Chapter 18

Estimating the Hazard Ratio

What is the hazard?

The hazard, or the hazard rate, is a rate-based measure of chance. Formal notation aside, the hazard at time t is defined as the limit of the following expression, when Δt tends to zero:

$$\frac{\text{Probability of an event in the interval } [t, t+\Delta t)}{\Delta t}$$

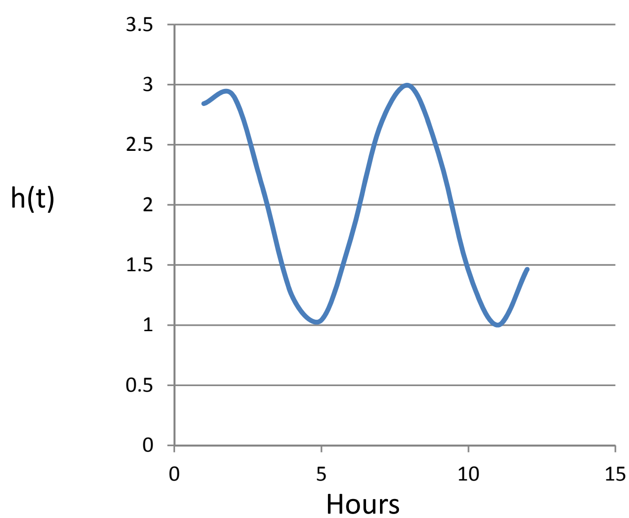
Writing the numerator as the ratio of the count of events (c) to the count of "at risk" (N), we can see that the expression above is indeed a rate — the number of events per unit of time-at-risk:

$$\frac{c \,/\, N}{\Delta t} = \frac{c}{N \,\Delta t}$$

Being the limit of the rate at Δt=0, the hazard may be viewed as the *instantaneous* rate at a time point. That is, the chance of something happening *at* a time, rather than *between* two times.

Since the hazard is defined at every time point, we may bring up the idea of a hazard function, $h(t)$ — the hazard rate as a function of time. This function is a theoretical idea (we cannot calculate an instantaneous rate), but it fits well with causal reality under the axiom of indeterminism. Anyone who felt, for example, risky and safe conditions while driving a car can imagine a hazard function with peaks and valleys at different moments. Figure 1 shows an example of what someone's hazard-of-death function might look like during some period (1AM till noon). The hazard at each moment is determined by the values that were taken by the causes of death at baseline.

Figure 1. Hypothetical hazard-of-death function



Cox regression

Cox regression is a regression model that enables us to estimate the hazard ratio (hazard rate ratio) — a measure of effect which may be computed whenever the time at risk is known. The model is named after the statistician who wrote the regression equation and proposed a method to solve it (to estimate the coefficients). For a reason that will be explained later, the model is also called "proportional hazards regression". Cox regression is shown next vis-à-vis three common regression models: linear, logistic, and Poisson.

Linear regression:	mean Y	$= \beta_0 + \beta_1 E$
Logistic regression:	log (odds)	$= \beta_0 + \beta_1 E$
Poisson regression:	log (rate)	$= \beta_0 + \beta_1 E$
Cox regression:	$\log h(t)$	$= \log h_0(t) + \beta_1 E$

A little algebra shows that the last equation may also be written as

$$h(t) = h_0(t) \times \exp(\beta_1 E)$$

The way to interpret the exposure coefficient, β_1, in Cox regression is similar to the way you interpret the exposure coefficient in any log model. It is the difference between the log-hazard per one unit increment in E, which is equivalent to the log of the hazard ratio:

$$\beta_1 = \log \text{(hazard ratio)}$$

Exponentiate the coefficient and you get the hazard ratio:

$$\text{hazard ratio} = \exp(\beta_1)$$

We observe, however, a key difference between Cox regression and other regression models. Instead of the usual intercept, β_0, we find a bizarre expression, $\log h_0(t)$, which looks like a time-varying intercept. Why is it there? What does it mean?

