Oxford University Clinical Research Unit

January 14, 2020

EXPLORING THE USE OF COMPOSITE ENDPOINTS IN CLINICAL TRIALS

Guadalupe Gómez Melis

thanks to Jordi Cortés and Marta Bofill

UNIVERSITAT POLITÈCNICA DE CATALUNYA

GRBIO

BIOstatnet

Primary endpoint in a randomized clinical trial (RCT)

- Outcome defined by the research question of interest.
- Should be important to patients, amenable to unbiased assessment and influenced by the treatment.
- Its ultimate goal is to demonstrate the efficacy of a new pharmaceutical drug or procedure

Primary endpoint

Variable (outcome) measuring the clinical evidence. Key decision for the study because

- efficacy of new treatment
- power
- sample size computation

are based on the **Primary Endpoint**

Medical improvements have led to:

- Decline in the incidence of clinically relevant outcomes
- Decline in mortality for several common disorders
- Improved standard of care

- Reduction in the number of relevant events
- Lower event rates
- Lower effect sizes

HENCE **relevant endpoints** are observed **less often** and the effect of treatment is **diminished**.

Combining several endpoints might be a solution.

Composite Endpoints (CE)

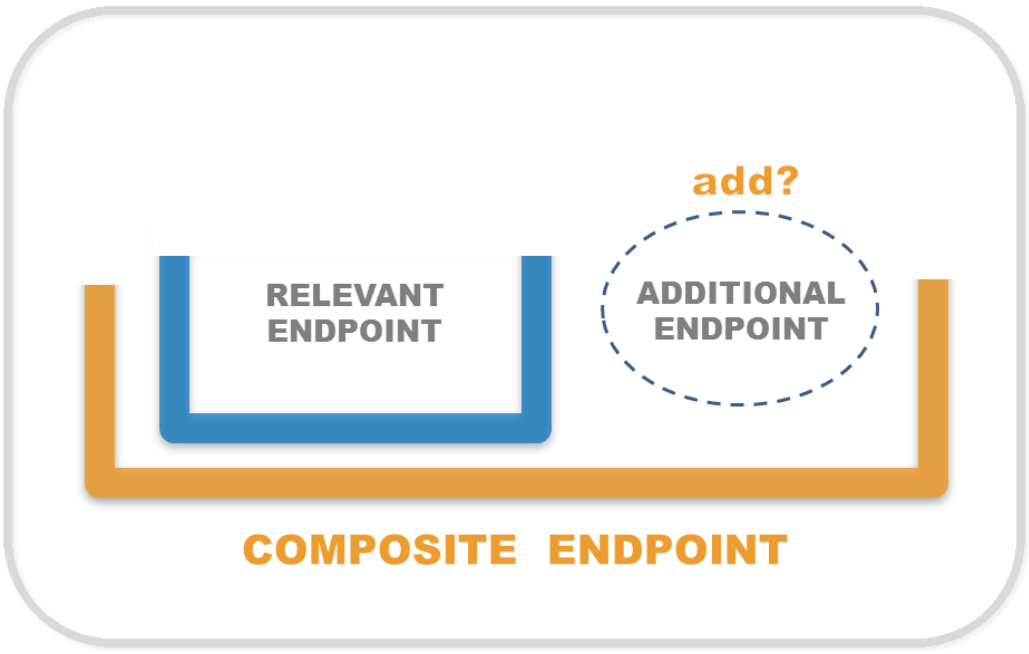
Composite endpoints, CE: Outcome defined as the combination of multiple distinct type of events into a single variable for a given period of follow-up.

PFS: progression-free survival in Oncology trials

$\mathcal{E}_1$: disease progression

$\mathcal{E}_2$: death

$\mathcal{E}_* = \mathcal{E}_1 \bigcup \mathcal{E}_2$: PFS



CE are very often used because:

1. Better description of the disease
2. Provides net clinical benefit of a therapy
3. Avoid adjustment for multiple comparisons
4. Achieves higher event rates
5. *Hopefully* improves statistical efficiency. If so, it would
   - need smaller sample sizes
   - achieve larger power
   - need shorter follow-up times

We will show that a thorough exploration of the Composite components and their association is needed to decide whether to use a CE as the Primary Endpoint

Concerns Using a CE

Main concern

Is the composite of outcome $\mathcal{E}_1$ or outcome $\mathcal{E}_2$ clinically meaningful?

Qualitative heterogeneity: Component endpoints could be dissimilar in patient's importance. (Ferreira-Gonzalez *et al.* BMJ, 2007)

Trials claiming treatment benefits on the CE could include components:

1. of widely varying importance,
2. Important components associated with control group lower event rates,
3. of greater importance associated with smaller effect sizes

Potential solutions to this problem:

Win Ratio (Pocock *et al*, Eur Heart J, 2012)

Desirability of outcome ranking (Oakes, Biometrika, 2016);

Ordering score (Follmann *et al*, Stat in Med, 2019)

Concerns Using a CE

Quantitative heterogeneity: Results on composite outcomes WOULD NOT GENERALLY IMPLY the same results on the individual components and viceversa

Differential treatment effects on components

- Benefits described for the CE might be "wrongly" presumed to relate to all the components (Freemantle *et al.* JAMA, 2003)
  Treatment effect on CE $\not\Longrightarrow$ Treatment effect on each component
- No treatment effect on CE $\not\Longrightarrow$ No treatment effect on one of the components. (Montori *et al.* BMJ, 2005).

Higher event rates and larger treatment effects on less important components $\Rightarrow$ misleading impressions of the impact of treatment

Regulatory Guidelines

- International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. **ICH-E9** guideline: Statistical Principles for Clinical Trials (CT) (1998)
  - *There should generally be only one primary variable*
  - *If a single primary variable cannot be selected from multiple measurements associated with the primary objective, another useful strategy is to integrate or combine the multiple measurements into a single or composite variable, using a predefined algorithm*
  - *...addresses the multiplicity problem without adjustment to type I error.*
- **EUnetHTA** Eur.Network for Health Technology Assessment guideline: Endpoints used for Relative Effectiveness Assessm. (2015)
- **FDA** Guidance for Industry: Multiple Endpoints in CT (2017)
- **ICH-E9 (R1)**: Estimands and Sensitivity Analysis in CT (2017)

Composite Endpoints in RCT should

- Be pre-specified in the protocol
- Individual components should be clinically meaningful
- Individual components should be of similar importance
- Expected effects on each component should be similar
- Clinically more important components should not affect negatively
- Include mortality whenever is appropriate

The decision to use a Composite Endpoint as the Primary Endpoint strongly depends on

- the anticipated event rates
- effects sizes
- association between the components

CE IN SELECTED THERAPEUTIC AREAS

1. **ONCOLOGY TRIALS**

   $\mathcal{E}_1$: disease progression

   $\mathcal{E}_2$: death

   $\mathcal{E}_* = \mathcal{E}_1 \bigcup \mathcal{E}_2$: PFS: progression-free survival

