $h = 0.05$, $N = 20$ and we get $(1 - 1.11 \times 0.05)^{20} \approx 0.319$.
- The exact one year survival probability is $e^{-1.11 \times 1} \approx 0.330$.
- We should get a better approximation through a finer split of the time scale.
- If $h = 0.01$, $N = 100$ and $(1 - 1.11 \times 0.01)^{100} \approx 0.328$.

# Regression models

- Recall the relationship $\mu = \lambda Y$. If $\alpha = \log \lambda$, we have the equivalent log-linear form

$$\log \mu = \alpha + \log Y,$$

where we call $\log Y$ an *offset* term.

- $\alpha$ is an unknown parameter, which we could estimate in an obvious way. (How?)
- Such a one-parameter model is not very interesting, but serves as a starting point to regression modeling.
- Consider now the expected number of events $\mu_1$ in $Y_1$ years of exposed person-time and the expected number of events $\mu_0$ in $Y_0$ years of unexposed person-time.
- The corresponding probability models are now

$$D_1 \sim \text{Poisson}(\mu_1) \quad \text{and} \quad D_0 \sim \text{Poisson}(\mu_0),$$

where $\mu_1 = \lambda_1 Y_1$ and $\mu_0 = \lambda_0 Y_0$.

# Combining the two models

- We may now *parametrize* the two log-rates in terms of an *intercept term* $\alpha$ and a *regression coefficient* $\beta$ as $\log \lambda_0 = \alpha$ and $\log \lambda_1 = \alpha + \beta$.
- By introducing an exposure variable $Z$, with $Z = 1$ ($Z = 0$) indicating the exposed (unexposed) person-time, we can express these definitions as a regression equation

$$\log \lambda_Z = \alpha + \beta Z \quad \Leftrightarrow \quad \lambda_Z = e^{\alpha + \beta Z}.$$

- This results in a single statistical model, namely

$$D_Z \sim \text{Poisson}\left(Y_Z e^{\alpha + \beta Z}\right).$$

- What is the interpretation of the regression coefficient?
- Now we have

$$\frac{\lambda_1}{\lambda_0} = \frac{e^{\alpha+\beta}}{e^{\alpha}} = \frac{e^{\alpha} e^{\beta}}{e^{\alpha}} = e^{\beta},$$

or $\beta = \log\left(\frac{\lambda_1}{\lambda_0}\right)$, that is, the log rate ratio.

# Likelihood for a rate ratio

- With the two Poisson ditributions $D_0 \sim \mathrm{Poisson}(Y_0 e^{\alpha})$ and $D_1 \sim \mathrm{Poisson}(Y_1 e^{\alpha+\beta})$, the log-likelihood becomes

$$l(\alpha, \beta) = D_0\alpha - e^{\alpha} Y_0 + D_1(\alpha + \beta) - e^{\alpha+\beta} Y_1.$$

- This may be maximized w.r.t. $\alpha$ and $\beta$ simultaneously.
- The maximum likelihood estimators do not necessarily have closed form solutions; this need not concern us, since the likelihood can be maximized, and the derivatives calculated, numerically.
- In fact, this is what a procedure such as the R `glm` function does.

# Reparametrizing rates

- The model can be easily extended to accommodate more than one covariate.
- For example, unadjusted comparisons of rates are susceptible to confounding; we can move on to consider confounder-adjusted rate ratios.
- Consider the following dataset:

**Table 22.6.** Energy intake and IHD incidence rates per 1000 person-years

| | Unexposed ($\geq$ 2750 kcals) | | | Exposed ($<$ 2750 kcals) | | | Rate |
|---|---|---|---|---|---|---|---|
| Age | Cases | P-yrs | Rate | Cases | P-yrs | Rate | ratio |
| 40–49 | 4 | 607.9 | 6.58 | 2 | 311.9 | 6.41 | 0.97 |
| 50–59 | 5 | 1272.1 | 3.93 | 12 | 878.1 | 13.67 | 3.48 |
| 60–69 | 8 | 888.9 | 9.00 | 14 | 667.5 | 20.97 | 2.33 |

- Introduce an exposure variable taking values $Z = 1$ (energy intake $< 2750$ kcals) and $Z = 0$ ($\geq 2750$ kcals), and an age group indicator taking values $X = 0$ $(40 - 49)$, $X = 1$ $(50 - 59)$ and $X = 2$ $(60 - 69)$.