The first question is easy to answer. It is there because the dependent variable is a function of time. We cannot simply write "$\log h(t) = \beta_0 + \beta_1 E$" as before. How can the dependent variable be a function of time, when time (t) is not included among the input variables? Some expression of time must appear on the right hand side of the equation.

As for the meaning of $\log h_0(t)$, it is not different from the meaning of any classic intercept: $\log h_0(t)$ takes the values of the dependent variable, $\log h(t)$, when E=0; or more generally, when all the independent variables take the value of zero. (That's the reason for the subscript "0".) Unfortunately, $\log h_0(t)$ is often called "the baseline hazard", a confusing term because "baseline" usually denotes the time at which follow up begins, not a zero value of variables. Moreover, when the zero value of one independent variable is meaningless (e.g., weight=0), the so-called baseline hazard is not quantifying any theoretical hazard. It is meaningless.

Why is Cox regression also called "proportional hazards regression"?

Since the hazard is a function of time, the hazard ratio, say, for exposed versus unexposed, is also a function of time; it may be different at different times of follow up. For example, if the exposure is some surgery (vs. no surgery), the hazard ratio of death may take values as follows:

Time since baseline	Hazard ratio
1 day	9
2 days	3.5
28 days	3.5
…	…
…	…
365 days	0.8

Cox regression, however, allows for only *one* hazard ratio, which is $\exp(\beta_1)$. The hazard ratio of death for surgery vs. no surgery is assumed to be the same *at any time since baseline*. The model may therefore be called "a constant hazard ratio model", but someone thought that "proportional" is a better word to describe a fixed ratio of two hazards over time. (When the ratio of two quantities is fixed, we may say that one quantity is proportional to the other, say, 1.5 times the other.)

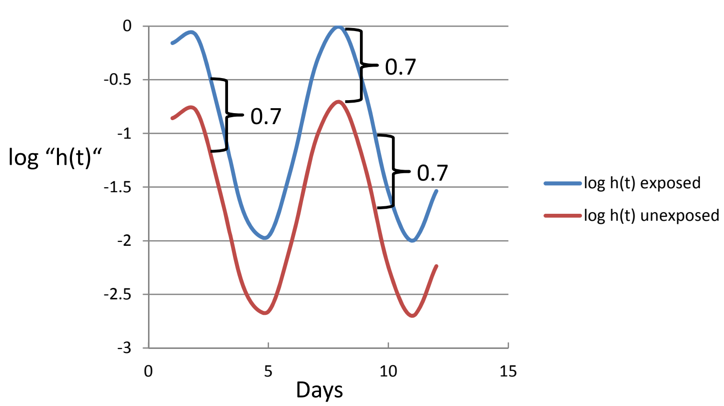
To get a visual impression of the proportional hazards feature, let's assume that E is a binary (0,1) exposure. Plugging in the value of E, we first derive two log-hazard functions:

For exposed (E=1): $\log h(t) = \log h_0(t) + \beta_1$

For unexposed (E=0): $\log h(t) = \log h_0(t)$

Not knowing the values of $\log h_0(t)$, we have no idea how to draw either function. But we do know that the two functions progress in the same direction, and that the distance between them at any point is β_1 — the difference in the log-hazard between exposed and unexposed (which is also the log of the hazard ratio). Figure 2 shows a hypothetical example where $\beta_1 = 0.7$. Note that the Y-axis is not truly a log-hazard, because we don't know the actual location of the functions on the Y-axis. We don't know the true value of the (log) hazard.

