Hypertrophic Cardiomyopathy and Competing Risks Models

Gabriel Demuth

February 24, 2017
What is this talk about?

- A brief summary of my internship at Mayo Clinic last semester
- An introduction to some basic survival analysis
  - With a particular focus on competing risks for medical data.
Hypertrophic Cardiomyopathy (HCM)

- Thickening of the heart muscle without obvious cause.
- Effects people of any gender, age, or fitness level.
- Is often undiagnosed, many people never experience symptoms.
- Some patients develop chest pain, shortness of breath, etc.
A very small number of HCM patients experience Sudden Cardiac Death (SCD), which occurs when the victim’s heart spontaneously begins a fatal arrhythmia.

- SCD is often associated with periods of intense physical exertion.
- It is difficult to determine which HCM patients are at risk of SCD.
- Preventative treatments for SCD are available.
  - Surgical operations to reduce heart muscle thickness (septal myectomy)
  - Implantable devices
The best available treatment for SCD is an Implantable Cardioverter Defibrillator (ICD), which can shock a patient’s heart out of arrhythmia.

- ICDs occasionally have false positive discharges. Patients describe this as like getting kicked in the chest by a horse.
- ICDs also may become infected, and need to be removed.

Giving most/all HCM patients ICDs is not viable. There is therefore considerable interest in correctly identifying the patients who are at greatest risk for SCD.
The goal of the project was to develop a real-time tool for doctors to assess SCD risk in HCM patients.

- Software would automatically look up the relevant information on a patient.
- Use an existing model to calculate an estimate of the probability of that patient experiencing SCD in the next 5 years.
- This would be displayed to the doctor while they are meeting with the patient as part of their EMR.
- Currently doctors go to an external website to calculate risk using this model.

http://www.doc2do.com/hcm/webHCM.html

My portion of this project was to develop a better/alternative risk model for SCD.

- Even if my model isn’t better, it can be integrated into the project to demonstrate model A/B testing.
The best existing model is one developed by the European Cardiac Society

- Cox proportional hazards model.
- Based on 7 covariates, includes one quadratic effect.
- Data from 6 European cardiac centers:
  - 3675 patients
  - 198 cases of SCD/SCD equivalent events recorded
- Used chained equations to impute missing values
Internal dataset collected at Mayo.
  ▶ 1945 patients in total
  ▶ 26 cases of SCD

Data was undocumented, with very high rates of missingness or nonsense values.
  ▶ I’m not going to outperform the European model with this...

But I can develop a better *theoretical* model
HCM-SCD data is time-to-event data, in this case time-to-SCD. It is simplest to represent this as a stochastic process. For patient $i$, let

$$X_i(t) = \begin{cases} 
0 & \text{if } i \text{ has not experienced SCD at } t \\
1 & \text{if } i \text{ has experienced SCD at } t 
\end{cases}$$

Time to event is then

$$T = \inf \{ t \mid X_i(t) \neq 0 \}$$

And we observe a sequence of event times $t^{(1)} < t^{(2)} < \cdots < t^{(n)}$.
Survival data is often subject to right censoring and left truncation.

- Right censoring is when, at the end of the study period, an event has not happened for a study subject.
- Left Truncation occurs when a patient must survive for some period of time before they can first be observed. Some patients may die of SCD before they are diagnosed with HCM.

The data I used is both left truncated and right censored. Standard survival methods handle these gracefully, so I won’t spend much time on either. Can define

$$C_i = \begin{cases} 
0 & \text{if } i \text{ has not experienced SCD by end of study} \\
1 & \text{if } i \text{ has experienced SCD by end of study}
\end{cases}$$

- European model used age as a covariate instead of a truncation time.
Risk modeling is usually based on the hazard function, which is defined as:

\[ \alpha_t = \lim_{\Delta t \downarrow 0} \frac{P(X_{t+\Delta t} = 1 \mid X_t = 0)}{\Delta t} \]

Which is the instantaneous change in the conditional probability of an event occurring in \( \Delta t \), given it has not happened up until time \( t \). The cumulative hazard is defined as

\[ A(t) = \int_0^t \alpha(u) \, du \]

Note that this is not a probability, but it is directly estimable from the observed data.
Recall that the goal of the project is to develop a model that can determine which patients are at a high risk of SCD, so we need to incorporate covariates.