2. **HIV STUDIES**

   $\mathcal{E}_1$: Virological failure, Initiation of new treatment

   $\mathcal{E}_2$: AIDS, Death

   $\mathcal{E}_*$: Loss of virological response

3. **TRANSMITTED DISEASES. PREVENTION TRIALS**

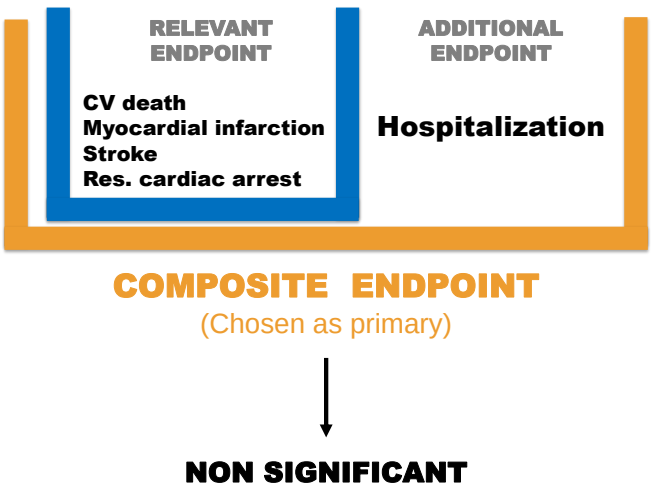
   $\mathcal{E}_1$: HIV infection

   $\mathcal{E}_2$: Hepatitis B infection

   $\mathcal{E}_*$: Either HIV or Hepatitis B Infection

4. CARDIOVASCULAR DISEASE STUDIES

Acute coronary syndrome patients blindly randomized (Tardif JC *et al.*, Lancet 2008)



$\mathcal{E}_1$: Cardiovascular death, myocardial infarction, stroke

$\mathcal{E}_2$: Hospitalization

$\mathcal{E}_*$: MACE: Major Adverse Cardiovascular Events

0. Standard of care ($n = 3066$)
   - 252 events $\mathcal{E}_1$
   - 277 events $\mathcal{E}_2$
   - 529 events $\mathcal{E}_* = \mathcal{E}_1 \cup \mathcal{E}_2$
1. SOC+Succinobucol ($n = 3078$)
   - 207 events $\mathcal{E}_1$
   - 323 events $\mathcal{E}_2$
   - 530 events $\mathcal{E}_* = \mathcal{E}_1 \cup \mathcal{E}_2$

1. Hospital admission component MASKED the mortality effect
2. Succinobucol might have shown a beneficial effect ($p = 0.029$) on $\mathcal{E}_1$

Relevant papers

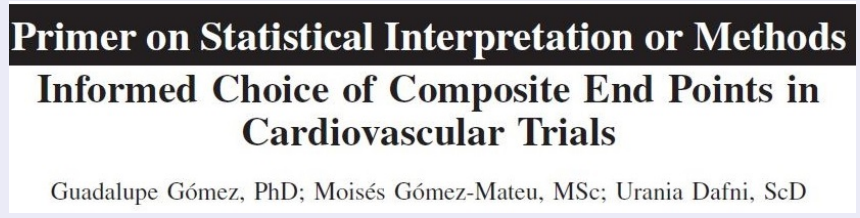
Time-to-event endpoint

Research Article

Statistics in Medicine

Statistical considerations when using a composite endpoint for comparing treatment groups

Guadalupe Gómez[a*†] and Stephen W. Lagakos[b‡]



Primer on Statistical Interpretation or Methods

Informed Choice of Composite End Points in Cardiovascular Trials

Guadalupe Gómez, PhD; Moisés Gómez-Mateu, MSc; Urania Dafni, ScD

JOURNAL OF BIOPHARMACEUTICAL STATISTICS
http://dx.doi.org/10.1080/10543406.2015.1094808

Taylor & Francis
Taylor & Francis Group

Selecting the primary endpoint in a randomized clinical trial: The ARE method

Oleguer Plana-Ripoll[a] and Guadalupe Gómez[b]

Selection of the primary end point in an observational cohort study

Guadalupe Gómez,[1] Oleguer Plana-Ripoll,[2] Urania Dafni[3,4]

Binary endpoint

RESEARCH PAPER

Biometrical Journal

Selection of composite binary endpoints in clinical trials

Marta Bofill Roig | Guadalupe Gómez Melis

RESEARCH ARTICLE

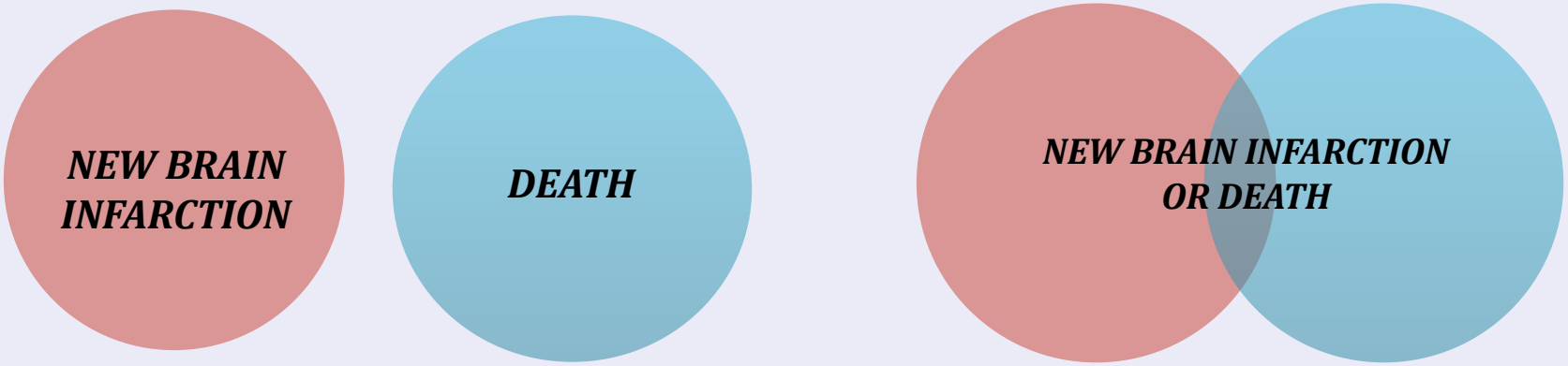
WILEY Statistics in Medicine

A new approach for sizing trials with composite binary endpoints using anticipated marginal values and accounting for the correlation between components

Marta Bofill Roig | Guadalupe Gómez Melis

BINARY CE IN INFECTIOUS DISEASES

ASPIRIN TRIAL FOR TBM IN HIV- ADULTS



1. $\mathcal{E}_1$ = New brain infarction by 60 days
2. $\mathcal{E}_2$ = Death by 60 days
3. $\mathcal{E}_*$ = Brain infarction or death by 60 days. Primary Efficacy Endpoint

A randomised double blind placebo controlled phase 2 trial of adjunctive aspirin for tuberculous meningitis in HIV-uninfected adults

Nguyen TH Mai[1,2], Nicholas Dobbs[3], Nguyen Hoan Phu[1,2], Romain A Colas[4], Le TP Thao[1], Nguyen TT Thuong[1], Ho DT Nghia[1,2], Nguyen HH Hanh[1,2], Nguyen T Hang[1], A Dorothee Heemskerk[1,5], Jeremy N Day[1,6], Lucy Ly[4], Do DA Thu[1], Laura Merson[6], Evelyne Kestelyn[1,6], Marcel Wolbers[1], Ronald Geskus[1,6], David Summers[3], Nguyen VV Chau[1,2], Jesmond Dalli[4], Guy E Thwaites[1,6]*

TIME-TO-EVENT CE IN INFECT. DISEASES

ARREST TRIAL OF RIFAMPICIN VS PLACEBO IN ADULTS WITH *Staphylococcus aureus (SA)* BACTERAEMIA.