Figure 2. Two log-hazard functions which are 0.7 log-hazard units apart



Switching now from log-hazard to hazard, we derive the corresponding hazard functions:

For exposed (E=1): $h(t) = \exp(\log h_0(t) + \beta_1) = \exp(\log h_0(t)) \times \exp(\beta_1) = h_0(t) \times \exp(\beta_1)$

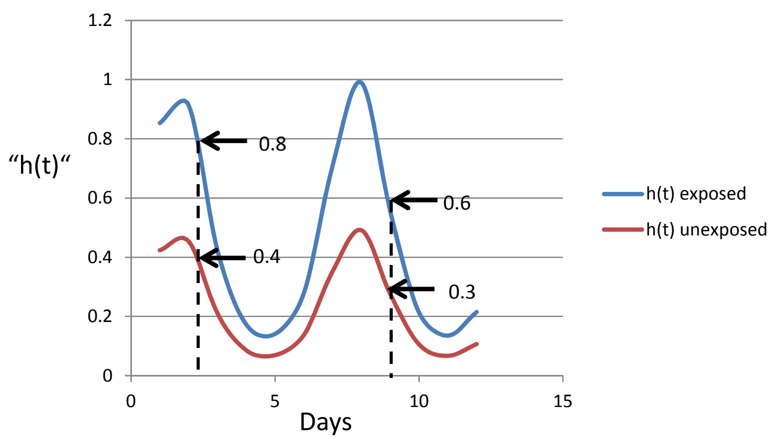
For unexposed (E=0): $h(t) = \exp(\log h_0(t)) = h_0(t)$

It is easy to see that at each time point the ratio of the hazard for exposed to the hazard for unexposed — the hazard ratio — is equal to $\exp(\beta_1)$, a constant:

$$h(t) \text{ in exposed } / \ h(t) \text{ in unexposed} = h_0(t) \times \exp(\beta_1) \ / \ h_0(t) = \exp(\beta_1)$$

Figure 3 shows the respective hazard functions for the log-hazard functions that were depicted in Figure 2 ($\beta_1 = 0.7$). At each time point the value of $h(t)$ for exposed is twice the value for unexposed: $\exp(0.7) \approx 2$. A constant difference of 0.7 between log-hazard functions (Figure 2) is equivalent to a constant ratio of about 2 between hazard functions (Figure 3). Notice that Figure 3 would have been identical to Figure 2 if the Y-axis were logarithmic.

Figure 3. Two hazard functions where the hazard for exposed is about twice the hazard for unexposed (hazard ratio ≈ 2)



Cox partial likelihood function

A regression model is useless without a method to estimate the coefficient of E, or more generally, the coefficients of all the independent variables. Similar to other regression models, the estimation in Cox regression requires two steps:

1) Construct a likelihood function (with the coefficients on the independent side): Likelihood=$f(\beta_1,\beta_2,\beta_3,\ldots)$

2) Find the maximum likelihood estimates — the values of the coefficients that maximize the value of the likelihood.

Here, however, we encounter a problem. Unlike other types of regression, the right hand side of Cox regression includes not only coefficients, but also a function of time, $\log h_0(t)$. How can we estimate that time-varying intercept? Don't we have to assume something about the shape of the so-called baseline hazard, the hazard function when all the independent variables take the value of zero?

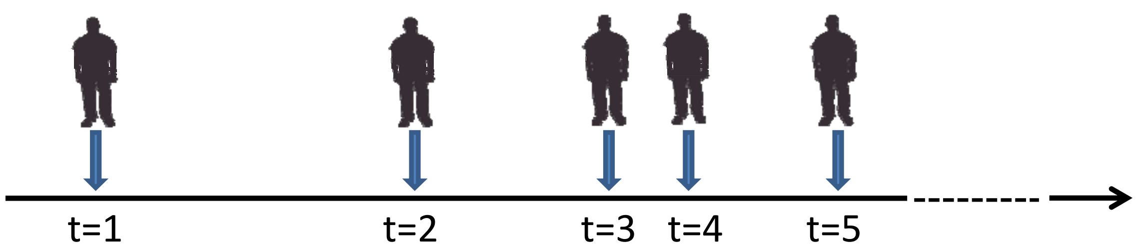
Fortunately, we can do without log $h_0(t)$ — even if it happens to be meaningful. Just as we didn't need the intercept, β_0, to estimate the effect of E from linear, logistic, or Poisson regression, we don't need log $h_0(t)$ to estimate the effect of E from Cox regression. As far as effect estimation is concerned, the intercept is always a nuisance term.

