Conferencia Española de Biometría
Sevilla, 15 de septiembre, 2017

Non-proportional hazards. Consequences on required sample sizes

jointly with KyungMann Kim, Moisès Gómez-Mateu

Outline of the Talk

1. Hazard Ratio of Composite Endpoints (CE)
2. Behaviour of Hazard Ratio over time $\mathrm{HR}_*(t)$
3. Measures r and R to evaluate non constancy of $\mathrm{HR}_*(t)$
4. **CompARE**: Interactive Web platform to study Composite Endpoints and get sample sizes

Hazard Ratio in RCT

HR routinely used as a measure to summarize treatment effect on time-to-event endpoints.

$\mathrm{HR}(t)$: hazard in treatment group divided by hazard in control group:

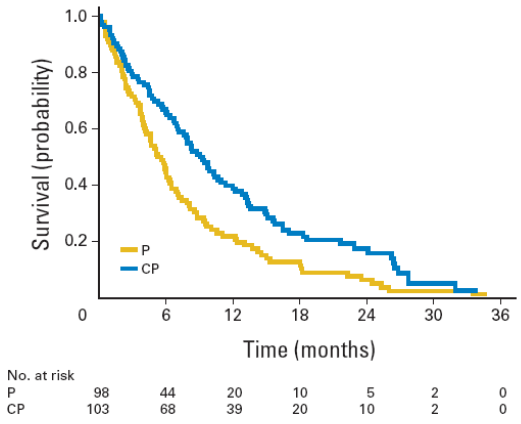
$$\mathrm{HR}(t) = \frac{\lambda^{(1)}(t)}{\lambda^{(0)}(t)} \quad \text{for all} \quad t$$

$\mathrm{HR}(t)$ depends on the study specific follow-up time.

When $\mathrm{HR}(t) \approx h \quad \forall t \implies$ h quantifies the survival-group difference because it might appropriately capture relative treatment effect between treatment arms.

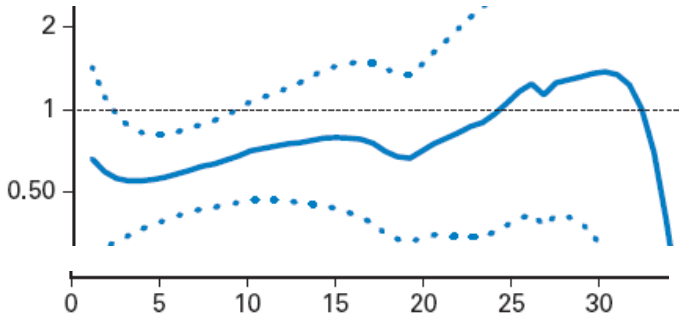
When $\mathrm{HR}(t) \neq h \implies$ the average of $\mathrm{HR}(t)$ over time is not a meaningful measure, cannot be translated into an understandable clinical benefit and common formulae to calculate sample sizes are not valid.

Motivating example



ECOG trial [1]**:**

Primary endpoint: Overall survival.
Control group: Single-agent pemetrexed (n=98).
Treatment group: Carboplatin+pemetrexed (n=103)
Hazard ratio: 0.62 (p-value=0.001)



Estimated HR [2]**:**

HR visually in favour of the new treatment early in the study and then approaching 1 afterwards.

[1] Zukin M, et al: Randomized phase III trial of single-agent pemetrexed versus carboplatin and pemetrexed in patients with advanced non-small-cell lung cancer and ECOG performance status of 2. J Cli Onc 31:2849-53,2013.
[2] Uno H, et al. Moving Beyond the Hazard Ratio in Quantifying the Between-Group Difference in Survival Analysis. J Clin Oncol. 2014 Aug 1;32(22):2380-5.

Non constant Hazard Ratios are being detected more frequently. Why is so?

