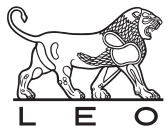
FMS/DsBs 2018

# Testing treatment equivalence in a survival setting

Joint work with Julie K. Furberg and Thomas Scheike

Christian Pipper

E-mail: tpbdk@leo-pharma.com



# Preliminaries

## Setting

Randomized clinical trial with two treatment arms and time from randomization to event of interest as endpoint.

## Objective

To assess if the difference in event progression between treatment arms is small enough to be clinically irelevant.

## Objective formalized

Let $T_m$, $m = 0,1$ denote the time from randomization to event for a person in treatment arm $m$ and let $S_m(t) = P(T_m > t)$ denote the survival functions. Then we are seeking to test:

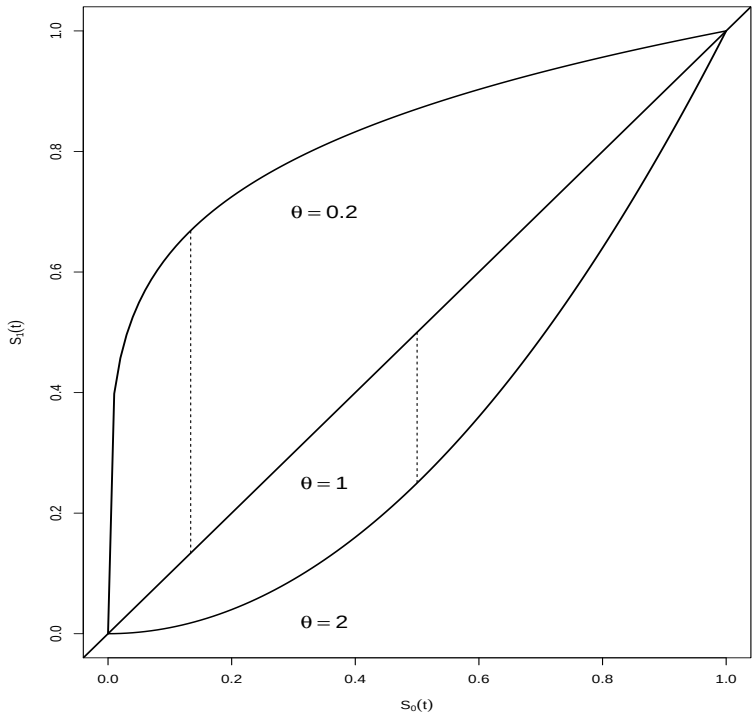
$$H_0 : \sup_{t \geq 0} |S_0(t) - S_1(t)| \geq \delta, \; H_a : \sup_{t \geq 0} |S_0(t) - S_1(t)| < \delta,$$

# The traditional approach

## Assume Cox regression

$$S_1(t) = S_0(t)^\theta, \quad t < \infty,$$

where $\theta$ denotes the hazard ratio




- $\sup_{t \geq 0} |S_0(t) - S_1(t)| = \sup_{u \in [0,1]} |u - u^\theta|$
- maximal distance is attained at $\theta^{\frac{1}{1-\theta}}$ and is symmetric in $\log(\theta)$
- $H_0$ Equivalent to testing

$$\tilde{H}_0 : |\log(\theta)| \geq \log(1+\varepsilon),$$

where
$\delta = (1+\varepsilon)^{-\frac{1}{\varepsilon}} - (1+\varepsilon)^{-\frac{(1+\varepsilon)}{\varepsilon}}$