# The original parameters

- There are now six rate parameters $\lambda_{ZX}$, corresponding to each exposure-age combination:

$$
\begin{array}{ccc}
 & Z=0 & Z=1 \\
X=0 & \lambda_{00} & \lambda_{10} \\
X=1 & \lambda_{01} & \lambda_{11} \\
X=2 & \lambda_{02} & \lambda_{12}
\end{array}
$$

- The corresponding statistical distributions are

$$
\begin{array}{lll}
 & Z=0 & Z=1 \\
X=0 & D_{00} \sim \text{Poisson}(Y_{00}\lambda_{00}) & D_{10} \sim \text{Poisson}(Y_{10}\lambda_{10}) \\
X=1 & D_{01} \sim \text{Poisson}(Y_{01}\lambda_{01}) & D_{11} \sim \text{Poisson}(Y_{11}\lambda_{11}) \\
X=2 & D_{02} \sim \text{Poisson}(Y_{02}\lambda_{02}) & D_{12} \sim \text{Poisson}(Y_{12}\lambda_{12})
\end{array}
$$

# Transformed parameters

- Now, we are not primarily interested in estimating six rates; rather, we are interested in the rate ratio between the exposure categories, adjusting for age.
- We could parametrize the rates w.r.t. the baseline, or reference, rate $\lambda_{00}$ which is then modified by the exposure and age (cf. Clayton & Hills 1993, p. 220).
- Define

$$\begin{array}{lll} & Z=0 & Z=1 \\ X=0 & \lambda_{00}=\lambda_{00} & \lambda_{10}=\lambda_{00}\theta \\ X=1 & \lambda_{01}=\lambda_{00}\phi_1 & \lambda_{11}=\lambda_{00}\theta\phi_1 \\ X=2 & \lambda_{02}=\lambda_{00}\phi_2 & \lambda_{12}=\lambda_{00}\theta\phi_2 \end{array}$$

- Now $\theta$ is the rate ratio within each age group (verify).

# Regression parameters

- As before, we can specify the reparametrization in terms of a link function and a linear predictor as

$$\log \lambda_{ZX} = \alpha + \beta Z + \gamma_1 \mathbf{1}_{\{X=1\}} + \gamma_2 \mathbf{1}_{\{X=2\}}.$$

- Since

$$\lambda_{ZX} = e^{\alpha + \beta Z + \gamma_1 \mathbf{1}_{\{X=1\}} + \gamma_2 \mathbf{1}_{\{X=2\}}},$$

  we have that $\lambda_{00} = e^{\alpha}$, $\theta = e^{\beta}$, $\phi_1 = e^{\gamma_1}$ and $\phi_2 = e^{\gamma_2}$.

- The rates are now given by the regression equation as

$$\begin{array}{lll} & Z = 0 & Z = 1 \\ X = 0 & \lambda_{00} = e^{\alpha} & \lambda_{10} = e^{\alpha+\beta} \\ X = 1 & \lambda_{01} = e^{\alpha+\gamma_1} & \lambda_{11} = e^{\alpha+\beta+\gamma_1} \\ X = 2 & \lambda_{02} = e^{\alpha+\gamma_2} & \lambda_{12} = e^{\alpha+\beta+\gamma_2} \end{array}$$

- The number of parameters has been reduced from six to four.

# Specification in terms of expected counts

- A Poisson model is always specified in terms of the expected event count: $D_{ZX} \sim \text{Poisson}(\mu_{ZX})$.
- The regression model for the expected count is specified by

$$\begin{aligned}\mu_{ZX} = Y_{ZX}\lambda_{ZX} &= Y_{ZX}\, e^{\alpha+\beta Z+\gamma_1 \mathbf{1}_{\{X=1\}}+\gamma_2 \mathbf{1}_{\{X=2\}}} \\ &= e^{\alpha+\beta Z+\gamma_1 \mathbf{1}_{\{X=1\}}+\gamma_2 \mathbf{1}_{\{X=2\}}+\log Y_{ZX}}.\end{aligned}$$

- We have obtained the model

$$D_{XZ} \sim \text{Poisson}\left(e^{\alpha+\beta Z+\gamma_1 \mathbf{1}_{\{X=1\}}+\gamma_2 \mathbf{1}_{\{X=2\}}+\log Y_{ZX}}\right).$$

- When fitting the model, log-person years has to be included in the linear predictor as an offset variable.

