Competing Risks

The textbook has a good discussion of multiple decrement life tables and we reproduce online the calculations in Boxes 4.1 and 4.2. Here is just a quick summary of some of the main ideas.

Continuous Time Formulation

We assume that there are J causes of failure and that every failure can be attributed to one and only one of these causes. We define a cause-specific hazard or force of mortality

$$\mu_j(x) = \lim_{dx \to 0} \frac{1}{dx} \Pr\{X \in (x, x+dx) | X > x \text{ and } J = j\}$$

As the limit of the probability of death due to cause i in a small interval after x given survival to x.

Using the law of total probability, the total force of mortality is a simple sum

$$\mu(x) = \sum_j \mu_j(x)$$

From the hazard we can calculate overall survival the usual way. We can also compute the probability of dying by age x due to cause j, also known as the *cumulative incidence* function

$$I_j(x) = \Pr\{X \leq x \text{ and } J = j\} = \frac{1}{l(0)} \int_0^x l(a) \mu_j(a) da$$

The integrand reflects the probability of surviving *all* causes up to age a times the conditional probability of then dying due to cause j. All of these functions reflect what happens when all causes are competing or acting at the same time.

One may also calculate a "survival function" based on cause j alone, or based on all causes other than j, but these are counterfactuals predicated on the assumption of independence of the underlying risks. Unfortunately this assumption cannot be verified empirically.

The Multiple Decrement Life Table

The additional data are simply counts of deaths by age due to cause j, say ${}_nD_x^j$, which add up to all deaths at that age. Let ${}_nR_x^j = {}_nD_x^j / {}_nD_x$ denote the proportion of deaths at ages x to $x+n$ due to cause j. The exposure remains the same, as everybody is exposed to all causes at any given time. We then calculate a life table using *all* causes of death as usual, but then add a few columns:

$${}_nq_x^j = {}_nq_x \; {}_nR_x^j$$

is the conditional probability of death at ages $x, x+n$ due to cause j, and

$$_nd_x^j = l_x\ _nq_x^j = \ _nd_x\ \ _nR_x^j$$

Is the number of deaths at ages $x, x+n$ due to cause j. Accumulating counts of deaths and dividing by l_0 estimates the incidence function, or cumulative probability of death due to cause j. The textbook also defines (in my opinion in a slight abuse of notation) l_x^j as the number of people age x who will eventually leave the life table due to cause j, obtained by accumulating $_nd_x^j$ from age x onwards.

The example in Box 4.1 distinguishes female deaths due to neoplasms and all other causes and estimates that 21.2% of female births will die of neoplasm if 1991 rates were to prevail.

Associated Single-Decrement Life Tables

Associated with cause j is a cause-specific force of mortality $\mu_j(x)$. The associated single-decrement life table attempts to estimate what survival would look like if this was the only cause operating. The honest answer is that we don't know, the question is a *counterfactual* that requires the strong assumption that eliminating the other causes would leave $\mu_j(x)$ unchanged. This is equivalent to assuming independence of the underlying risks.

Then there is a technical problem concerning how to construct the life table. The textbook discusses three approaches.

1. We can calculate a cause-specific rate $_nM_x^j$ dividing deaths dues to cause j by total exposure, equate that to the life table rate $_nm_x^j$ and proceed "as usual", making assumptions about $_na_x$, which may of course be different from the assumptions made for the overall table.

2. Assume that cause-specific risks are constant within an age group and convert the rates to probabilities accordingly. This is my preferred approach. The textbook notes that is "logically consistent and easy to apply" and preferable when the assumption is tenable, which of course is only true for small age intervals.

3. Follow Chiang in assuming that the cause-specific force of mortality is proportional to the overall force of mortality in the age interval x to $x+n$, so $\mu_j(a) = R_j\ \mu(a)$ when $a \in (x, x+n)$. The proportionality factor is estimated using $_nR_x^j$, the proportion of all deaths in the age interval that are due to cause j. This is a proportional hazards assumption and leads to estimating the probability of surviving the interval when only cause j is operating as

$$_np_x^j = \ _np_x{}^{_nR_x^j}$$

From these survival probabilities we can get the survival function as well as the conditional probabilities of death. Calculating time lived, which we need to get expectation of life, requires assumptions about $_na_x$. A common solution is to keep the values in the full life table, but the textbook notes that if only one cause is operating the age distribution in the interval will be younger, and recommends a graduation approach, using formulas 4.6 and 4.8. The former is based on fitting a quadratic to age at death over three adjacent intervals. The latter interpolates between two extremes based on the cases when 0 or 100% of the deaths in the interval are due to cause j.

IMHO the differences between these approaches are relatively minor and pale in comparison to the real problem, which is not knowing how the force of mortality for cause j would change if all other causes were eliminated.

Cause-Deleted Life Tables

This is formally exactly the same problem, but instead of assuming $\mu_j(x)$ is the only force operating, we assume that this cause has been eliminated, but all *other* causes continue to operate as before. In other words the overall force of mortality is now $\mu(x) - \mu_j(x)$. A cause-deleted life table is thus exactly the same as the associated single-decrement life table for all other causes.

Exactly the same issues noted above apply. The key assumption is that eliminating a cause of death would leave the other forces of mortality unchanged. The life table would change, of course, but only because exposure would change.

The website reproduces the calculations in Box 4.2 using Chiang's method and confirms that if neoplasm were eliminated but the rates for all other causes remained unchanged, life expectancy would increase from 78.92 to 82.46 years, a gain of 3.54 years.

I also show that using instead the simpler assumption that all forces of mortality are constant within each age interval leads to almost identical estimates.

It is worth noting that if there is heterogeneity and people at high risk of cancer are also at high risk of dying of other causes, the increase in life expectancy might be less than calculated. On the other hand if those survivors benefit from lower future death rates for other causes the gain might be larger than estimated. In short, the calculation is just a counterfactual based on the assumption that eliminating some causes of death doesn't change others and all age-specific death rates remain the same.