Chapter 9

Other Models and Topics

Recurrent Events

- So far we have assumed that subjects can only experience the event of interest once
  - Original application of survival analysis was to analyze time to death
- There are applications where the event can occur more than once.
  - Examples: heart attacks among patients with heart disease, relapses for MS patients

Simple Solutions

- One way to avoid the issue of recurrent events is to focus on the first occurrence of the event.
  - Advantage: Can use the methods we have already learned
  - Disadvantage: Are not using all of the information that we have

- Another approach is to perform separate analyses for each successive event
  - First, analyze time until first event for all subjects.
  - Then, analyze the time until the second event for those subjects who had at least two (the survival time is the time between the first and second events)
  - Continue until you run out of events or sample size gets too small.

Simple Solutions

- Problems with this approach:
  - Have to do lots of analyses if subjects had lots of events
  - Will get several sets of coefficients that will need to be interpreted
  - Will be inefficient if the effect of covariates does not change with event number

Andersen-Gill Model

- The counting process approach is the simplest way of modeling recurrent events
  - Also known as the Andersen-Gill (AG) model.
  - Each subject represented by series of observations
    - $(0, t_1]$, $(t_1, t_2]$, $(t_2, t_3]$, ….
    - Number of observations is the number of observed/censored events the subject had
  - $\log(h(t|\beta,x)) = \log(h_0(t)) + \beta x$

Andersen-Gill Model

- Assumptions for Andersen-Gill:
  - Events are assumed to be independent
  - Hazard does not change after an event occurs
  - Subject contributes to the risk set for an event as long as they are still in the study
- Andersen-Gill does not take into account the order of events

Prentice-Williams-Peterson Model

- The Prentice-Williams-Peterson model is similar to the AG model except the baseline hazard function varies by event
- Two choices of time scales for the Prentice-Williams-Peterson model:
  - Use total time from beginning of study (PWP-CP model)
    - Subjects represented in the same way as in AG model
    - $(0, t_1]$, $(t_1, t_2]$, $(t_2, t_3]$, ….
    - $\log(h_s(t|\beta,x)) = \log(h_{0s}(t)) + \beta x$ ← Have a stratum for each event

Prentice-Williams-Peterson Model

- Use the "gap time", i.e., the amount of time between events (PWP-GT model)
  - $(0, t_1]$, $(0, t_2 - t_1]$, $(0, t_3 - t_2]$, ….
  - $\log(h_s(t|\beta,x)) = \log(h_{0s}(t-t_{s-1})) + \beta x$

t_{s-1} is the time of the previous event

- Assumptions for Prentice-Williams-Peterson:
  - Subjects are not at risk for the k^{th} event until the $(k-1)^{th}$ event has occurred.
  - Shape of the hazard function allowed to vary with event number

Wei-Lin-Weissfeld Model

- The Wei-Lin-Weissfeld (WLW) model allows for each event to have its own baseline hazard function.
  - Each subject represented by a series of observations
    - Number of observations is the same for each subject and is the number of possible events
    - $(0, t_1]$, $(0, t_2]$, $(0, t_3]$, ….
  - $\log(h_s(t|\beta,x)) = \log(h_{0s}(t)) + \beta x$ ← Have a stratum for each event

Wei-Lin-Weissfeld Model

- Assumptions for Wei-Lin-Weissfeld:
  - Each event is treated as a separate process
  - Subjects are considered to be at risk for each event, no matter how many events they actually had
    - For example, a subject is treated as being at risk for a fourth event even if he or she only had two.

Adjusting for Correlation

- Each of these models (AG, PWP, WLW) are fit by letting each subject have multiple observations in the data set.
  - Events from the same subject will be correlated
  - Standard errors calculated using the information matrix will be incorrect

Adjusting for Correlation

- The robust variance estimator or the “sandwich estimator” can be used to account for the dependence among multiple event times.
  - The robust variance estimator is

$$\hat{R}\left(\hat{\beta}\right) = \hat{V}\left(\hat{\beta}\right)\left[\hat{L}^T\hat{L}\right]\hat{V}\left(\hat{\beta}\right)$$

where $\hat{V}\left(\hat{\beta}\right)$ is the estimate of the variance calculated using the information matrix and $\hat{L}$ is the vector of score residuals