# Fitting the model

- The data as frequency records are entered into R as:

```
d <- c(4,5,8,2,12,14)
y <- c(607.9,1272.1,888.9,311.9,878.1,667.5)
z <- c(0,0,0,1,1,1)
x <- c(0,1,2,0,1,2)
```

- The model is specified as

```
model <- glm(d ~ z + as.factor(x) +
                 offset(log(y)),
             family=poisson(link="log"))
```

- The `as.factor(x)` term specifies that we want to estimate separate age group effects (rather than assume that the $X$-variable modifies the log-rate additively).

# Results

```
Call:
glm(formula = d ~ z + as.factor(x) + offset(log(y)),
    family = poisson(link = "log"))

Deviance Residuals:
       1         2         3         4         5         6
 0.73940  -0.58410   0.04255  -0.77385   0.42800  -0.03191

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)    -5.4177     0.4421 -12.256  < 2e-16 ***
z               0.8697     0.3080   2.823  0.00476 **
as.factor(x)1   0.1290     0.4754   0.271  0.78609
as.factor(x)2   0.6920     0.4614   1.500  0.13366
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 14.5780  on 5  degrees of freedom
Residual deviance:  1.6727  on 2  degrees of freedom
AIC: 31.796

Number of Fisher Scoring iterations: 4
```

# Proportional hazards

- From the model output, we may calculate estimates for the original rate parameters (per 1000 person-years) as

$$\begin{array}{lll} & Z=0 & Z=1 \\ X=0 & \hat{\lambda}_{00}=4.44 & \hat{\lambda}_{10}=10.59 \\ X=1 & \hat{\lambda}_{01}=5.05 & \hat{\lambda}_{11}=12.05 \\ X=2 & \hat{\lambda}_{02}=8.86 & \hat{\lambda}_{12}=21.20 \end{array}$$

- Note that the rate ratio stays constant across the age groups. This is forced by the earlier model specification.
- This is a modeling assumption, namely the *proportional hazards* assumption.
- Compare these estimates to the corresponding six empirical rates. Is assuming proportionality of the hazard rates justified? How could one test this? Or relax this assumption?

# Basic concepts

# Time-to-event outcome

- In survival analysis, the outcome data are realized values for a pair of random variables $(T_i, E_i)$, where $T_i$ represents the observed time when something happened, and $E_i$ the type of the event that occurred at $T_i$.
- Usually, we have to consider at least two types of events, namely the outcome event of interest (say, $E_i = 1$), and censoring (say, $E_i = 0$), that is, termination of the follow-up due to some other reason than the outcome event of interest.
- However, we are not interested in modeling the censoring events; we are only interested in what characterizes the outcome events.
- To express this, suppose that the observed time is given by $T_i = \min\{\tilde{T}_i, C_i\}$, where $\tilde{T}_i$ and $C_i$ are latent event and censoring times.
- We can now define the event indicator as $E_i = \mathbf{1}_{\{T_i = \tilde{T}_i\}}$.

# Hazard function

- The hazard function is defined in terms of the latent event time as

$$\lambda(t) \equiv \lim_{h \to 0} \frac{P(t \leq \tilde{T}_i < t + h \mid \tilde{T}_i \geq t)}{h}.$$

- Corresponding to the previous discussion, the probability interpretation of this is

$$\lambda(t)\,\mathrm{d}t = P(t \leq \tilde{T}_i < t + \mathrm{d}t \mid \tilde{T}_i \geq t).$$

- The probability $P(\tilde{T}_i \geq t) \equiv S(t)$ is known as the survival function.