- New therapies have different modes of action: the effect of the intervention can diminish after a period of time
- Phase III trials are much larger $\Rightarrow$ more chances to detect Non constant hazard ratios $\equiv$ Non Proportionals Hazards (NPH)
- To test smaller treatment effects with new and better therapies Composite endpoints (CE) used more often
- CE $\Rightarrow$ NPH

COMPOSITE ENDPOINTS $\mathcal{E}_* = \mathcal{E}_1 \cup .... \cup \mathcal{E}_k$

1. **CANCER CLINICAL TRIALS**

   T_1 time to Disease progression ($\mathcal{E}_1$)

   T_2 time to Overall survival ($\mathcal{E}_2$)

   T_* time to PFS: Progression-free survival ($\mathcal{E}_*$)

2. **CARDIOVASCULAR STUDIES**

   T_1 time to Death ($\mathcal{E}_1$)

   T_2 time to MI ($\mathcal{E}_2$)

   T_* time to earlier event between Death and MI ($\mathcal{E}_*$)

COMPOSITE ENDPOINTS $\mathcal{E}_* = \mathcal{E}_1 \cup .... \cup \mathcal{E}_k$

1. **CANCER CLINICAL TRIALS**
   - T_1 time to Disease progression ($\mathcal{E}_1$)
   - T_2 time to Overall survival ($\mathcal{E}_2$)
   - T_* time to PFS: Progression-free survival ($\mathcal{E}_*$)
2. **CARDIOVASCULAR STUDIES**
   - T_1 time to Death ($\mathcal{E}_1$)
   - T_2 time to MI ($\mathcal{E}_2$)
   - T_* time to earlier event between Death and MI ($\mathcal{E}_*$)

SURVIVAL COMPOSITE ENDPOINTS (CE)

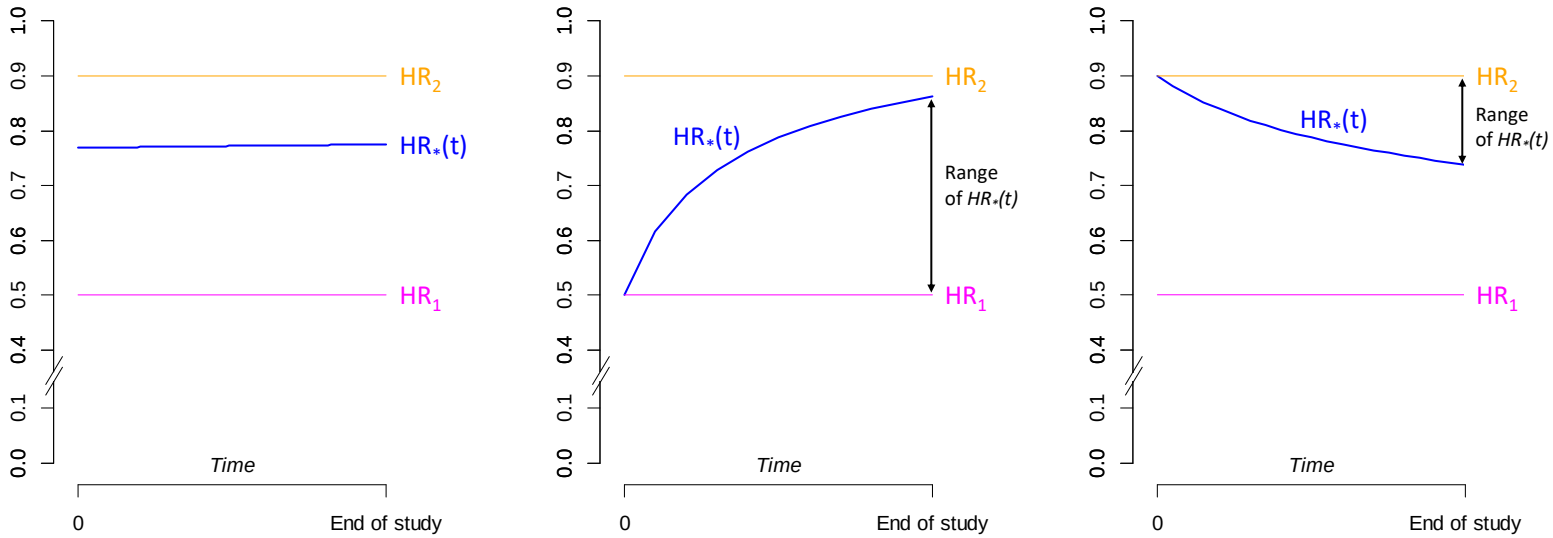
$T_* = \min\{T_1, T_2\}$ being T_j time from randomization to $\mathcal{E}_j$

$\mathrm{HR}_*(t)$ depends on:

- Joint law of (T_1, T_2): Frank's copula and Spearman's correlation between T_1 and T_2
- Laws of T_1 and T_2: Weibull densities, Probabilities of observing T_1 and T_2 in group 0 and constant HR_1 and HR_2

Behavior of the hazard ratio over time

Even if HR_1 and HR_2 are constant $\not\Rightarrow$ $HR_*(t)$ constant.



