Non-proportional Hazards: What, Me Worry?

When the hazards in a treated and control group are proportional in time to event trials, the log-rank test is the optimal approach to comparing the groups and the Cox model fits perfectly. Of course, proportionality rarely if ever holds exactly, raising the question of when we should worry enough to abandon these widely used methods and opt for other approaches. This talk deals with some of the common causes of non-proportionality, focusing on three classes of reasons. First, in some situations the action of the experimental intervention leads to early harm and late benefit. Examples are trials of invasive surgery and, more recently, CAR-T. Second are cases where the study population at randomization is at non-constant risk and the time to event tends to be shorter in those of higher risk status. In these trials, the risk profile of the two groups, experimental and control, changes over time. When the treatment is highly effective, the treated group can experience higher hazard than the control as the study progresses. A third source of non-proportionality can arise when a secular change occurs in the population at large. Examples might include a new standard of care introduced during the trial, or a pandemic that affects the two treatment groups differently, Kaplan-Meier curves set time zero at the time of randomization, thereby erasing the ability to visualize secular time from them. The talk discusses ways of determining if the hazards are non-proportional, methods for identifying the cause of failure of proportionality, and considerations for when to replace the log-rank and Cox model with another approach. Some people argue to use Restricted Mean Survival Time or parametric models when faced with non-proportional hazards. Spoiler alert – I will argue that in most cases, the log-rank and Cox are appropriate as long as we interpret the hazard ratio sensibly.