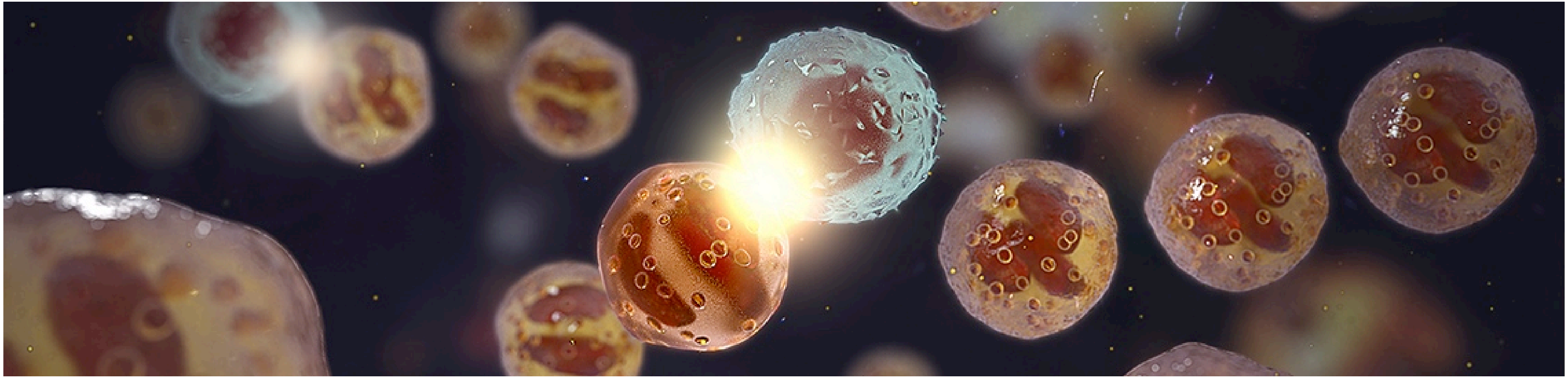
Testing treatment effect when hazards are non-proportional

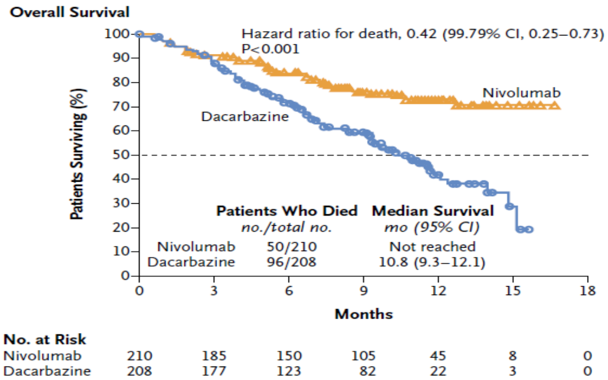
David Wright, PhD, Head Statistical Innovation, AZ

3rd EFSPI Workshop on Regulatory Statistics, Basel

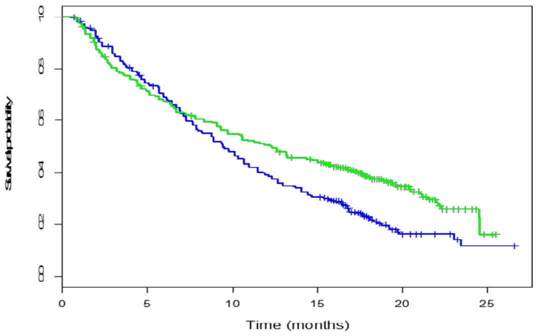
25 September 2018



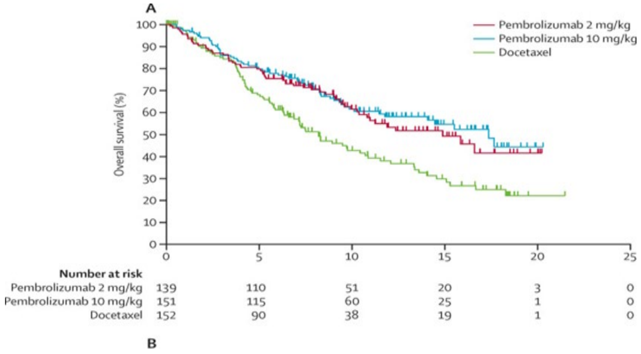
Recent IO trials have brought the concept of NPH in the forefront….