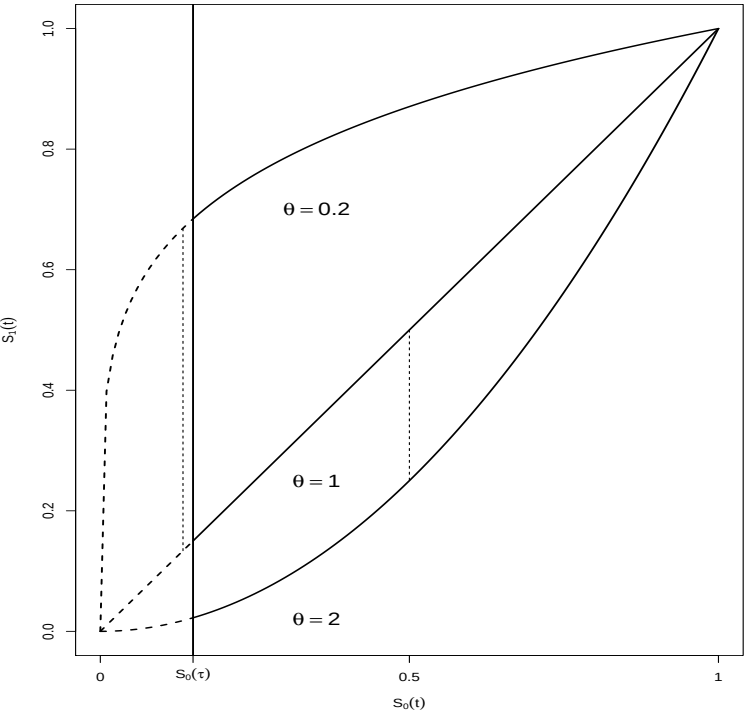
# A practical challenge

- So far we have forgotten all about end of follow-up and assumed infinite follow-up.
- In practice we always have a finite end of follow-up $\tau$.
- This implies that we only have data evidence to assess equivalence of event progression up to end of follow-up

$$H_0^\tau : \sup_{t\in[0,\tau]} |S_0(t) - S_1(t)| \geq \delta,\ H_a^\tau : \sup_{t\in[0,\tau]} |S_0(t) - S_1(t)| < \delta.$$

- We can of course close our eyes to this and follow the traditional strategy.
- The consequence ,however, will be that we may fail to declare equivalence solely based on our model beliefs beyond end of follow-up ($H_0^\tau \subseteq H_0$)
- The cautious would claim that we can never reject $H_0$ unless we know that $\max(S_0(\tau),S_1(\tau)) < \delta$.

# A solution




- Assume Cox regression until end of follow-up:

$$S_1(t) = S_0(t)^\theta, \quad t < \tau.$$

- Show that

$$m_\tau = \sup_{t \in [0,\tau]} |S_0(t) - S_1(t)|$$

$$= \left| \max\left\{ S_0(\tau), \theta^{\frac{1}{1-\theta}} \right\} - \max\left\{ S_0(\tau), \theta^{\frac{1}{1-\theta}} \right\}^\theta \right|.$$

- Base assessment of $H_0^\tau$ directly on $m_\tau$

# A solution: continued

- Estimate $m_\tau$ by plugging in the KM estimate of $\hat{S}_0(\tau)$ and the hazard ratio estimate $\hat{\theta}$ from Cox regression.
- Show that

$$\sqrt{n}(\hat{m}_\tau - m_\tau) \xrightarrow{\mathcal{D}} \mathcal{N}(0,\xi^2),$$

  where $\xi$ can be consistently estimated.
- Test equvivalence by

$$Z = \frac{\sqrt{n} \cdot (\hat{m}_\tau - \delta)}{\hat{\xi}}$$

  and reject $H_0^\tau$ when

$$Z < q_\alpha,$$

  where $q_\alpha$ denotes the $\alpha$-quantile in the standard normal distribution.

# Assessing performance by simulation

- $S_0(t) = exp(-t)$, $\theta = 1.5$, $\tau = 0.2$
- $\delta = 0.135$
- $m_\tau = 0.078$, $m_\infty = 0.148$ so $H_0^\tau$ is false and $H_0$ is true

Estimated probability of rejecting the null hypothesis.

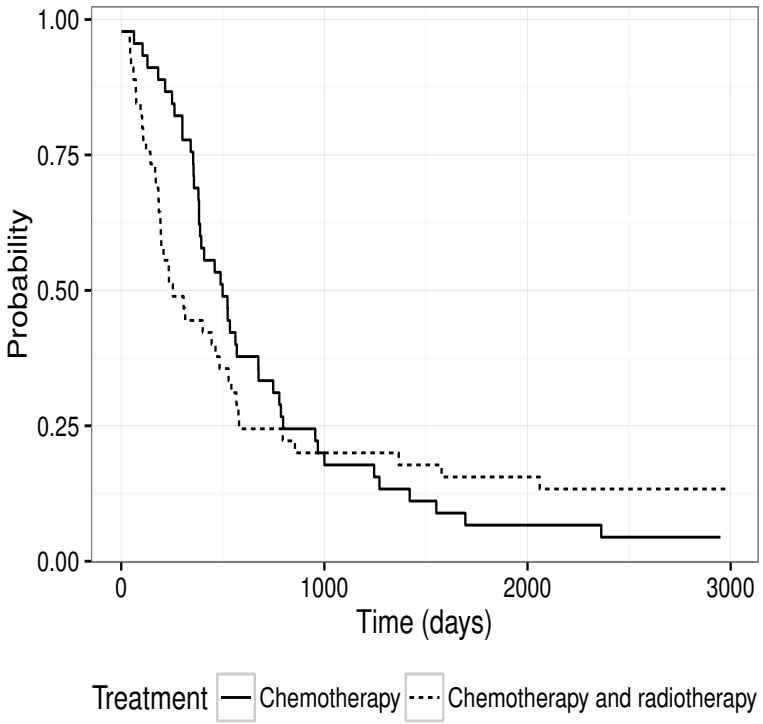
| Sample size, $n$ | P(reject $H_0^\tau$), Modified equivalence test | P(reject $H_0$), Classical equivalence test |
|---|---|---|
| 5000 | 1.000 | 0.013 |
| 1000 | 0.711 | 0.029 |
| 500 | 0.453 | 0.033 |
| 200 | 0.231 | 0.039 |
| 100 | 0.036 | 0.030 |