Model Interpretations

- Which model is most appropriate depends on the application.
  - Look at each model's assumptions to see if those assumptions are reasonable for the data
- Andersen-Gill assumes events are independent and ignores event order.
  - A tenth event is assumed to have the same baseline hazard function as a first event.
  - Possible applications: Time to catching a cold, time until next bug hits your windshield

Model Interpretations

- For the Prentice-Williams-Peterson models, a subject is not at risk for a k^{th} event until they've had the $(k-1)^{th}$ event.
  - This assumption is useful for settings where the events occur in a "causal path" and each event affects the risk of having another event.
  - PWP-CP uses time defined from the beginning of the study (full time course), while PWP-GT "resets the clock" at each event

Model Interpretations

- The Wei-Lin-Weissfeld model subjects are at risk for all events for as long as they are under observation.
  - Used for “unordered events of the same type”.
  - The WLW model treats each event as its own process. WLW is sometimes called a “marginal model”.
  - Possible applications: Time to onset of infection in HIV-positive patients, family studies where each family member is at risk of the event

Summary

- AG is probably the most common recurrent event model, perhaps due to its simplicity

- Textbook recommends using PWP-CP (if you are interested modeling over the entire observation period) or PWP-GT (if you are interested in the time since last event).

Competing Risks

- So far we have assumed that there is only one event of interest
  - Usually, subjects are followed until either the event of interest occurs or until censoring
  - Sometimes the event can occur more than once (recurrent events).

- In some applications, the subject may “fail” due to more than one cause
  - Examples: Want to study the factors that affect time to initiation of dialysis in patients with chronic kidney disease. Some patients may die due to complications before they can start dialysis
  - May not want to treat death due to kidney disease as a censoring event, since it is unlikely to be independent of the outcome of interest

Competing Risks

- When follow-up may terminate due to more than one type of event, the different types of events are called competing risks.

- In a competing risks analysis, the occurrence of one type of event prevents us from observing the other events for that individual
  - A person who dies of a heart attack is no longer at risk of dying of a stroke.

Competing Risks

- One solution to dealing with competing risks is to use a composite endpoint.
  - Example: Cardiovascular studies often use a composite endpoint called MACE (major adverse cardiovascular event) which can include stroke, myocardial infarction, and death
  - Advantage: Can use the methods we have already learned

Competing Risks

- Disadvantage: Will be unable to determine if the covariates have different effects on the various events

- If we want to take into account the different types of events, we need a model that describes progression to each of the competing risks.

Notation

- Let T_k be the unobservable time to the occurrence of the k^{th} competing risk.
  - Assume that the K event types are mutually exclusive and that subjects are not followed after an event occurs.

- We can only observe the first event, so the observed survival time is

$$T = \min(T_1, \ldots T_K)$$

Notation

- We can also define an indicator C, which tells us which of the K events occurred:
  - $C = k$ if $T = T_k$ $k = 1, \ldots, K$

- Example: Competing risks are initiation of dialysis (1) and death due to kidney disease (2)
  - T_1 is the time until initiation of dialysis
  - T_2 is the time until death due to kidney disease
  - T is the observed survival time

Notation

- If the patient lives long enough to start dialysis, $T = T_1$ and $C = 1$.
  - T_2 is unknown
- If the patient dies from kidney disease before he or she starts dialysis, then $T = T_2$ and $C = 2$.
  - T_1 is unknown

- If the patient is lost to follow-up or dies from another cause (e.g., car accident), T is the time of censoring and $C = 0$.
  - As before, we consider censoring to be independent of the events.
  - This allows us to treat censoring as a separate process.

The Cause-Specific Hazard Function

- We need to generalize the hazard, cumulative hazard, and survival functions to handle more than one event.

- The cause-specific hazard function for risk *k* is

$$h_k(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T \leq t + \Delta t, C = k \mid T \geq t)}{\Delta t}$$

  - This is the conditional probability that an event of type *k* occurs between *t* and *t*+*Δt*, given that the subject had not already experienced an event by time *t*.

$$h(t) = \sum_{k=1}^{K} h_k(t)$$

The Cause-Specific Hazard Function

- Cause-specific cumulative hazard function:

$$H_k(t) = \int_0^t h_k(u)du$$

- Overall cumulative hazard:

$$H(t) = \sum_{k=1}^{K} H_k(t)$$

The Cumulative Incidence Function

- The cumulative incidence function (CIF) or cause-specific distribution function for an event type is defined as

$$F_k(t) = P(T \leq t, C = k) = \int_0^t h_k(u) \cdot S(u)\, du$$

  - Estimates the probability that the kth event type occurs before time *t* and that it occurs before any of the competing causes of failure
  - This represents the cumulative proportion of failures due to cause *k* at time *t*.