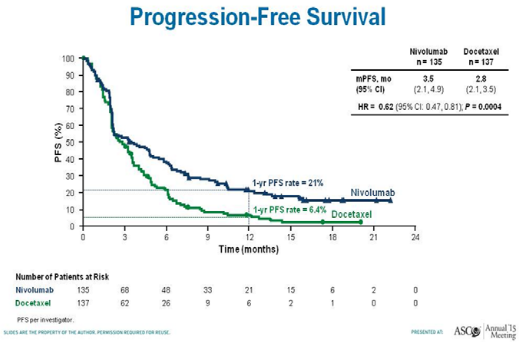
OS: Nivo in melanoma



OS: Nivo in NSCLC

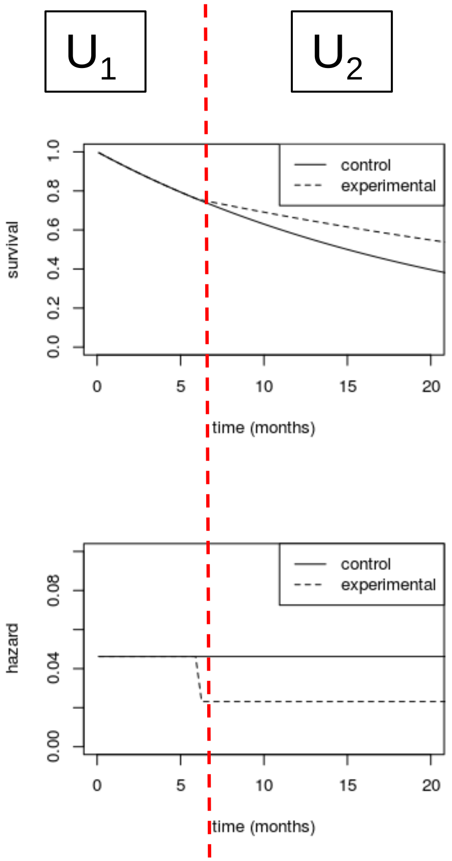


OS: Pembro in NSCLC



PFS: Nivo in NSCLC

Log rank test



$$U_1 = \sum_{t_i<6} O_{Ci} - EC_i$$

$$U_2 = \sum_{t_i>6} O_{Ci} - EC_i$$

The log-rank statistic, $U = U_1 + U_2$, may have very low power because $E(U_1) \approx 0$.

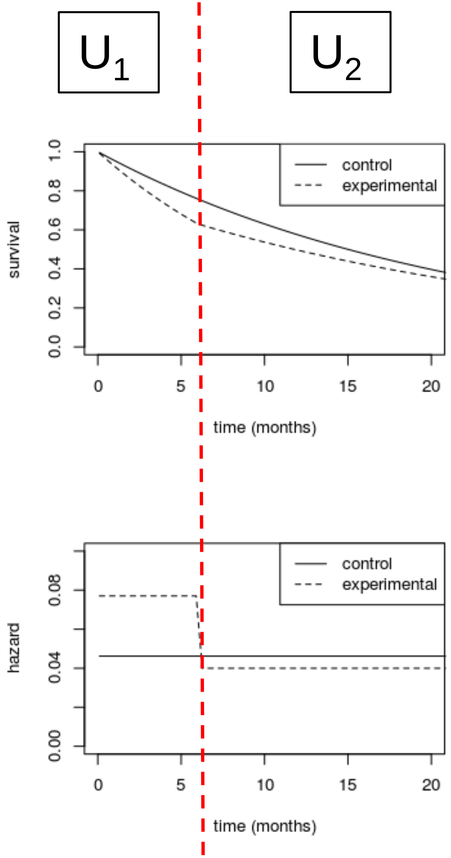
Since we expect $E(U_2) > 0$, there has been interest in using a weighted log-rank statistic, e.g.,

$$U_W = 0 \times U_1 + 1 \times U_2,$$

and claim a significant positive result when e.g.,

$$U_W / se(U_W) > 1.96.$$

Weighted log-rank test



The problem is that we can find situations where $E(U_2) > 0$, yet survival on the experimental arm is worse at all time points.

In this case pr($U_W / se(U_W) > 1.96$) may far exceed 2.5%.

We have proposed a "modestly weighted log-rank test" that avoids this problem but still manages to improve power over the standard log-rank test in delayed effect scenarios.

Magirr, D., & Burman, C. F. (2018). Modestly Weighted Logrank Tests. *arXiv:1807.11097*.

Discussion points

Which null hypothesis should we consider: $S_E(t) = S_C(t)$ or $S_E(t) \leq S_C(t)$?

Should hypothesis testing and estimation match up, or is it acceptable to use different approaches?