- $HR_*(t)$ lies between the hazard ratios of its components
- For two given constant hazard ratios, we can anticipate that the CE will show no less treatment effect than the less effective component.

Under which circumstances $HR_*(t)$ may potentially result in greater departure from constancy?

What is the impact on sample size of erroneoulsly using a constant summary?

Assessment of how $\mathrm{HR}_*(t)$ deviates from constancy

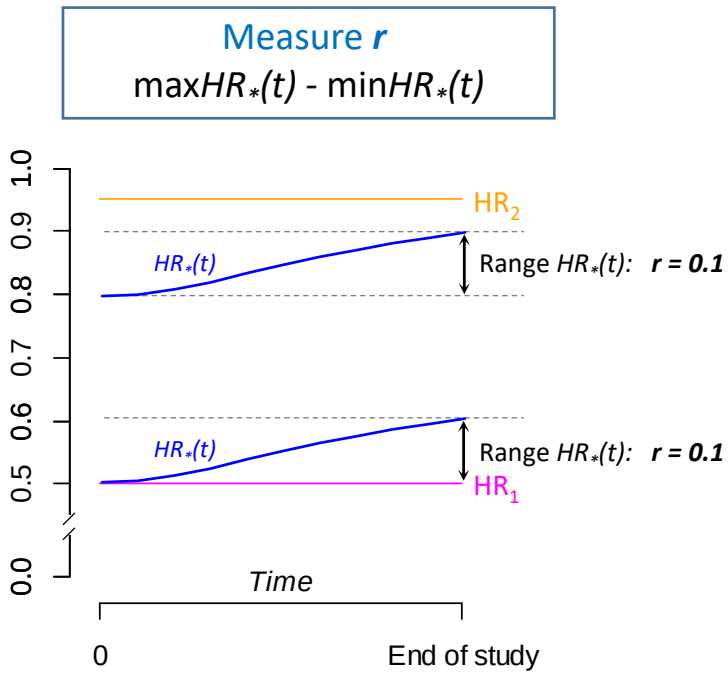
When $\mathrm{HR}_*(t)$ varies over time, to capture treatment effect we migth use:

- $MHR_* = \max_t \mathrm{HR}_*(t) \equiv$ minimum detectable effect
- $mHR_* = \min_t \mathrm{HR}_*(t) \equiv$ maximum detectable effect
- $aHR_* =$ average of $\mathrm{HR}_*(t) \equiv$ average detectable effect

Assesment might be based on

1. Range $r = MHR_* - mHR_*$
2. R Relative difference in sample size

Appropriateness of range r



If $\mathrm{HR}_*(t) \approx h_*$ constant $\iff MHR_* \approx mHR_* \approx aHR_* \iff r \approx 0$

Different curves $\mathrm{HR}_*(t)$ might entail equal range r with different consequences in needed sample size if aHR_* would be used as summary

Consequences depend on:

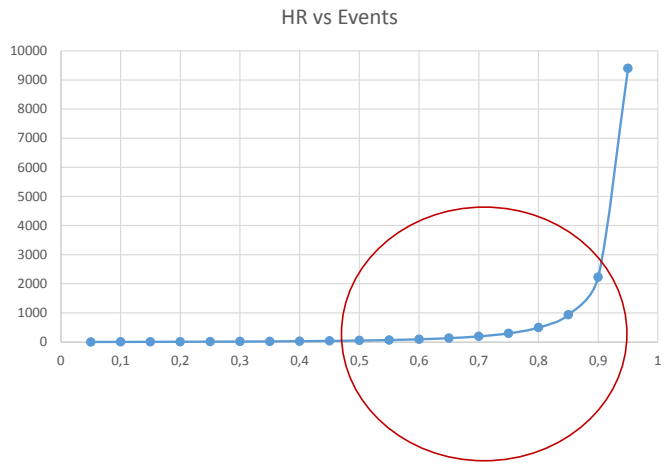
1. Value of aHR_*
2. Laws of T_1 and T_2