- A common way to do this is through a Cox proportional hazards model. This assumes a parametric model for covariate effects, and an (unspecified) non-parametric baseline hazard $\alpha(t)$.

For a covariate $Z_i$, the Cox model assumes that

$$
\alpha(t, Z_i) = e^{\beta Z_i} \alpha(t)
$$

$$
\Rightarrow A(t, Z_i) = e^{\beta Z_i} A(t)
$$

Note that when $Z_i = 0$ (baseline level), the hazard is simply $\alpha(t)$. The $\beta$ parameters can be estimated by conditional likelihood, without specifying a form for $\alpha(t)$.

- Estimates of cumulative hazard at baseline or any covariate level may be obtained.
In this particular application, we are mostly interested in $P(\text{SCD in 5 years})$, not in the hazard itself. However the probability can be recovered from the cumulative hazard:

$$P(T \leq t) = 1 - \exp \left( - \int_0^t \alpha(u) \, du \right)$$

$$= 1 - \exp (-A(t))$$
Estimation for Proportional Hazards

Estimation is done using a partial likelihood.

- We observe a sequence of event time $t_1 < t_2 < \cdots < t_n$.
- For event time $t_l$, let $R(t_l)$ be the set of patients who are at risk at that time.
  - $R(t_l)$ contains all patients who have yet to experience an event, or are censored prior to $t_l$.
- Given an event occurs at time $t_l$, the conditional probability that individual $i$ experiences the event is
  $$
  \frac{\alpha_0(t_l) \exp(\beta Z_i)}{\sum_{k \in R(t_l)} \alpha_0(t_l) \exp(\beta Z_k)} = \frac{\exp(\beta Z_i)}{\sum_{k \in R(t_l)} \exp(\beta Z_k)}
  $$
It is convenient to define an at-risk function \( Y_i(t) \), where
\[
Y_i(t) = \begin{cases} 
0 & \text{if } i \text{ is not at risk} \\
1 & \text{if } i \text{ is at risk} 
\end{cases}
\]

Then the full likelihood may be written as
\[
L(\beta) = \prod_{t=t(1)}^{t(n)} \prod_{i=1}^{n} \left\{ \frac{\exp(\beta Z_i) Y_i(t)}{\sum_{k=1}^{N} \exp(\beta Z_k) Y_k(t)} \right\}^{I_{\text{event at time } t}}
\]

Which may then be maximized in \( \beta \).

- Truncation makes this very slightly more complicated.
- Methods exist for handling ties in event times.
Competing Risks

Note that the above model does not account for other causes of death.

Suppose a patient dies of a stroke.

- We could consider them censored at time of death $C_i = 1$, but this implicitly assumes that they could still experience SCD.

- This assumes that, if you were to wait long enough, every single HCM patient would eventually experience SCD.

Which are obviously not correct in this setting. The correct approach is to consider a *Competing Risks Model*, which allows for other causes of death.
Competing Risks State Diagram

HCM Patient
(j = 0)

Death by SCD
(j = 1)

Other CoD
(j = 2)
Let $X_i(t) \in \{j = 0, 1, 2\}$ give the state of patient $i$ at time $t$, as before, where $j = 0$ indicates that the patient has yet to experience an event, $j = 1$ denotes SCD, and $j = 2$ denotes death by any other cause. $C_i$, the censoring indicator, is as before. We can then model cause-specific hazards for each endpoint. Again using a proportional hazards assumption

$$\alpha_1(t, Z_i) = e^{\beta_1 Z_i} \alpha_1(t)$$

$$\alpha_2(t, Z_i) = e^{\beta_2 Z_i} \alpha_2(t)$$

Note that $Z_i$ may have a different effect on each of the transition hazards, via transition specific $\beta_j$ parameters.
This is essentially the same as before, except that we include both endpoints in the likelihood:

$$L(\beta) = \prod_{t=t(1)}^{t(n)} \prod_{i=1}^{N} \prod_{j=1}^{2} \left\{ \frac{\exp(\beta_j Z_i) Y_i(t)}{\sum_{k=1}^{N} \exp(\beta_j Z_k) Y_k(t)} \right\} I_{\text{event of type } j \text{ at time } t}$$

- Likelihood is maximized separately for $\beta_1$, $\beta_2$ parameters.
- A patient who experiences event 1 is treated as censored in the event 2 likelihood and vice versa.
- Estimates are therefore the same as the single endpoint model
- The difference is in how event probabilities are calculated.
As before, we can calculate the probability of an event from the cumulative hazards.