We loose out on equivalence by trying to assess equivalence beyond end of follow-up based solely on model beliefs.

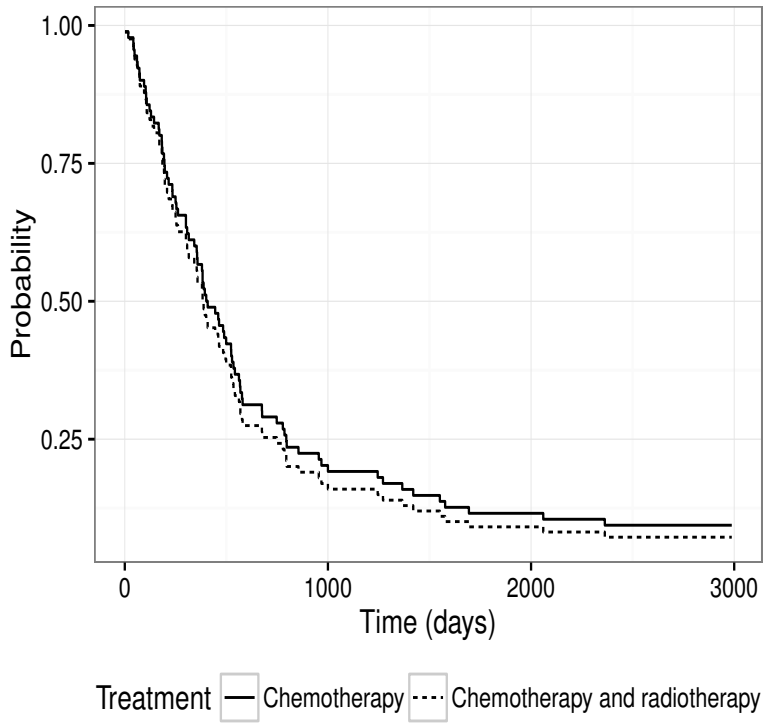
# A challenge about model assumptions

Consider the following example on surival of 90 gastric cancer patients 1:1 randomly assigned to two treatment arrms

## Kaplan Meier Curves




## Survival: Cox regression




Cox model yields bad fit!

# A non-parametric equivalence test

For situations such as the Gastric cancer example a non-parametric equivalence test would be attractive

## Simple idea

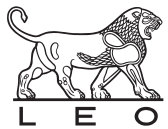
To assess $H_0^\tau$ simply use:

$$\hat{m}_\tau = \sup_{t \in [0,\tau]} |\hat{S}_0(t) - \hat{S}_1(t)|,$$

where $\hat{S}_m$ denotes the Kaplan-Meier estimator in the $m$th treatment arm.

## Theoretical Challenge

The asymptotic distribution is very complicated!



# Asymptotics: Preliminaries

- Denote $f_n(t) = \hat{S}_0(t) - \hat{S}_1(t)$, $f(t) = S_0(t) - S_1(t)$, $G_n(t) = \sqrt{n}\{f_n(t) - f(t)\}$
- Note that $G_n$ converges to a tight zero mean Gaussian process $\mathcal{W}$ with a variance/covariance $\xi(s \wedge t)$ that can be consistently estimated by $\hat{\xi}(s \wedge t)$
- For a bounded function $g$ on $[0,\tau]$ denote the signed sets of extremal points by

$$\mathcal{E}^{\pm}(g) = \{s \in [0,\tau] |\ g(s) = \pm \sup_{t \in [0,\tau]} |g(t)|\}.$$

# Asymptotics: Result

For the non-parametric estimator of $m_\tau$ the following asymptotic result holds:

$$\sqrt{n}(\hat{m}_\tau - m_\tau) \xrightarrow{\mathcal{D}} \max[\max_{t\in\mathcal{E}^+(f)}\{\mathcal{W}(t)\}, \max_{t\in\mathcal{E}^-(f)}\{-\mathcal{W}(t)\}]$$

For the special case $\mathcal{E}^{\pm}(f) = \{t^*\}$, where $f$ has exactly one extremal point, the following asymptotic result holds:

$$\sqrt{n}(\hat{m}_\tau - m_\tau) \xrightarrow{\mathcal{D}} \mathcal{N}(0,\xi(t^*)).$$

Moreover, if $f(t^\star) \neq 0$, then for any sequence $t_n^\star \in \mathcal{E}^{\pm}(f_n)$ we have

$$t_n^\star = t^\star + o_P(1)$$

and a consistent estimator of $\xi(t^*)$ is given by $\hat{\xi}(t_n^\star)$.

