

# We describe methodology to **benchmark** time-to-event outcomes producing **fair comparisons** between **hierarchical units**, such as trials being combined in IPD meta-analysis.

## Standardised Contrasts Across Trials in Individual Patient Data Meta-Analysis With Time-to-Event Outcomes

Alessandro Gasparini<sup>1\*</sup>, Michael J. Crowther<sup>1</sup>, Justin M. Schaffer<sup>2</sup>



### Objectives

Data from randomised controlled clinical trials with a time-to-event outcome is often combined in individual patient data meta-analyses (IPD-MA) to summarise the overall treatment effect. When the goal is to fairly benchmark and compare different trials, adjusting for observed and unobserved heterogeneity is key. To this end, we extend previous work to time-to-event outcomes to allow obtaining standardised survival curves and contrasts.

### Methods

We assume a two-levels proportional hazards survival model for an IPD-MA, where  $i$  denotes subjects and  $j$  trials:

$$h(t_{ij}) = h_0(t_{ij})e^{X_{ij}\beta + b_j}$$

with fixed-effects covariates  $X$ , regression coefficients  $\beta$ , random intercept  $b_j$ , baseline hazard  $h_0$ . Other classes of multilevel survival models (e.g., accelerated failure time) could be used as well. Then, we define the case-mix adjusted counterfactual survival probability, at time  $t$ , for the  $j^{\text{th}}$  trial:

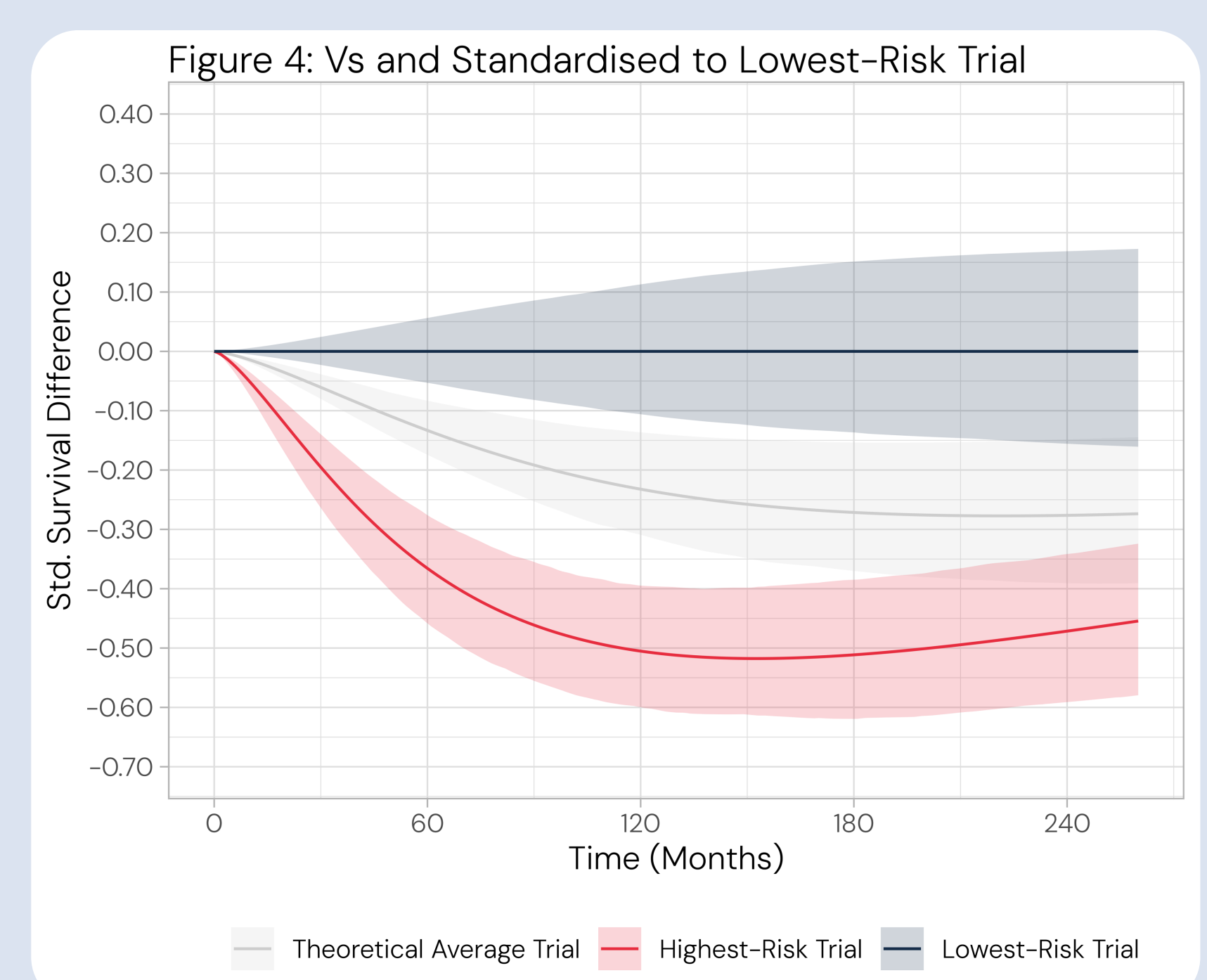
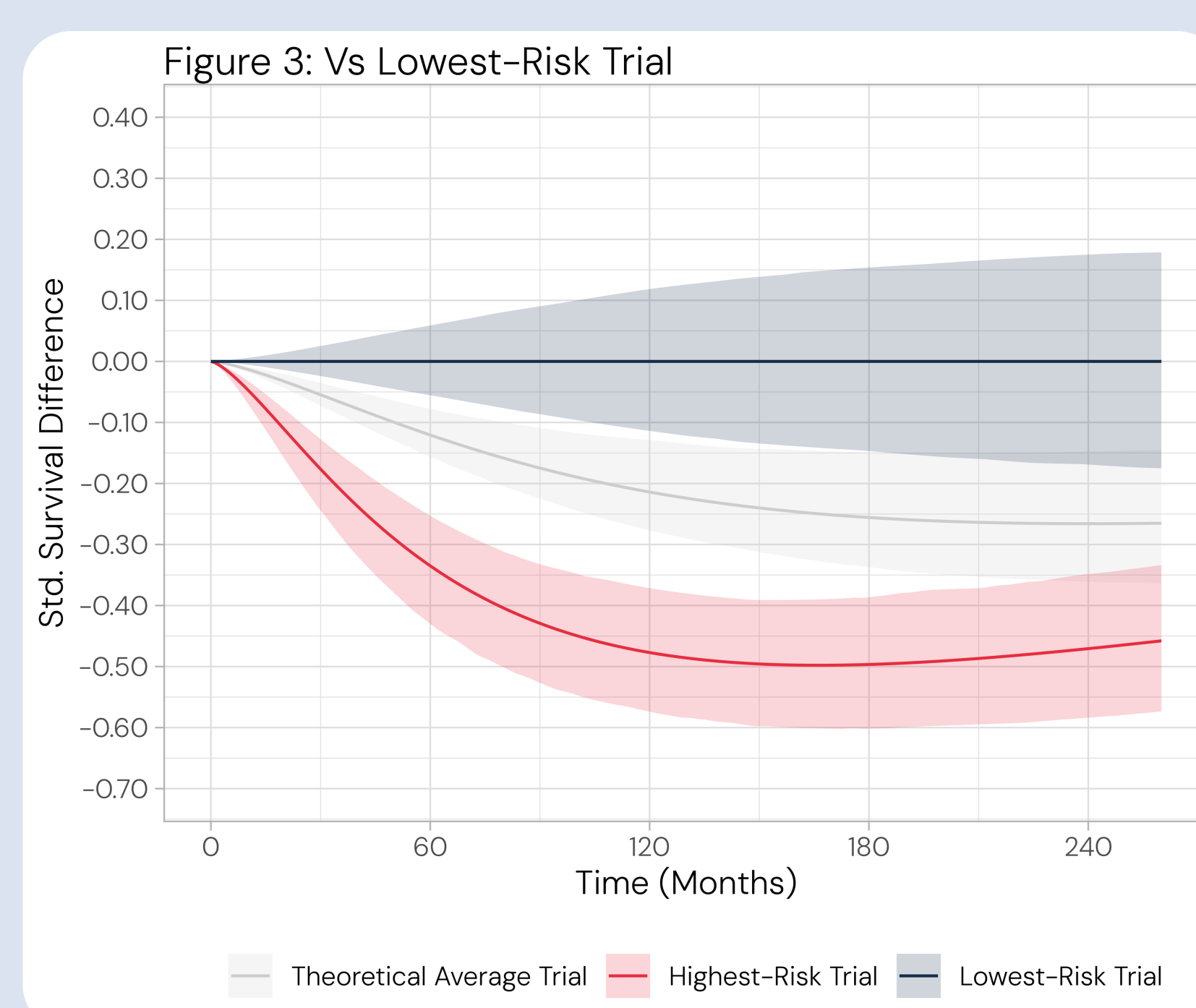