- But since we have two endpoints, we need to be precise about what event we want to estimate.

Some possible events:

1. \( P(T \leq t) \), the probability that the patient experiences an event by time \( t \)

2. \( P(T \leq t, X_i(T) = j) \), the probability that the patient experiences event \( j \) by time \( t \).
Calculating Event Probabilities

Let

\[ \alpha_0(t) = \alpha_1(t) + \alpha_2(t) \]

\[ A_0(t) = \int_0^t \alpha_0(u)du = A_1(t) + A_2(t) \]

Note that \( \alpha_0(t) \) is the hazard of any event at time \( t \), and \( A_0(t) \) the cumulative any-event hazard. Then

\[ P(T \leq t) = 1 - \exp(-A_0(t)) \] (1)

And

\[ P(T \leq t, X_i(T) = j) = \int_0^t P(T > u)\alpha_j(u)du \]

\[ = \int_0^t \exp\left(-\int_0^v \alpha_1(v) + \alpha_2(v)dv\right)\alpha_j(u)du \] (2)

So the probability of a patient suffering SCD depends on both hazards.
The European Cardiac Society study uses a proportional hazards model, but does not account for competing risks. What are the effects of this?

- Estimation of covariate effects is unbiased.
- Estimation of the SCD hazard is unbiased.
- Estimates of the probability of SCD are biased.
  - Specifically, will tend to be too high, because the single-endpoint model does not account for the $-\alpha_2(t)$ term in the exponent.
Recall that my real data is very bad, so these numbers are not medically relevant.

- But they can give a sense of the importance of accounting for competing risks.

I fit a proportional hazards model, using a single covariate:

- Covariate is Syncope (fainting), which is an indicator variable.
- Fit both single endpoint and competing risks models.
  - Competing risks model uses separate baseline hazards for SCD, Other CoD
  - Separate covariate effects on each hazard
### Model Results

#### Single Endpoint Version:

<table>
<thead>
<tr>
<th></th>
<th>coef</th>
<th>exp(coef)</th>
<th>se(coef)</th>
<th>z</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Syncope</td>
<td>0.895</td>
<td>2.448</td>
<td>0.466</td>
<td>1.92</td>
<td>0.055</td>
</tr>
</tbody>
</table>

#### Competing Risks Version:

<table>
<thead>
<tr>
<th></th>
<th>coef</th>
<th>exp(coef)</th>
<th>se(coef)</th>
<th>z</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Syncope.SCD</td>
<td>0.895</td>
<td>2.448</td>
<td>0.466</td>
<td>1.92</td>
<td>0.055</td>
</tr>
<tr>
<td>Syncope.Other</td>
<td>0.117</td>
<td>1.124</td>
<td>0.160</td>
<td>0.73</td>
<td>0.466</td>
</tr>
</tbody>
</table>

Note that the estimates of the effect of syncope on SCD are the same between models. Estimation is done independently for each endpoint.
Competing Risks: Cumulative Hazard Estimates

Proportional Hazards Estimates

- SCD, no Syncope
- SCD, Syncope
- Other, No Syncope
- Other, Syncope

Age vs. Cumulative Hazard

Cumulative Hazard
Competing Risks: Cause-Specific Event Probability Estimates

The image shows a competing risks analysis with a subdistribution function graph. The graph compares the cumulative incidence of events across different categories:
- SCD, no Syncope
- SCD, Syncope
- Other, no Syncope
- Other, Syncope

The x-axis represents time, ranging from 20 to 100, while the y-axis represents the subdistribution function, ranging from 0.0 to 1.0.
Estimated Survival Function

![Survival Function Graph]

- Survival Probability on the y-axis.
- Age on the x-axis.
- Two lines represent different conditions:
  - Solid line: No Syncope
  - Dashed line: Syncope

The graph illustrates the estimated survival function over age for individuals with and without syncope.
Conclusions

▶ Data was a mess.
▶ I learned a lot about survival analysis, which was not an area of statistics I had previously studied.
▶ I got to be a part of a very cool project.
▶ And found a significant shortcoming in the best model available.
▶ So even though my final model was not usable, it still contributed something of value to the team.
Citations