# Constructing a valid $\alpha$-level equivalence test

- Assess equivalence by

$$Q = \frac{\sqrt{n}\{\hat{m}_\tau - \delta\}}{\sqrt{\xi(\hat{t}_n^\star)}},$$

  where $t_n^\star$ is the smallest extremal point
- Reject $H_0^\tau$ when

$$Q < q_\alpha,$$

  where $q_\alpha$ denotes the $\alpha$-quantile in the standard normal distribution.
- One may show that this is an asymptotically valid $\alpha$-level test
- It is strictly below $\alpha$ level when there is more than one extremal point in $\mathcal{E}^-(f) \cup \mathcal{E}^+(f)$

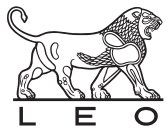
# Testing equivalence for the Gastric cancer example

- With $\delta = 0.2$ We compute $Q = 1.72$ corresponding to a p-value of $\Phi(Q) = 0.95$.
- According to this test there is no evidence of equivalence.
- This we would also expect from looking at the KM-curves
- Using our Cox based proposal we compute $Z = -1.96$ corresponding to a p-value of $\Phi(Z) = 0.025$.
- According to this test there is evidence of equvialence
- This we would definitely not expect from looking at the KM-curves!

# Closing remarks

- We have proposed a Cox based equivalence test that incorporates end-of-follow-up
- It has the advantage of only assessing equivalence during the follow-up period instead of basing assessment on modelling beliefs beyond data evidence.
- We have also proposed a simple (to understand) non-parametric test that is applicable when modelling assumptions such as proportional hazards are not appropriate
- Hopefully we have helped raise awareness that end-of-follow-up should be an integral ingredient in equivalence testing based on time to event data :-)

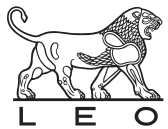
Thanks for your attention!



# Closing remarks

- We have proposed a Cox based equivalence test that incorporates end-of-follow-up
- It has the advantage of only assessing equivalence during the follow-up period instead of basing assessment on modelling beliefs beyond data evidence.
- We have also proposed a simple (to understand) non-parametric test that is applicable when modelling assumptions such as proportional hazards are not appropriate
- Hopefully we have helped raise awareness that end-of-follow-up should be an integral ingredient in equivalence testing based on time to event data :-)

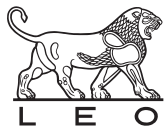
Thanks for your attention!



# Closing remarks

- We have proposed a Cox based equivalence test that incorporates end-of-follow-up
- It has the advantage of only assessing equivalence during the follow-up period instead of basing assessment on modelling beliefs beyond data evidence.
- We have also proposed a simple (to understand) non-parametric test that is applicable when modelling assumptions such as proportional hazards are not appropriate
- Hopefully we have helped raise awareness that end-of-follow-up should be an integral ingredient in equivalence testing based on time to event data :-)

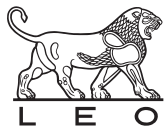
Thanks for your attention!



# Closing remarks

- We have proposed a Cox based equivalence test that incorporates end-of-follow-up
- It has the advantage of only assessing equivalence during the follow-up period instead of basing assessment on modelling beliefs beyond data evidence.
- We have also proposed a simple (to understand) non-parametric test that is applicable when modelling assumptions such as proportional hazards are not appropriate
- Hopefully we have helped raise awareness that end-of-follow-up should be an integral ingredient in equivalence testing based on time to event data :-)

Thanks for your attention!



# Closing remarks

- We have proposed a Cox based equivalence test that incorporates end-of-follow-up
- It has the advantage of only assessing equivalence during the follow-up period instead of basing assessment on modelling beliefs beyond data evidence.
- We have also proposed a simple (to understand) non-parametric test that is applicable when modelling assumptions such as proportional hazards are not appropriate
- Hopefully we have helped raise awareness that end-of-follow-up should be an integral ingredient in equivalence testing based on time to event data :-)

Thanks for your attention!