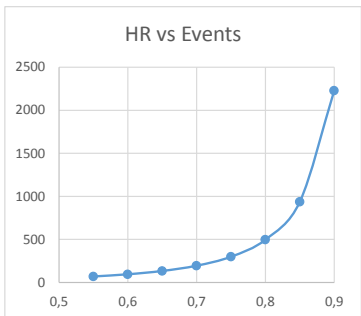
1 Influence of value of aHR_*

If $\alpha = 0.05$ and $1 - \beta = 0.8$ and if $HR_*(t) \approx \mathrm{aHR}_*$

$$\text{needed events } e_* = 4\left(\frac{z_\alpha + z_\beta}{\log(\mathrm{aHR}_*)}\right)^2 ; \text{sample size } \quad n_* = \frac{2e_*}{p_*^{(0)} + p_*^{(1)}} \qquad 1$$

HR	Events
0,05	3
0,1	5
0,15	7
0,2	10
0,25	13
0,3	17
0,35	22
0,4	29
0,45	39
0,5	51
0,55	69
0,6	95
0,65	133
0,7	194
0,75	299
0,8	497
0,85	936
0,9	2228
0,95	9400





Even small differences of the anticipated value for aHR_* can dramatically change the required number of events: for $\mathrm{aHR}_* = 0.85 \implies e_* = 936$ while $\mathrm{aHR}_* = 0.75 \implies e_* = 299$

[1] Schoenfeld (Biometrika, 1981)

❷ Laws of T_1 and T_2

$p_1^{(0)} = 0.15$, $p_2^{(0)} = 0.5$, $\mathrm{HR}_1 = 0.6$, $\mathrm{HR}_2 = 0.9$, $\rho = 0.3$ and
The same $r = 0.06$.

For $\alpha = 0.05$, $1 - \beta = 0.8$ and if $\mathrm{HR}_*(t)$ is summarized by aHR_*

(A) T_1 increasing hazards, T_2 exponential

- $e_{aHR_*} = 4\left(\frac{z_\alpha + z_\beta}{\log(aHR_*)}\right)^2 =$ **1230** events

(B) T_1 exponential, T_2 decreasing hazards

- $e_{aHR_*} =$ **1094** events

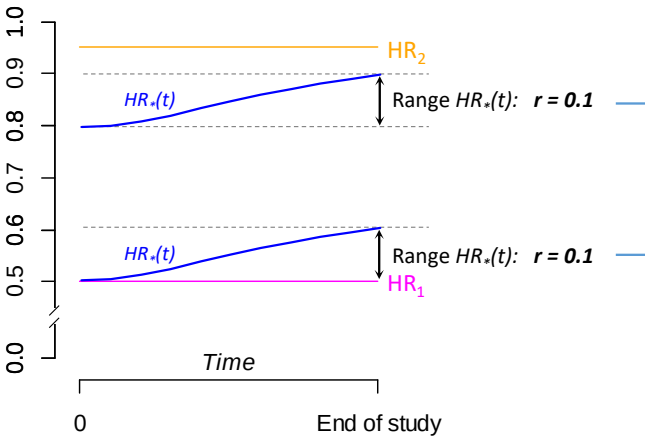
Range r as a measure of the departure of $\mathrm{HR}_*(t)$ is not convenient if we want to take into account impact on sample size

Relative measure R

$$R = \left(\frac{\log(aHR_*)}{\log(MHR_*)} \right)^2 = \frac{e_{MHR_*}}{e_{aHR^*}} = \frac{n_{MHR_*}}{n_{aHR^*}}$$

Measure *r*
max$HR_*(t)$ - min$HR_*(t)$

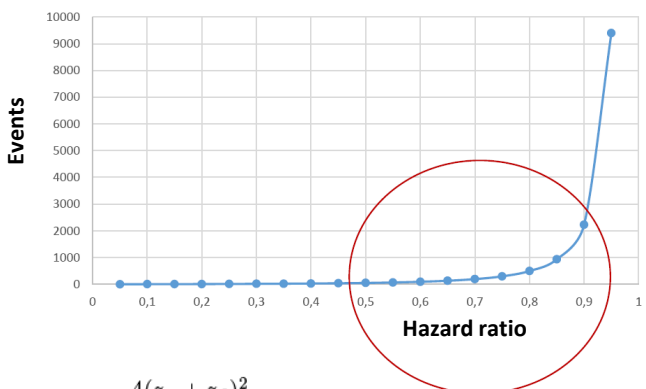
Measure *R*
$[\log(\text{mean}HR_*(t))/\log(\max HR_*(t))]^{2\,(+)}$



***R* = 1,89**
Relative difference between required number of events of **89%**

***R* = 1,34**
Relative difference between required number of events of **34%**

Hazard ratio vs events



$$e = \frac{4(z_\alpha + z_\beta)^2}{(\ln(HR))^2}$$ *α=0.05, β=0.2*

(+) ***R***: Equivalent to the ratio of required number of events for meanHR* vs maxHR*.