Realizing the last point, Cox suggested a radical idea back in the 1970s. He proposed to estimate the coefficient(s) using a *partial* likelihood function which does not include log $h_0(t)$. If you like analogies, it is similar to estimating the coefficient of E in logistic regression, without estimating the intercept. (In fact, that's exactly what we do when we fit a conditional logistic regression model to data from an individually matched case-control study.)

According to a circulated gossip, Cox's solution of the regression equation was belittled by many when it was presented for the first time at a statistics conference. Those who belittled his idea are probably still hiding somewhere, if they are still around, because partial likelihood has become a standard tool in statistics, and Cox's seminal paper on this topic is counted among the most cited papers in science. I suspect that Cox's critics at that time have learned the lesson that many arrogant minds haven't learned yet: It is the duty of the scholar to try to tear apart an idea on substantive arguments, but it is foolish to dismiss an idea because "it doesn't sound right to my brilliant mind".

Back to partial likelihood. A likelihood function tells us something about the likelihood of the observed data as a function of the coefficients. Here, part of the observed data is a sequence of events during some follow-up time. Figure 4 shows a hypothetical example.

Figure 4. The first five events in a cohort study, or a trial



Assuming independent events, the likelihood of observing n events is the product of the likelihood of observing each event. But what is that single-event quantity? Simple hand-waving (and some math) suggests that the likelihood of an event that was observed at time t is given by the following proportion of hazards:

$$\frac{h(t)\text{ for the person who had the event}}{\text{Sum of }h(t)\text{ for all those who were at risk at that time}}$$

To construct the partial likelihood function (L_p), write the product of the likelihood of observed events ($L_p = L_{t=1} \times L_{t=2} \times \ldots$), substituting the expression above for each event.[a] Then, replace each $h(t)$ with the right hand side of Cox regression [in our case, with $h_0(t) \times \exp(\beta_1 E)$]. Finally, plug in the values of the independent variables (in our case, the value of E), and you got the function — the partial likelihood as a function of the coefficients. The maximum partial likelihood estimates of the coefficients may be found by some trial-and-error algorithm.

What happened, though, to the time-dependent intercept, log $h_0(t)$? Does it appear in the likelihood function?

No, it does not. To see why not, let's derive the likelihood of the first event in Figure 4. We'll assume that the person in the figure was exposed, and that 100 people were at risk at that time, 30 of whom were exposed and 70 were not.

Using the alternative expression of Cox regression, $h(t) = h_0(t) \times \exp(\beta_1 E)$, we first derive the hazard at the first time point, $h(t=1)$, for those 100 people at risk:

For every exposed person (E=1): $h(t=1) = h_0(t=1) \times \exp(\beta_1)$

For every unexposed person (E=0): $h(t=1) = h_0(t=1)$

The likelihood of the first event is the hazard for the exposed person (to whom it happened) divided by the sum of the hazard for 100 people: 30 exposed and 70 unexposed. In notation:

$$\text{Likelihood of event 1} = \frac{h_0(t=1) \times \exp(\beta_1)}{30 \times h_0(t=1) \times \exp(\beta_1) + 70 \times h_0(t=1)} = \frac{\exp(\beta_1)}{30 \exp(\beta_1) + 70}$$

As you see above, $h_0(t)$ is cancelled in the likelihood term for the first event (and for any event). Therefore, the partial likelihood is a function of the coefficient(s) alone. It is neither a function of follow-up time nor a function of the time-at-risk.