- The CIF is also sometimes called the subdistribution function because it is not really a CDF.

The Cumulative Incidence Function

- The overall distribution function is

$$F(t) = \sum_{k=1}^{K} F_k(t)$$

The Cumulative Incidence Function

- The estimator for the CIF is

$$\hat{F}_k(t) = \sum_{t_j \le t} \hat{h}_k(t_j) \cdot \hat{S}(t_{j-1})$$

  - where $\hat{S}(\cdot)$ is the Kaplan-Meier estimate and

$$\hat{h}_k(t_j) = \frac{d_{jk}}{n_j}$$

  d_{jk} is the number of subjects who failed from cause *k* at time t_j.

- $\hat{F}_k(t)$ will be a step function and will only increase when a subject experiences the k^{th} event.
- The variance of $\hat{F}_k(t)$ is derived using the delta method (see p. 331 of textbook).

Hypothesis Tests for Competing Risks

- To assess the effect of a categorical covariate on the cumulative incidence function:
  - Stratify on that covariate and estimate separate $F_k(t)$ for each stratum
  - Use Gray's test to test the null hypothesis

    H_0: $F_{k1}(t) = F_{k2}(t) = \ldots = F_{ks}(t)$ against the alternative H_1: At least two CIF are different
  - $F_{ks}(t)$ is the cumulative incidence function for the s^{th} stratum

Hypothesis Tests for Competing Risks

- To assess the effect of covariates on survival in the presence of competing risks, we have two options:
  - Consider how the covariate affects the cause-specific hazard, or
  - Consider how the covariate affects the cumulative incidence function.

- Note that testing for differences in the cumulative incidence function is not the same as testing for differences in the cause-specific hazard.

Hypothesis Tests for Competing Risks

- The cumulative incidence function depends on both the overall survival function and the cause-specific hazard function.

$$F_k(t) = \int_0^t S(u)h_k(u)du$$

Hypothesis Tests for Competing Risks

- Example: Want to look at the risk of bone fracture in a set of patients.
  - Event of interest is fracture, competing risk is death.
  - Suppose age has no effect on the risk of fracture, but being older does increase the risk of death
    - This would mean there would be a lower cumulative incidence of fracture in older subjects.
  - Test for difference in the cause-specific hazard for fracture would be not significant, but a test for difference in the cumulative incidence function would be significant.

Hypothesis Tests for Competing Risks

- To assess the effect of a categorical covariates on the cause-specific hazard:
  - Stratify on the covariate.
    - Let $h_{ks}(t)$ be the cause-specific hazard function for the s^{th} stratum
  - Use the log-rank test to test the hypotheses

$$H_0: h_{k1}(t) = h_{k2}(t) = \ldots = h_{kS}(t) \text{ vs } H_1: \text{Not } H_0$$

    - Treat any subjects who did not have the k^{th} event as censored.

Hypothesis Tests for Competing Risks

- To assess the effect of a categorical covariate on the cumulative incidence function:
  - Stratify on that covariate and estimate separate cumulative incidence functions for each stratum
    - Let $F_{ks}(t)$ be the cumulative incidence function for the s^{th} stratum
  - Use Gray's test to test the hypotheses

H_0: $F_{k1}(t) = F_{k2}(t) = \ldots = F_{ks}(t)$ vs H_1: At least two CIF differ

    - Grey's test is sort of a modified chi-square test.

Hypothesis Tests for Competing Risks

- Similar to the log-rank and Wilcoxon test for Kaplan-Meier estimates of the survival function, we can only do these tests if we have one, categorical covariate.
  - If we want to use continuous or multiple covariates, we need another method.

Cox Model for Cause-Specific Hazard

- The Cox proportional hazards model can be extended to model the effects of covariates on the cause-specific hazard:

$$\log(h_k(t|\beta_k,x)) = \log(h_{0k}(t)) + \beta_k x$$

  - $h_{0k}(t)$ is the baseline cause-specific hazard for event type *k.*
  - This model allows for cause-specific estimates of the effects of covariates.
  - If β_k is the same for all *k*, this model reduces to the stratified Cox model.