- It puts more **weight** for departures with higher hazard ratios.
- It is **easy to interpret**.

Scenarios for studying R

Parameters				
$p_1^{(0)}, p_2^{(0)}$	0.1	0.3	0.5	
HR_1, HR_2	0.6	0.7	0.8	0.9
ρ	0.1	0.3	0.5	
Distribution	(Decr. hazards)	(Exponential)	(Incr. hazards)	
β_1, β_2	0.5	1	2	
Number of scenarios*	**3 888**			

Table:

$p_1^{(0)}$, $p_2^{(0)}$: probabilities of observing $\mathcal{E}_1$ and $\mathcal{E}_2$ in the control group

ρ: Spearman's rank correlation between T_1 and T_2

HR_1, HR_2: constant hazard ratios for $\mathcal{E}_1$ and $\mathcal{E}_2$

β_1, β_2: shape parameters of the Weibull distribution for T_1 and T_2.

Departure from constancy in terms of R

R ratio of required number of events if using constant hazard ratio equal to aHR_* versus MHR_*

	R		
	Minimum	**Median**	**Maximum**
Treatment effect (HR_1, HR_2: hazard ratios for $\mathcal{E}_1$, $\mathcal{E}_2$)			
$HR_1 = HR_2$	1	1.05	1.35
$\lvert HR_1 - HR_2 \rvert = 0.1$	1	1.2	3.49
$\lvert HR_1 - HR_2 \rvert = 0.2$	1	1.49	8.18
$\lvert HR_1 - HR_2 \rvert = 0.3$	1	2.06	15.65
Laws of each component			
Both decreasing hazards ($\beta_1 = \beta_2 = 0.5$)	1	1.04	1.23
Both exponential ($\beta_1 = \beta_2 = 1$)	1	1.04	1.28
Both increasing hazards ($\beta_1 = \beta_2 = 2$)	1	1.06	1.44
Different behaviour in hazards ($\beta_1 \neq \beta_2$)	1.01	1.39	15.65
Correlation			
Weak ($\rho = 0.1$)	1	1.07	14.97
Mild ($\rho = 0.3$)	1.01	1.13	15.19
Moderate ($\rho = 0.5$)	1.01	1.18	15.65
Global	**1**	**1.15**	**15.65**

Criterion to decide whether aHR_* is not a meaningful summary for $HR_*(t)$

$HR_*(t)$ is remarkably non-constant and aHR_* should not be used to assess treatment differences and plan the RCT if $R > 1.25$ because required sample size using MHR_* is 25% larger than that using aHR_*

	Treatment effect				
Laws of each component	$HR_1 = HR_2$	$\lvert HR_1 - HR_2 \rvert = 0.1$	$\lvert HR_1 - HR_2 \rvert = 0.2$	$\lvert HR_1 - HR_2 \rvert = 0.3$	
Both decreasing hazards	0%	0%	0%	0%	**0%**
Both exponential	0%	0%	0%	4%	**0%**
Both increasing hazards	0%	0%	3%	11%	**2%**
Different behaviour on hazards	2%	65%	93%	98%	**60%**
	2%	**43%**	**62%**	**67%**	**41%**

Table: Percentages of scenarios where $R > 1.25$

WARNING:

Be aware if different laws are governing the behaviour of T_1 and T_2 and if expected treatment effects on each outcome are quite apart