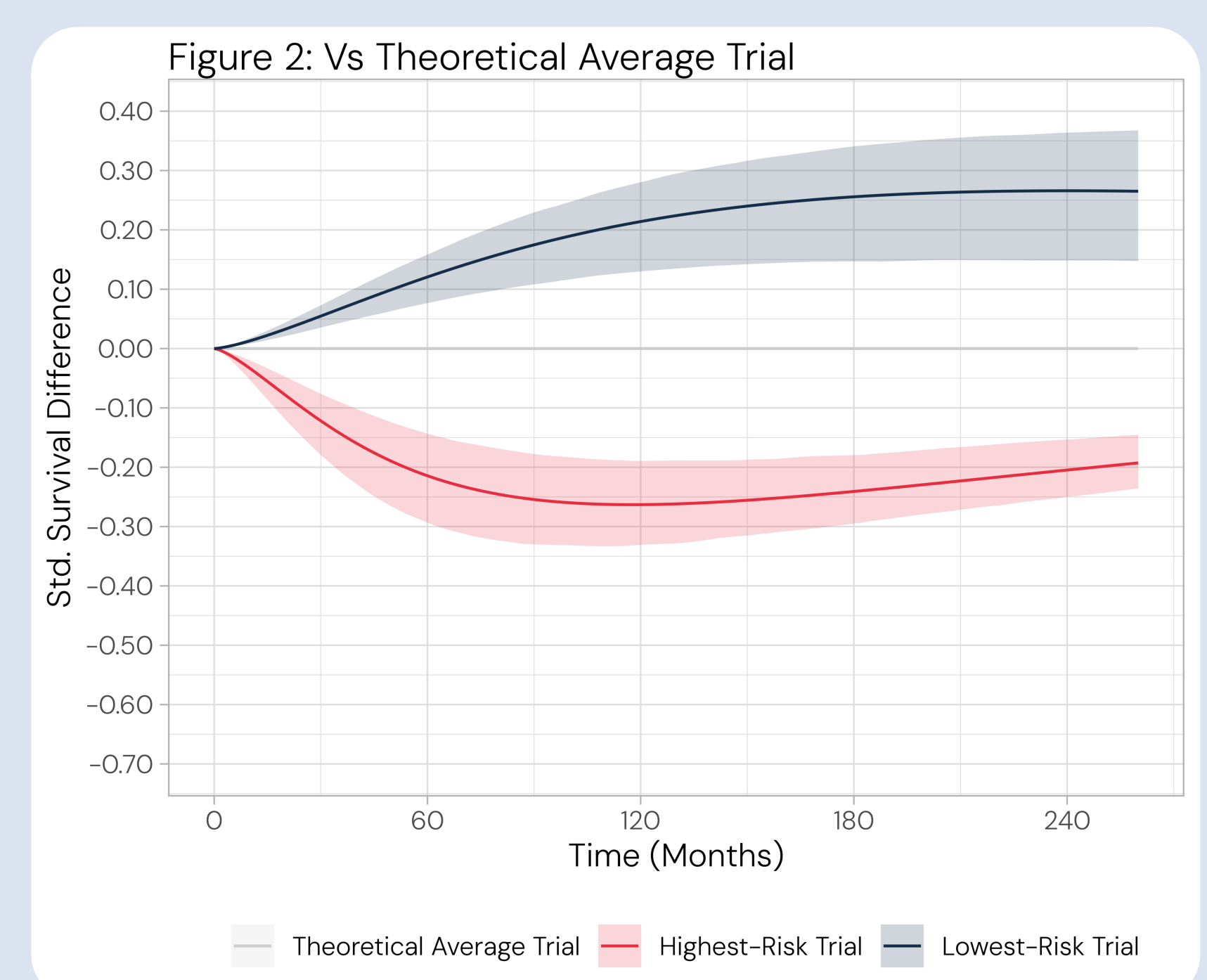