1. $\mathcal{E}_1$ = Treatment failure or Disease Recurrence.
   T_1 = First time, from randomization, to earlier event between Treatment failure or Disease Recurrence before week 12
2. $\mathcal{E}_2$ = Death (All causes). T_2 time to Death before week 12
3. T_* time, from randomization, to earlier event between Treatment failure, Disease Recurrence or Death before week 12

Adjunctive rifampicin for *Staphylococcus aureus* bacteraemia (ARREST): a multicentre, randomised, double-blind, placebo-controlled trial

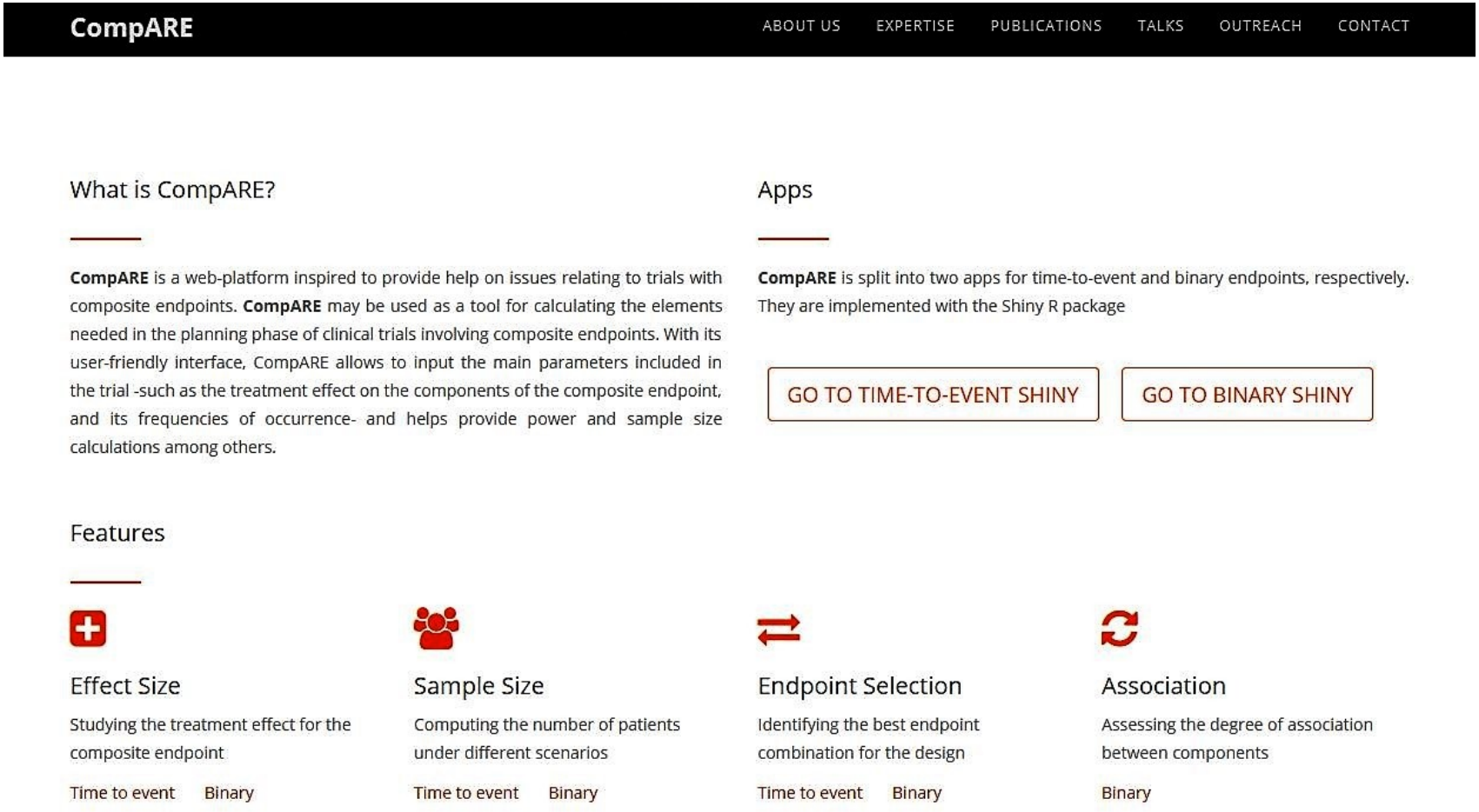
*Guy E Thwaites, Matthew Scarborough, Alexander Szubert, Emmanuel Nsutebu, Robert Tilley, Julia Greig, Sarah A Wyllie, Peter Wilson, Cressida Auckland, Janet Cairns, Denise Ward, Pankaj Lal, Achyut Guleri, Neil Jenkins, Julian Sutton, Martin Wiselka, Gonzalez-Ruiz Armando, Clive Graham, Paul R Chadwick, Gavin Barlow, N Claire Gordon, Bernadette Young, Sarah Meisner, Paul McWhinney, David A Price, David Harvey, Deepa Nayar, Dakshika Jeyaratnam, Tim Planche, Jane Minton, Fleur Hudson, Susan Hopkins, John Williams, M Estee Török, Martin J Llewelyn, Jonathan D Edgeworth, A Sarah Walker, on behalf of the United Kingdom Clinical Infection Research Group (UKCIRG)**

Lancet 2018; 391: 668–78
Published Online
December 14, 2017
http://dx.doi.org/10.1016/S0140-6736(17)32456-X

Summary

Background ***Staphylococcus aureus*** **bacteraemia is a common cause of severe community-acquired and hospital-acquired infection worldwide. We tested the hypothesis that adjunctive rifampicin would reduce bacteriologically confirmed treatment failure or disease recurrence, or death, by enhancing early *S aureus* killing, sterilising infected foci and blood faster, and reducing risks of dissemination and metastatic infection.**

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



Using CompARE to discuss

- **Aspirin Trial** to illustrate Composite **Binary** Endpoints
- **ARREST Trial** to illustrate Com- posite **Time-to-Event** Endpoints

Aspirin trial (1)

HYPOTHESIS: Aspirin prevents TBM-related brain infarction through its anti-thrombotic, anti-inflammatory and pro-resolution properties

- **Population**: Tuberculous meningitis HIV-unifected adults
- **Placebo (0)**: SoC: Antituberculosis drugs and dexamethasone
- **Intervention (1)**: SoC + Aspirin (1000mg)
- **Endpoint (ε_1)**: New brain infarction by 60 days
- **Endpoint (ε_2)**: Death by 60 days
- **Composite endpoint (ε_*)**: Brain infarction or death by 60 days

Phase 2 primary objective was to demonstrate

Safety, tolerability and potential **efficacy** of 81mg and **1000**mg aspirin when added to dexamethasone for the first 60 days of TBM treatment.