Web for Studying Composite Endpoints

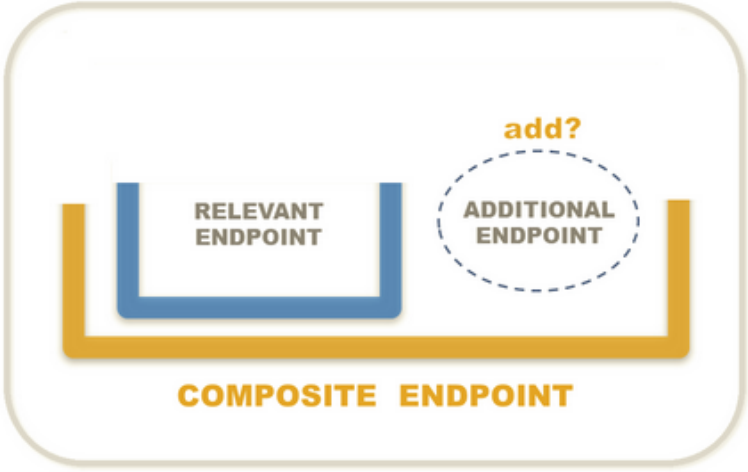
http://cinna.upc.edu/compare/

Home | ARE value | ARE by correlation (plot) | ARE by correlation (table) | Sample Size | Study of HR*(t) | Help

Welcome to Compare platform.

This website will help you to:

* Analyze whether you should use a Composite endpoint as Primary endpoint.
* Sample size for time-to-event data.
* Compare different scenarios depending on your candidate endpoints.
* Get helpful numerical and intuitive graphical results.



Sample Size as a function of correlation

Home | ARE value | ARE by correlation (plot) | ARE by correlation (table) | Sample Size | Study of HR*(t) | Help

Total sample size using Endpoint 1:

6811

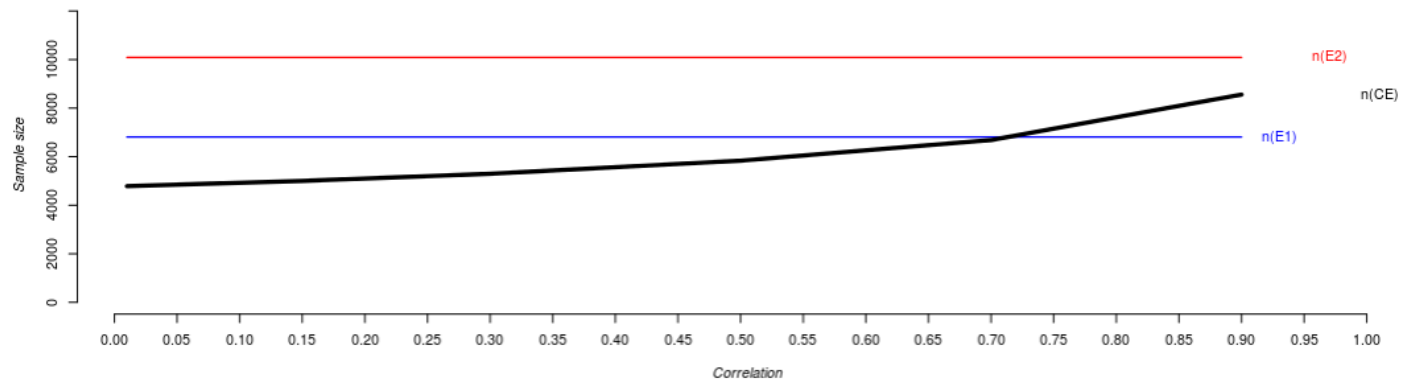
Total sample size using Endpoint 2:

10086

Total sample size using Composite Endpoint:

5834

Sample size depending on different correlations



$\alpha = 0.05,\ 1-\beta = 0.8,\ p_1^{(0)} = 0.05,\ \mathrm{HR}_1 = 0.7 \Rightarrow \mathbf{n_1} = \mathbf{4560}$ patients

If $p_2^{(0)} = 0.05,\ \mathrm{HR}_2 = 0.8 \Rightarrow \mathbf{n}_* = \begin{cases} \mathbf{3699}, & \text{if } \rho = 0.1 \\ \mathbf{4149}, & \text{if } \rho = 0.5 \end{cases}$

WRAPPING UP

- Composite Endpoints (CE) are very often used as PE in phase 3 RCT
- Determination of sample size (SS) is fundamental
- Hazard Ratio summaries are often used
- $HR_*(t)$ often changes over time SS for CE

TAKE HOME MESSAGE: Before planning a RCT with a CE, take into account the expected treatment effect for each component and the laws for the times to each of these events to decide if a constant summary is a good quantification of the treatment effect

ONGOING RESEARCH

1. Alternative summary measures when PH fails and CE are being used
2. Web interfaces **CompARE** for Binary and Time-to-event CE

See you in Barcelona in JULY 2018 !!!!!!


THANK
YOU