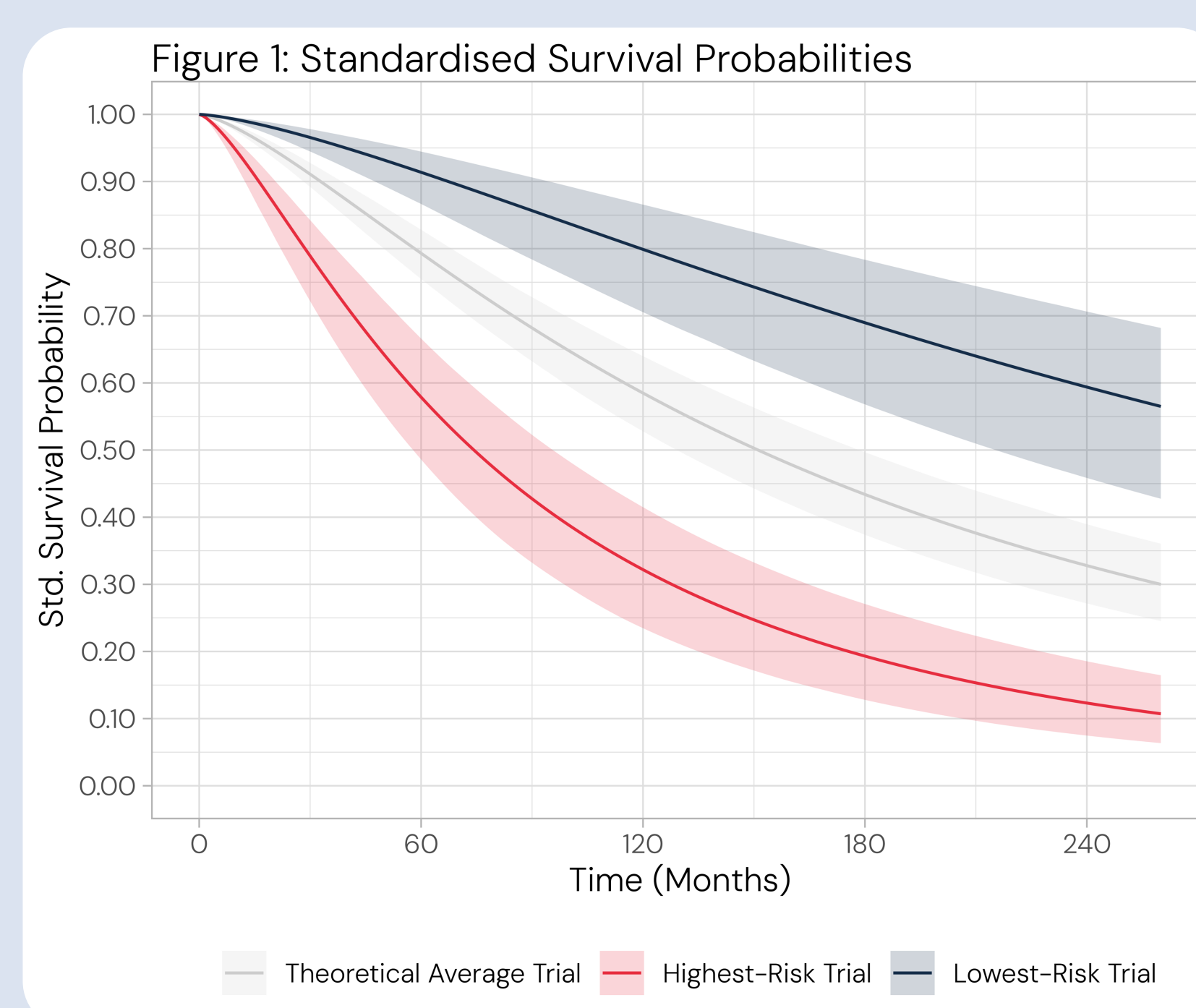
$$S^j(t) = E[S(t | X, b = b_j^*)]$$