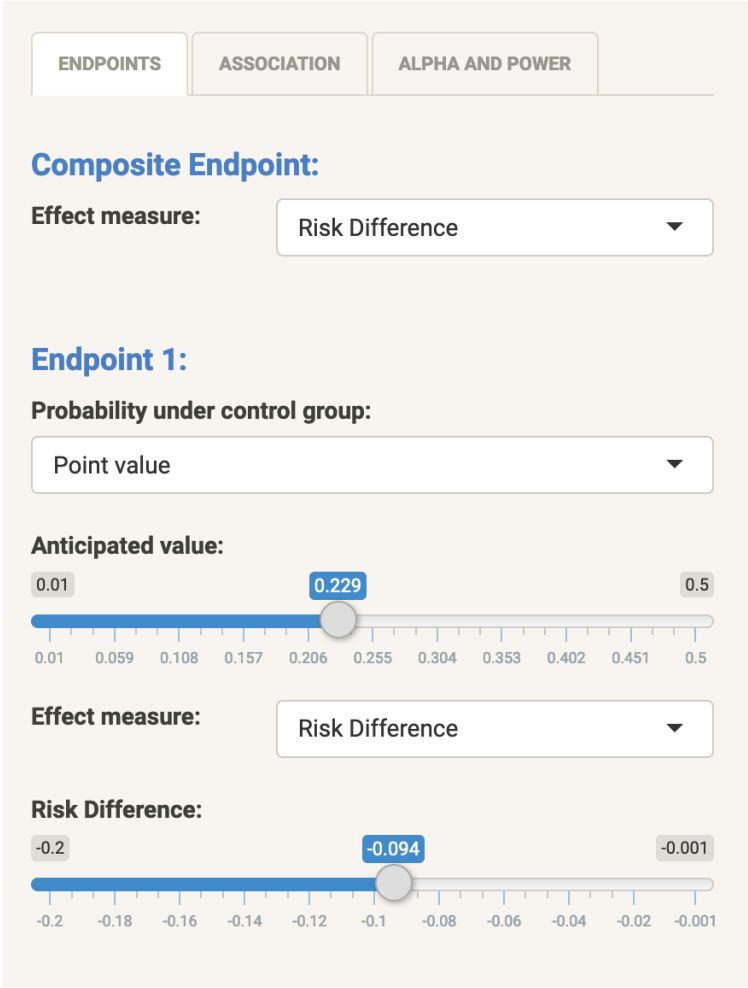
The trial showed:

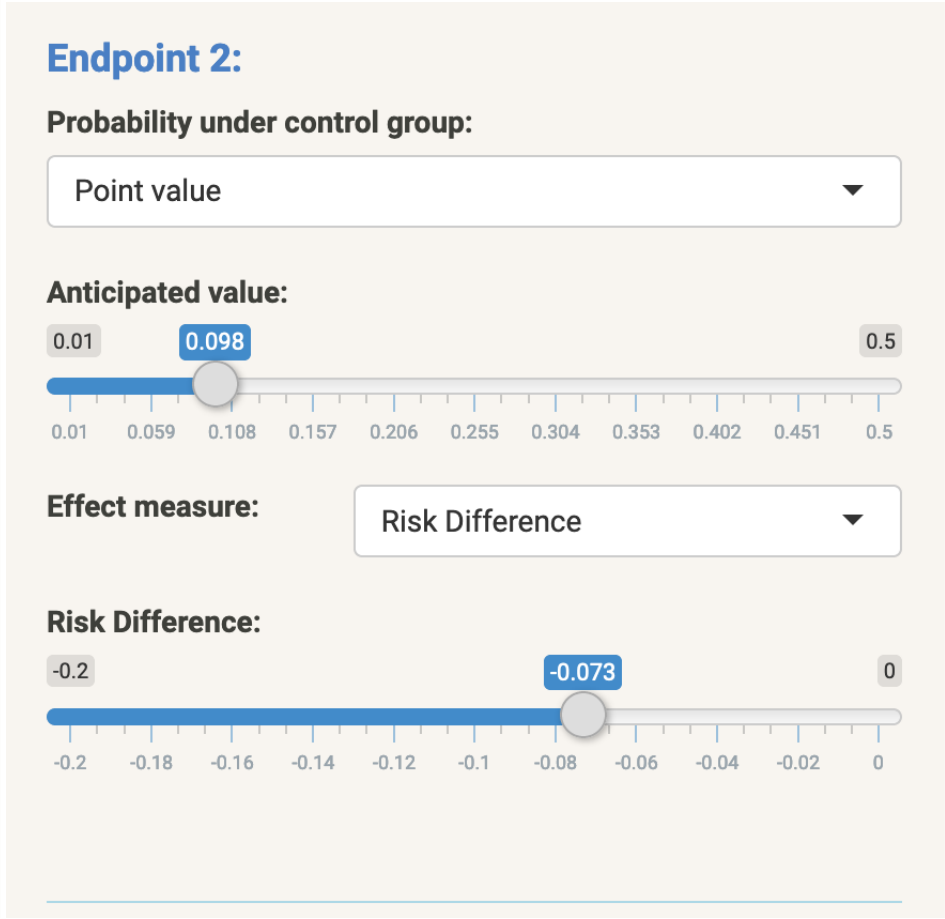
- Daily aspirin can be given safely
- (Non significant) risk reductions of new brain infarction or death (ε_*) in both 81 mg and 1000 mg aspirin versus placebo

Aspirin Trial: Inputs to design a new RCT

Endpoints

Endpoint 1: New brain infarction by 60 days, **Endpoint 2**: Death by 60 days, **CE**: Brain infarction or death by 60 days





Aspirin Trial: Summaries (for Corr = 0.15)

SUMMARY FOR THE COMPOSITE COMPONENTS

Values	Explanation
0.23	Event rate of the Endpoint 1 in control group
0.14	Event rate of the Endpoint 1 in treatment group
0.53	Odds ratio for Endpoint 1
0.59	Risk ratio for Endpoint 1
-0.09	Risk difference for Endpoint 1

Values	Explanation
0.10	Event rate of the Endpoint 2 in control group
0.03	Event rate of the Endpoint 2 in treatment group
0.24	Odds ratio for Endpoint 2
0.26	Risk ratio for Endpoint 2
-0.07	Risk difference for Endpoint 2