Cox Model for Cause-Specific Hazard

- The partial likelihood is

$$L(\beta) = \prod_{i=1}^{n} \prod_{k=1}^{K} \left(\frac{e^{\beta_k x_i}}{\sum_{j \in R(t_i)} e^{\beta_k x_j}} \right)^{I(c_i = k)}$$

$$= \prod_{k=1}^{K} \prod_{i=1}^{n} \left(\frac{e^{\beta_k x_i}}{\sum_{j \in R(t_i)} e^{\beta_k x_j}} \right)^{I(c_i = k)} = \prod_{k=1}^{K} L(\beta_k)$$

  - Can maximize L(β) by maximizing each of the $L(\beta_k)$ separately.
  - That is, this model can be fit by fitting K proportional hazards models, where for each model all other event types are treated as censored.

Cox Model for Cause-Specific Hazard

- This model can be used to test whether the cause-specific hazards are constant across strata
  - H_0: $h_{k1}(t) = h_{k2}(t) = \ldots = h_{ks}(t)$ vs H_1: Not H_0

- Can also be used to test whether the effects of covariates on different event types are significantly different.
  - H_0: $\beta_1 = \beta_2 = \ldots = \beta_k$ against the alternative H_1: At least two β_k differ:
  - Fit a Cox model using a combined endpoint
    - C = 1 if any of events 1, 2, … k occurred; C = 0 is censored.

Cox Model for Cause-Specific Hazard

- Compare the likelihood under combined endpoint model to the sum of the likelihoods for the cause-specific models using the partial likelihood ratio test.
  - This method only works if there are no tied event times or if we use Breslow's method for handling ties

Fine and Gray Model for the CIF

- Fine and Gray developed a model that can be used to assess the effects of covariates on the cumulative incidence function.
- Define the subdistribution hazard

$$\tilde{h}_k(t) = \frac{d}{dt}\log(1 - F_k(t))$$

  - $\tilde{h}_k(t)$ is the hazard of the subdistribution function.
  - Note that $\tilde{h}_k(t)$ and $h_k(t)$ are not the same because their risk sets differ.

Fine and Gray Model for the CIF

- Fine and Gray model

$$\log\left(\tilde{h}_k(t|\beta)\right) = \log(\tilde{h}_{k0}(t)) + \beta x$$

Sample Size and Power

- It is important to keep statistical power and sample size in mind when designing a study
  - If the sample size is too large, the study will waste money and other resources
  - If the sample size is too small, the study will be underpowered and a potentially useful treatment may be discarded
    - Wastes the time of participants

Sample Size and Power

- A Type I error occurs when the null hypothesis is rejected when it is actually true.
  - P(Type I error) = α is the significance level of a hypothesis test.
- A Type II error occurs when the null hypothesis is not rejected when it is actually false.
  - 1 - P(Type II error) = 1 - β is the power of a test.

		Truth
Decision	H_0 True	H_1 True
Fail to reject H_0	Correct	Type II Error
Reject H_0	Type I Error	Correct

Power for Survival Analysis

- In survival analysis settings, the power depends on the number of *observed* events, not the overall number of subjects in the study
  - Recall that in Cox proportional hazards model, only subjects who were not censored contribute to the likelihood
- So when determining the sample size for a study when the outcome is time to an event, need to answer two questions:
  - How many events do we need to observe to have the desired amount of power?
  - How many subjects do we need to enroll in order to observe that number of events?

Determining Number of Events Needed

- Simplest, most frequent case: Want to compare the survival time of patients in two groups
  - Want to test H_0: $S_1(t) = S_2(t)$ vs H_1: $S_1(t) \neq S_2(t)$
  - Example: Do patients on the new treatment live longer than patients on the old treatment?

- Can be shown that the log-rank test is equivalent to the score test from a Cox regression model with a single dichotomous covariate
  - That is, can fit a Cox proportional hazards model with only one covariate:

$$\log h(t|\beta, x) = \log h_0(t) + \gamma x$$

    - x = 0 if subject is in Group 1; x = 1 if subject is in Group 2
  - A test of difference in survival times for the two groups is equivalent to testing H_0: $\gamma = 0$ against H_1: $\gamma \neq 0$.