with the expectation taken over the fixed effects  $X$ . The value  $b^*$  denotes best linear unbiased predictions (BLUPs) for the random effects, which is fixed in the above equation. This quantity can be interpreted as the *counterfactual survival probability if the entire IPD-MA population was exposed to the performance of the  $j^{\text{th}}$  trial. We can estimate this quantity using regression standardisation, where  $N$  denotes total sample size:*

$$\hat{S}^j(t) = 1/N \sum_{ij} \hat{S}(t | X_{ij}, b = b_j^*)$$

We can extend this further by manipulating the reference population to standardise against ( $X$ ) or by defining contrasts (e.g., comparing  $j^{\text{th}}$  vs  $k^{\text{th}}$  hierarchical unit). If adjusting for case-mix variables is sufficient to control for confounding, and under usual causal inference assumptions, these contrasts of standardised survival probabilities have a causal interpretation.

We illustrate this methodology using a publicly available subset of the 3CIA database of COPD patients. We fit a multilevel Weibull survival model with a cohort-level random intercept, adjusting for age, forced expiratory volume in 1 second (FEV1), dyspnoea score, and calculate several standardised measures.



### Results

Figure 1 depicts case-mix adjusted, standardised survival probability for a subset of trials: the lowest-risk, highest-risk, and theoretical average trial ( $b^*=0$ ). These are fairly and directly comparable because we standardised over the case-mix distribution of the entire IPD-MA population.

If we contrast this with the theoretical average trial or the trial with the lowest risk, we obtain Figures 2 and 3, respectively. As expected, the lowest-risk trial has consistently above-average (better) survival and the highest-risk has below-average (worse) survival, up to approximately 25% and -20% at 20 years of follow-up, respectively. Standardising to the case-mix covariates distribution of the lowest-risk trial (Figure 4) did not significantly change the results.

### Conclusion

This methodology can be used to compare performance across trials in IPD-MA or hierarchical modelling settings. Under usual assumptions in causal inference, standardised survival probability differences (e.g., in Figures 2, 3, 4) can be interpreted as causal risk differences.

1. Red Door Analytics AB, Stockholm, Sweden

2. Department of Cardiothoracic Surgery, Baylor Scott & White, The Heart Hospital, Plano, Texas, USA

\* E-mail: <alessandro.gasparini@reddooranalytics.se>