SUMMARY FOR THE COMPOSITE ENDPOINT

Values	Explanation
0.29	Event rate of the Composite Endpoint in control group
0.15	Event rate of the Composite Endpoint in treatment group
0.44	Odds ratio for the Composite Endpoint
0.52	Risk ratio for the Composite Endpoint
-0.14	Risk difference for the Composite Endpoint

Aspirin Trial: Effect size

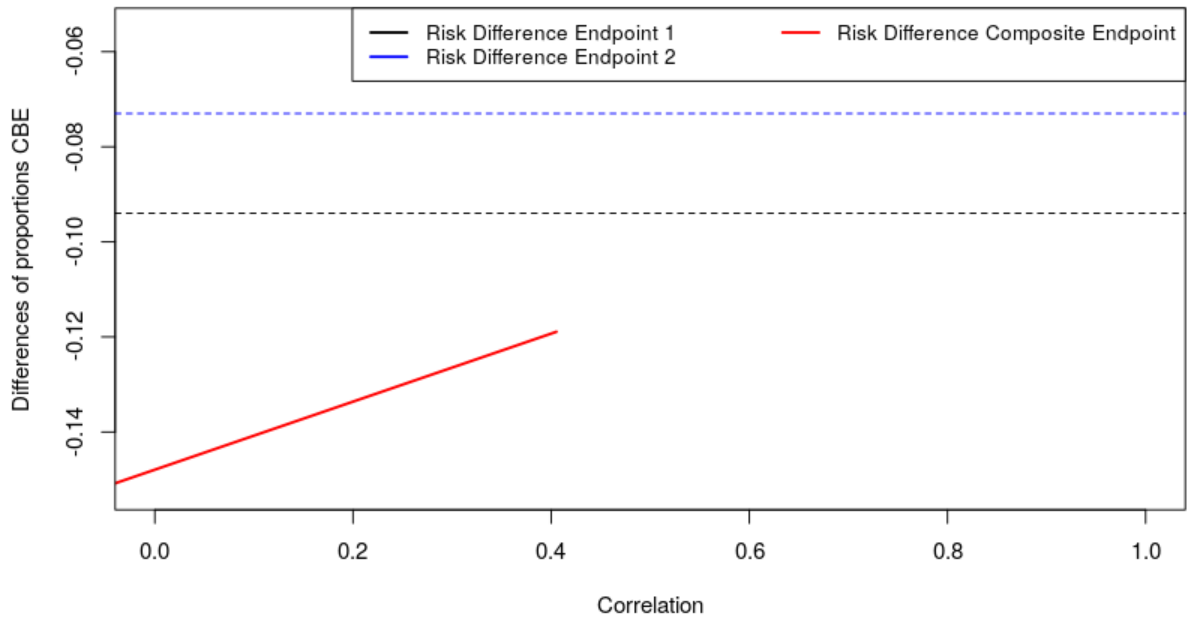
RISK DIFFERENCE WHEN THE CORRELATION VALUE IS NOT KNOWN

Effect size bounds

Lower Bound: -0.152

Upper Bound: -0.119

Risk Difference depending on the correlation



Aspirin Trial: Sample Size

SAMPLE SIZE WHEN THE CORRELATION VALUE IS NOT KNOWN

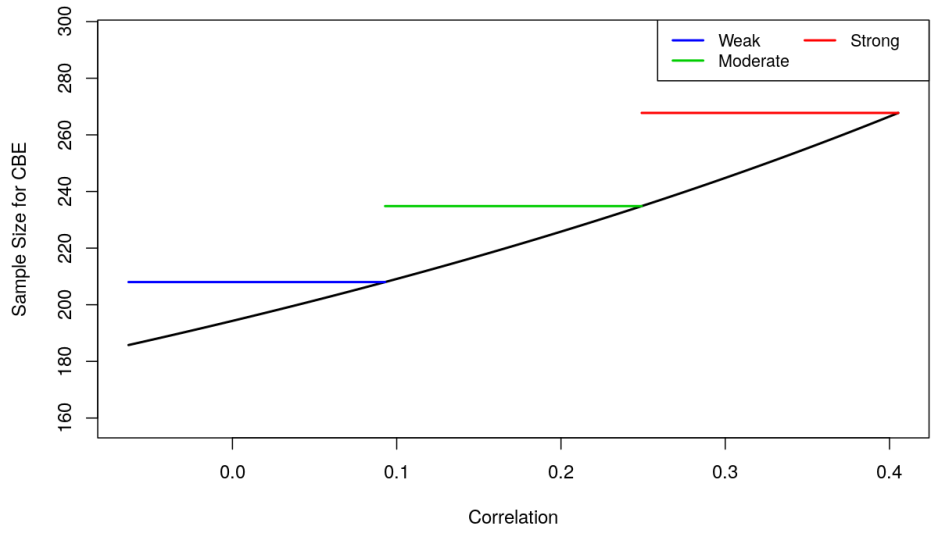
Sample size bounds

Lower Bound: 186

Upper Bound: 268

Sample size according to different correlation categories

Weak	Moderate	Strong
208	235	268



For $\alpha = 0.05$, power$= 0.80$. Sample size is highly sensitive to the association (max correlation$=0.41$):

- $\rho = 0.00 \implies n = 194$
- $\rho = 0.09 \implies n = 208$
- $\rho = 0.15 \implies n = 217$
- $\rho = 0.25 \implies n = 225$
- $\rho = 0.4 \implies n = 266$

Sample size in the RCT was $81 = 41$ (placebo)$+40$(asp.)

Aspirin Trial: Sample Sizes for Prob(Infarction) $\in (0.21, 0.25)$ and Prob(Death)$\in (0.08, 0.12)$ varying correlation between infarction and death

Endpoint 1:

Probability under control group:

Interval plausible values

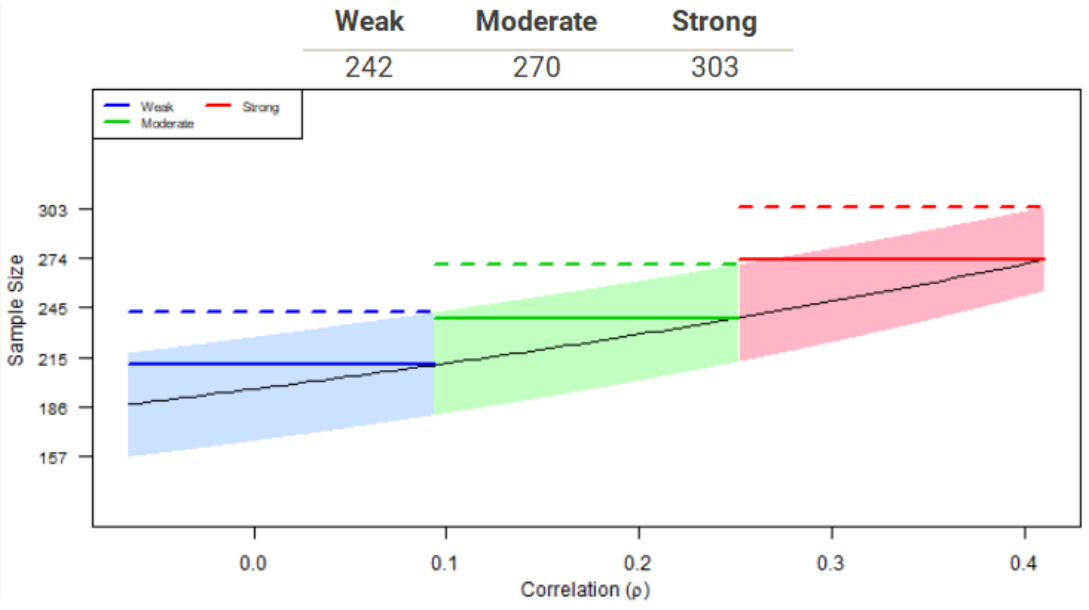
Lower: 0.21 **Upper:** 0.25

Endpoint 2:

Probability under control group:

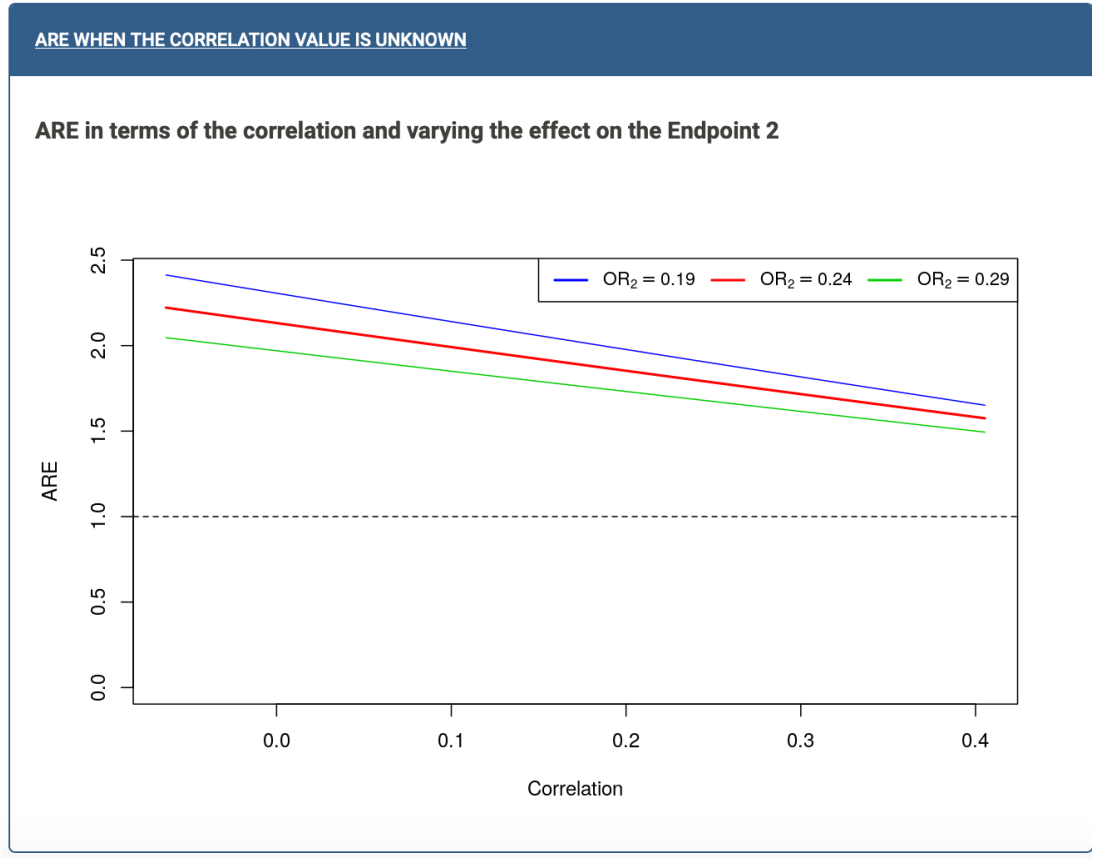
Interval plausible values

Lower: 0.08 **Upper:** 0.12



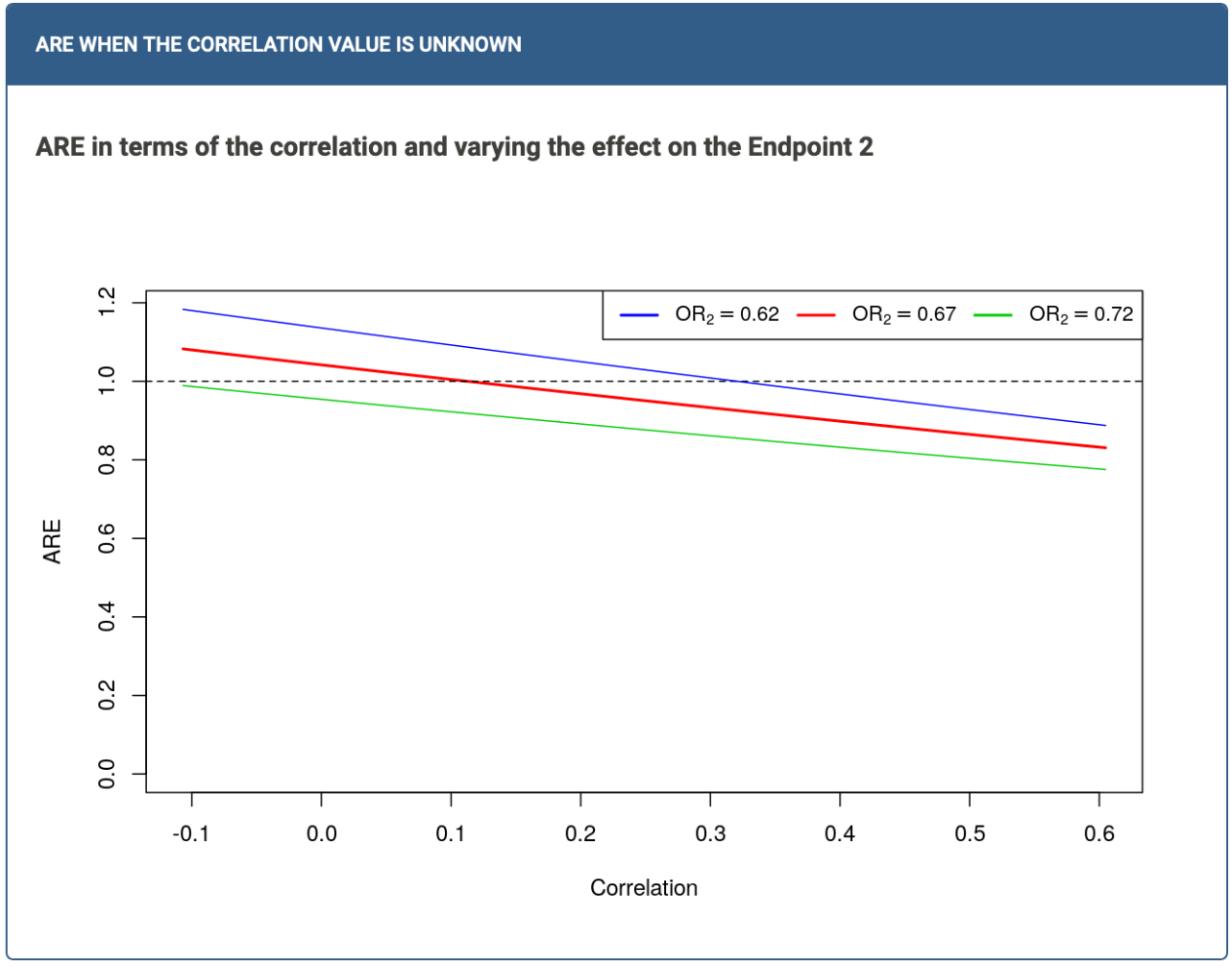
Aspirin Trial: It is **more efficient** to add death to New MRI-proved brain infarction