# Connection between hazard and survival functions

- Now

$$P(t \leq \tilde{T}_i < t + \mathrm{d}t \mid \tilde{T}_i \geq t) = \frac{P(t \leq \tilde{T}_i < t + \mathrm{d}t)}{P(\tilde{T}_i \geq t)}$$

$$\Leftrightarrow \lambda(t) = \frac{f(t)}{S(t)},$$

where $f(t)$ is the density function of the event time distribution, interpreted through

$$f(t)\,\mathrm{d}t = P(t \leq \tilde{T}_i < t + \mathrm{d}t).$$

# Connection between hazard and survival functions (2)

- Note that $S(t) = 1 - F(t)$ and $f(t) = \frac{\mathrm{d}F(t)}{\mathrm{d}t}$, where $F(t) \equiv P(\tilde{T}_i \leq t)$.
- Further, $\frac{\mathrm{d}[\log F(t)]}{\mathrm{d}t} = \frac{f(t)}{F(t)}$ and $-\frac{\mathrm{d}[\log S(t)]}{\mathrm{d}t} = \frac{f(t)}{S(t)} = \lambda(t)$.
- Because $S(0) = 1$, this gives us again the fundamental relationship

$$S(t) = \exp\left\{ -\int_0^t \lambda(u)\,\mathrm{d}u \right\},$$

where $\int_0^t \lambda(u)\,\mathrm{d}u \equiv \Lambda(t)$ is the cumulative hazard.

# Counting process notation

- We will occasionally encounter counting process notation, which is an alternative way to represent the framework.
- What is a process?
- The *counting process* $\{\tilde{N}_i(t), t \geq 0\}$ for the outcome event of interest is defined through

$$\tilde{N}_i(t) = \mathbf{1}_{\{\tilde{T}_i \leq t\}}.$$

- In survival analysis, the counting process only counts to one, as we only consider the first event.
- The *at-risk process* $\{Y_i(t), t \geq 0\}$ (needed later) is defined through

$$Y_i(t) \equiv \mathbf{1}_{\{T_i \geq t\}}.$$

# Counting process jump

- Whether an event happens exactly at time $t$ for individual $i$ is recorded by the counting process jump

$$\mathrm{d}\tilde{N}_i(t) \equiv \tilde{N}_i(t^- + \mathrm{d}t) - \tilde{N}_i(t^-).$$

- We can now define the hazard function equivalently through

$$\begin{aligned} P(\mathrm{d}\tilde{N}_i(t) = 1 \mid \tilde{N}_i(t^-) = 0) &= E[\mathrm{d}\tilde{N}_i(t) \mid \tilde{N}_i(t^-) = 0] \\ &= P(t \leq \tilde{T}_i < t + \mathrm{d}t \mid \tilde{T}_i \geq t) \\ &= \lambda(t)\,\mathrm{d}t. \end{aligned}$$

- We can understand hazard models as modeling of the expected counting process jump.

# Competing risks

- The survival model generalizes straightforwardly to situation where we may have more than one mutually exclusive event type of interest.
- The time $\tilde{T}_i$ refers to the time of the first event of interest (of any type), but in addition we introduce a latent event type indicator taking values $\tilde{E}_i \in \{1, \ldots, J\}$.
- Equivalently, we could introduce the cause-specific counting processes $\tilde{N}_{ij}(t)$, $j = 1, \ldots, J$.

# Cause-specific hazards

- We may now define cause-specific hazard functions for each event type $j \in \{1, \ldots, J\}$ through

$$\lambda_j(t) \equiv \lim_{h \to 0} \frac{P(t \leq \tilde{T}_i < t + h, \tilde{E}_i = j \mid \tilde{T}_i \geq t)}{h}.$$

- The sub-density function corresponding to event type $j$ is given by the relationship

$$\begin{aligned} f_j(t)\,\mathrm{d}t &= P(t \leq \tilde{T}_i < t + \mathrm{d}t, \tilde{E}_i = j) \\ &= P(t \leq \tilde{T}_i < t + \mathrm{d}t, \tilde{E}_i = j \mid \tilde{T}_i \geq t) P(\tilde{T}_i \geq t) \\ &= \lambda_j(t)\,\mathrm{d}t \exp\left\{ -\int_0^t \textstyle\sum_{k=1}^J \lambda_k(u)\,\mathrm{d}u \right\}, \end{aligned}$$

where the overall survival term is the probability that none of the events occurred by time $t$.