Time-at-risk is needed only to identify the "risk set", the set of people who were at risk at the time of each event. The actual event time does not matter. For instance, as long as the risk set at t=1 comprised 30 exposed and 70 unexposed, that time point could be one day, or three weeks, or 14 months since baseline. Likewise, if the risk set at t=2 comprised 90 people, say, evenly split between exposed and unexposed, it does not matter whether the second event happened two days or 15 months after the first event. All that matters is who had the event and who was at risk at each event time. When these parameters are fixed, the *spacing* makes no difference.

The partial likelihood, as constructed above, does not allow for coinciding events (called "ties"), but there are statistical methods to handle the problem. If ties are uncommon, you can solve the problem by adding a trivial error: change a date. For example, if time is counted in days and two

[a] In many texts, the likelihood of an event is called "probability". There is a subtle point here which is usually ignored. Since time is continuous, event probabilities form a probability density function, which means that the probability of an event *at any time point* must be zero. In practice, however, time is treated as a discrete variable (hours, days), so the computed probability is not truly a time point probability. For example, when follow-up is counted in days, the probability of an event on a given date means the probability of it happening during a 24-hour interval. All of this surely sounds "a little different" from the theoretical proportion of hazards (which are *instantaneous* rates).

events happened on day 178, change the date of one event to be day 177 or day 179. Those who object to this inelegant solution should think about the following point: If the results are sensitive to such trivial alteration of data, the problem of ties must be trivial as compared with the bigger data problem we have (perhaps short follow-up, or sparse data). Furthermore, you can change the date both ahead and backward to see if the results are similar.

In classic Cox regression, people who already had the event are excluded from the risk set, just like the exclusion of prevalent disease at baseline. Therefore, the hazard and the hazard ratio are "conditional" measures. For example, the hazard at t=2 is conditional on not having the event before $t=2$. For reasons that are beyond the scope of this text, conditioning may not be a good idea in both cases.

Lastly, we may now understand why the likelihood is called "partial". The "full" likelihood should take into account not only observed events, but also observed "non-events". The latter are ignored when the likelihood function is constructed. For example, we did not consider the 99 likelihoods for 99 people who remained event-free at t=1.

On the proportional hazards assumption

I explained earlier why Cox regression is called "proportional hazards regression". It is time to explain why this descriptor is misleading, if not a misnomer. Cox regression doesn't have to be a "proportional hazards regression" at all. If you want to allow the hazard ratio to be different at each time point, simply fit the following model:

$$\log h(t) = \log h_0(t) + \beta_1 \text{E} + \beta_2 \text{E}t$$

where Et is not the name of a movie, but the product "exposure x follow-up time". In this model the hazard ratio is no longer a constant. It is a function of time: HR = exp ($\beta_1 + \beta_2 t$). In fact, you have just invoked the "non-proportional hazards assumption" in Cox regression!

Don't want to allow the hazard ratio to vary so much? That's easy. Categorize the follow-up time (two intervals, three intervals, any k intervals); replace k intervals by k-1 dummy variables; and fit a similar model with k-1 product terms. Now the hazard ratio is forced to be constant only within each interval.

But the issue is much deeper than fitting different models. Reading scientific literature, you get the impression that scientists are extremely worried about possible violation of the proportional hazards assumption. Actually, some of them seem to be obsessed with it, which is funny from one point of view and serious from another.

It is funny because the same scientists regularly impose a comparable assumption without blinking an eye. Consider, for example, the following multi-variable logistic regression model where E is the exposure and Q, R, S, and T are covariates for conditioning:

$$\text{log odds (D=1)} = \beta_0 + \beta_1 \text{E} + \beta_2 \text{Q} + \beta_3 \text{R} + \beta_4 \text{S} + \beta_5 \text{T}$$

Analogous to "the proportional hazards model", this model may be called "the proportional odds model". Instead of imposing proportionality of the hazard over time points, the model imposes proportionality of the odds over covariates' values. The ratio of the disease odds in exposed to the disease odds in unexposed (the odds ratio) is assumed to be identical for any value of Q, for any value of R, for any value of S, and for any value of T. Yet I have read hundreds of papers in which such a model was fit — without calling it "the proportional odds model" and without worrying about possible violation of "the proportional odds assumption". The same scientists who pay careful attention to a proportionality assumption in one model (Cox) regularly ignore it in other models (logistic, Poisson, log-probability). How come? I will try to answer this question later.