Risk Difference $\mathcal{E}_2 \in (-0.078, -0.067) \Leftrightarrow OR_2 \in (0.19, 0.29)$



Aspirin Trial: It is **NOT more efficient** to add death to New MRI-proved brain infarction

Risk Difference $\mathcal{E}_2 \in (-0.035, -0.025) \Leftrightarrow OR_2 \in (0.62, 0.72)$



Aspirin Trial: Wrapping up

If a large phase 3 trial of adjunctive aspirin for TBM is planned, the CE of new brain infarction and death is recommended (ARE>1), and based on the phase 2 trial, 226 patients would be needed.

KEY QUESTIONS

How much sample size is needed?

The total sample size (2 groups) that you need for having 0.8 power at significance level 0.05 when using *Endpoint 1* is **410** subjects; when using *Endpoint 2*, you need **262** subjects; and when using the *Composite endpoint*, you need **226** subjects.

What is the expected risk difference for the composite endpoint?

Considering a correlation between the components of **0.2**; the effect on the *Composite endpoint* is **-0.134**. In case that the correlation value is not known, the effect ranges between **-0.152** and **-0.119**.

How much correlation can you expect between the components?

The correlation between *Endpoint 1* and *Endpoint 2* could take values between **-0.06** and **0.41**.

Is the use of the Composite Endpoint more efficient than the use of Endpoint 1?

Yes, because the value of the ARE (**1.85**) is greater than 1.

ARREST TRIAL (2)

HYPOTHESIS: Adjunctive rifampicin would reduce bacteriologically confirmed treatment failure or disease recurrence or death by enhancing early *S aureus* killing, sterilising infected foci and blood faster and reducing the risk of disseminations and metastatic infection

- **Population**: Adults with *SA* bacteraemia who had received $\leq 96h$ of active antibiotic therapy
- **Placebo (0)**: Standard antibiotic therapy (SoC)
- **Intervention (1)**: Adjunctive rifampicin
- **Outcome (ε_1)**: Bacteriologically confirmed treatment failure or disease recurrence by 12 weeks
- **Outcome (ε_2)**: All cause deaths by 12 weeks
- **Composite outcome (ε_*)**: Bacteriologically confirmed treatment failure or disease recurrence or death (all cause) by 12 weeks

(2) Guy E Thwaites *et al*. Lancet 2018: 301:668-78

ARREST trial: Time-to-event framework

- **Endpoint (T_1)**: Time to first among treatment failure or disease recurrence before 12 weeks
- **Endpoint (T_2)**: Time to death before 12 weeks
- **Composite endpoint (T_*)**: Time to first among treatment failure, disease recurrence or death (all cause) by 12 weeks

Phase 3 RCT primary objective was to demonstrate

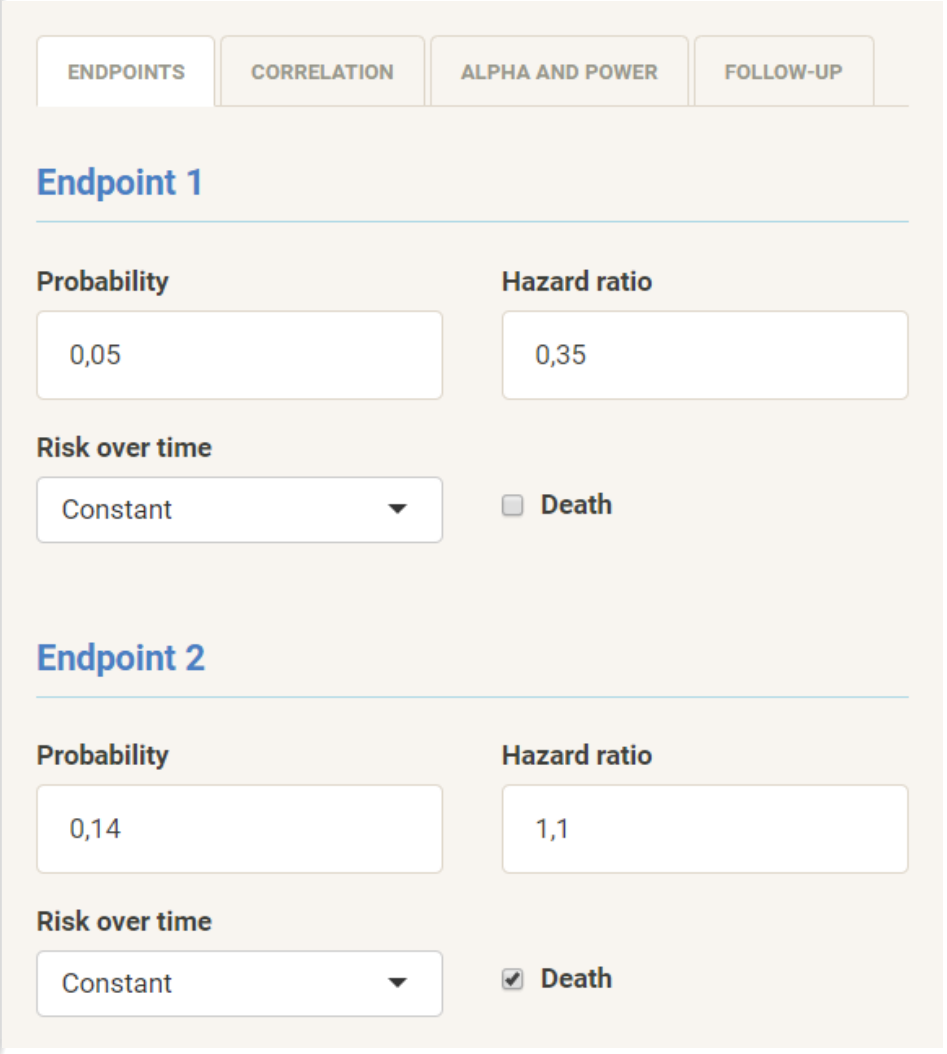
Efficacy of adjunctive rifampicin in reducing bacteriologically confirmed treatment failure or disease recurrence or death. The trial showed:

- Non significant reduction on *SA* related events, (T_*)
- Small significant reduction in bacteriologically and clinically defined disease recurrences

ARREST: Input Parameters

Endpoints

- **Endpoint 1**: Time to first among treatment failure or recurrence < 12 wks.
  $\mathcal{E}_1$ event rate (group 0)=0.054 and $\mathrm{HR}_1 = 0.35$
- **Endpoint 2**: Time to (all cause) death < 12 wks.
  $\mathcal{E}_2$ event rate (group 0)=0.183 and $\mathrm{HR}_2 = 1.1$
- **CE**: Time to first among *SA* related failures and all causes death < 12 wks.