Determining Number of Events Needed

- Schoenfeld's method of estimating power:
  - The number of events needed, *m*, to detect a given hazard ratio (e^{θ}) at significance level α and power (1-β)·100% is

$$m = \frac{\left(z_{\alpha/2} + z_{\beta}\right)^2}{\theta^2 \cdot \pi \cdot (1 - \pi)}$$

    - π is the proportion of the subjects who are in Group 1.
    - Hazard for Group 2 is e^{θ} times the hazard for Group 1
  - Round *m* up to the nearest integer.

Determining Number of Events Needed

- Comments:
  - Schoenfeld's formula is an approximation and may slightly underestimate the required number of events
  - The number of events needed is minimized when subjects are allocated equally to the two groups (i.e., $\pi = 0.5$).
  - The more imbalance there is in the group sizes, the more events we will need to observe to have the same power.

Determining Number of Subjects Needed

- Now, we need to figure out how many subjects we need to enroll in order to observe approximately *m* events.
  - This quantity depends on the probability that an event will occur.
- The number of subjects, n, we need to enroll is given by

$$n = \frac{m}{P(Event\ occurs)}$$

Estimating the Probability of an Event

- The probability that a subject will have an event depends on:
  - The underlying survival function
  - The length of time the subject is at risk

- A subject's survival function depends on group assignment.
  - Let $\bar{S}$(t) be a weighted average of the survival functions for the two groups:

$$\bar{S}(t) = \pi \cdot S_1(t) + (1 - \pi) \cdot S_2(t)$$

  - 1-$\bar{S}$(t) is the probability that an event will occur during t units of time.

Estimating the Probability of the Event

- Most studies will start with an accrual period during which subjects are recruited into the study, followed by a predetermined follow-up period.
  - Let *a* be the length of the accrual period and *f* be the length of the follow-up period.
  - Overall the study will last for *a+f* units of time
- Example: Plan to have a two-year enrollment period and then follow subjects for three more years.
  - Study could begin enrolling subjects on Jan 1, 2017 and continue enrollment through Dec 31, 2018.
  - Subjects followed until Dec 31, 2021, when the study ends.
  - a = 2 years, f = 3 years

Estimating the Probability of the Event

- Note the amount of time each subject is in the study (also called the exposure time) can vary
  - A subject who is recruited on Jan 1, 2017 could be followed for up to five years.
  - A subject who is recruited on Dec 31, 2018 could be followed for up to only three years.
  - Even if these subjects were identical in every other way, the first subject has a higher probability of experiencing the event than the second subject.

Estimating the Probability of an Event

- So *f* is the minimum exposure time for a subject and *a+f* is the maximum exposure time.
  - To keep things simple, assume the rate of recruitment is constant during the accrual period.
  - Under this assumption, the average exposure time is

$$\frac{a}{2} + f$$

Estimating the Probability of an Event

- We use the accrual period, follow-up period and overall survival function to estimate the probability that an event occurs

$$P(Event\ occurs) \approx 1 - \frac{1}{6}\left[\bar{S}(f) + 4 \cdot \bar{S}\left(\frac{a}{2} + f\right) + \bar{S}(a + f)\right]$$

Power for Survival Analysis

- So in order to determine the sample size, we need:
  - Need to know the length of the accrual and follow-up periods and how subjects will be allocated to groups (π)
  - The significance level, α, and the desired power, $100 \cdot (1-\beta)\%$
    - Usually $\alpha = 0.05$ or 0.01 and power is 80% or 90%
  - The effect size or hazard ratio (HR) we want to be able to detect
  - Estimates of the survival probabilities at the minimum, maximum and average exposure times

Power for Survival Analysis

- In the cases where we are comparing a new treatment to an old treatment, usually have a good idea of what survival looks like with the old treatment
  - Use previous studies to estimate $\hat{S}_1(t)$
  - Our estimate of $S_2(t)$ is $\hat{S}_2(t) = \left(\hat{S}_1(t)\right)^{e^{\theta}}$
- What if we don't have estimates of the survival probability at minimum, maximum, and average exposure times?
  - Often only have estimate of survival at one time point
    - Example: You know the one-year survival rate is 46%
  - Common solution is to use an exponential model to get remaining survival probabilities