The model above is called a main effects model. This model, and similar log models, claim that none of the covariates modifies the exposure effect on the disease, which amounts to proportionality of the odds (or the rate, or the probability) across the values of the covariates. Is there a comparable idea for time? Does Cox regression, without time-containing product term(s), claim no effect modification by time? Here we get into a serious, frequently overlooked, issue.

First, time is not a modifier of any effect because a modifier must be a causal variable, and time causes nothing. Any time variable (age, period, birth year) that is associated with an outcome merely substitutes for an unknown list of causal variables. If interested, you may read more on this topic in my commentary on period and cohort effects (posted on my website).

Second, according to an axiom of causality, all effects operate between a time point exposure and a time point outcome, which implies that a causal parameter might depend on the time interval between the two variables. For instance, the effect of some surgery (vs. no surgery) on death might be different *at* 24 hours post-surgery, *at* 157 hours, and *at* 8760 hours (three years post-surgery). If so, the so-called effect *over* a time interval, say, *by* three years since surgery, is not truly a causal parameter. It is some kind of an average of unknown true effect sizes at different time points. To use a metaphor, the so-called effect of surgery on death *by* three years may be as informative as the average price of some stock between 2007 and 2009.

From this perspective, a model with a constant hazard ratio is equivalent to a naïve theory — not an assumption — that the effect of a time point exposure on a time point outcome is identical for different intervals between the two variables. This theory may be explored and challenged not only in Cox regression but also in other models, provided that follow-up data are available. For instance, we may fit logistic regression models to trial data on surgery and death, truncating the end-date at different times (e.g., 24 hours since baseline, 157 hours since baseline). The estimated odds ratios for different length intervals may tell us something about the truth of the "same effect" theory.

How often do you see scientists fit such a series of logistic regression models, or even entertain them? Rarely. How often do you see scientists address the very same issue in Cox regression? Often. What is the explanation for that discordant behavior? One author proposed that it's a matter of linguistics and psychology. Since the words "proportional hazards" often show up in the name of the model — Cox *proportional hazards* regression — scientists and statisticians feel compelled to address "the proportional hazards assumption". If so, the solution is simple: take these words out. Call it "Cox regression", which is both shorter and more accurate. (Cox is credited not only with the regression equation, but also with its solution.)

In my view, however, the explanation is deeper than word choice and psychology. We observe here a common disconnect between statistical ideas as regularly taught by statisticians, and causal ideas, which are rarely taught to statisticians and scientists. We observe here what may be called "mechanical use of statistics", an ailment of modern science. To demonstrate my point, consider a typical "let me reassure you" statement that many authors write after running Cox regression:

"The proportional hazards assumption was tested by adding interaction terms with time. The coefficients of these terms were not statistically significant ($p>0.05$)."

Three components of this statement indicate superficial understanding of both science and statistics. First, as you already understand, "proportional hazards" is not an assumption but a (naïve) causal theory which claims that the effect of a time point exposure on some outcome is identical at future time points. Second, whatever the null hypothesis states (no effect or no interaction), rejection of the null adds insignificant knowledge, because the complementary of the null is "everything but null" — essentially a useless piece of knowledge. Third, the lack of statistical significance ($p>0.05$) provides evidence for only one thing: that testing of the null was a waste of time. Large *p*-values provide no reassurance that the hazard ratio is indeed constant, because the lack of evidence against the null is not evidence for the null. Try to memorize the last sentence, which too many try to forget.

Does the last paragraph sound wrong to you? Do you find it hard to believe that it's all true? If so, ask your teachers of statistics to write a rebuttal. Chances are they wouldn't. And please don't settle for spoken words. They evaporate as soon as they leave the mouth.