ARREST: Input ALL Parameters

ENDPOINTS | CORRELATION | ALPHA AND POWER | FOLLOW-UP

Endpoint 1

Probability: 0,05

Hazard ratio: 0,35

Risk over time: Constant

☐ **Death**

Endpoint 2

Probability: 0,14

Hazard ratio: 1,1

Risk over time: Constant

☑ **Death**

ENDPOINTS | CORRELATION | ALPHA AND POWER | FOLLOW-UP

Relationship between endpoints

Correlation: 0,85

Type: Spearman's rho

Copula: Frank

Alpha and Power

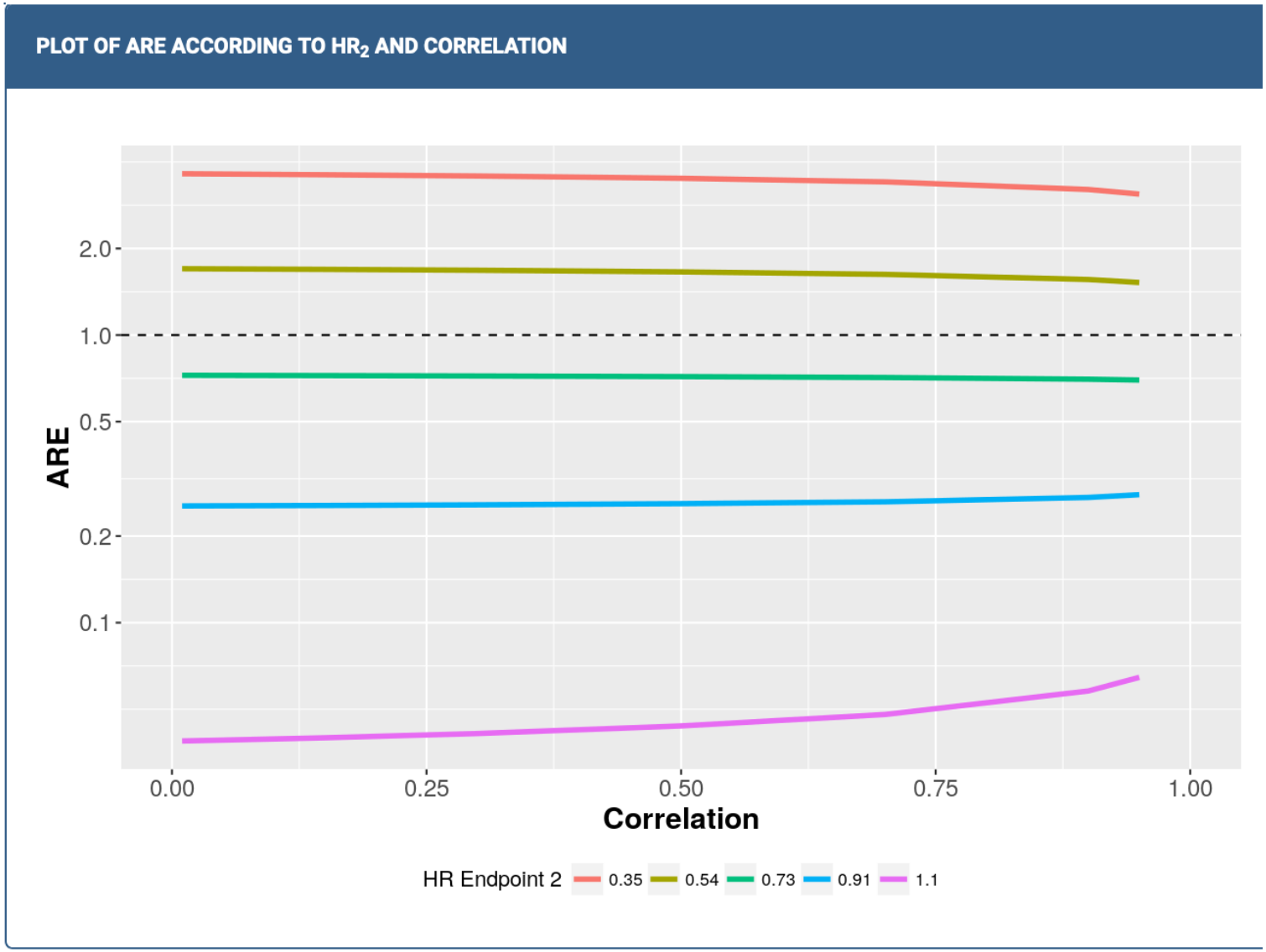
Significance level: 0,05

Power: 0,8

Formula: Schoenfeld

ARREST: Adding Death to *SA* related events

Fixed values in this plot: Prob (failure or recurrence|SoC)=0.05; Prob (death|SoC)=0.18; effect size for failure or recurrence = $\mathrm{HR}_1 = 0.35$



ARREST: Wrapping up

- I have been using
  - **Endpoint 1**: Time to first among treatment failure and disease recurrence
  - **Endpoint 2**: Time to (all cause) death
  - **CE**: Time to first among *SA* related failures and all causes death

  but perhaps other combinations could have been meaningful.
- Based on those, the needed correlation between *SA* related failures and all cause death is 0.85 for a $HR_* = 0.96$ as the one observed. In this case adding all cause of death to *SA* related failures would have not been recommended.
- For stronger effect sizes on mortality ($HR_2 < 0.7$) the CE would have been advised

SUMMARIZING

When planning a CT, the decision of whether or not to use a composite endpoint has to be based, other than clinical relevance, on a careful study of the anticipated component values:

- the probabilities of observing the events in the control group,
- the effect size for each component
- the association between the two events for each group. Seldomly reported and not easy to guess!!
- on the corresponding efficiency (ARE)

If your study involves several outcomes, you are interested in their union and you need to know:

- Probability of occurrence of their union
- Odds Ratio of the CE
- Survival and hazard functions for time to the first event

DO NOT HESITATE AND USE CompARE

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

This work is partially supported by MTM2015-64465-C2-1-R and MDM-2014-0445 from Spanish Ministry of Economy and Competitiveness (Spain) and 2017 SGR 622 (GRBIO) from Generalitat de Catalunya