Seeking Harmony: Estimands and Sensitivity Analyses for Randomized Clinical Trials

Devan V. Mehrotra*
Biostatistics and Research Decision Sciences

* devan_mehrotra@merck.com

ASA-BIOP Regulatory-Industry Statistics Workshop
Washington DC, 9-13-18

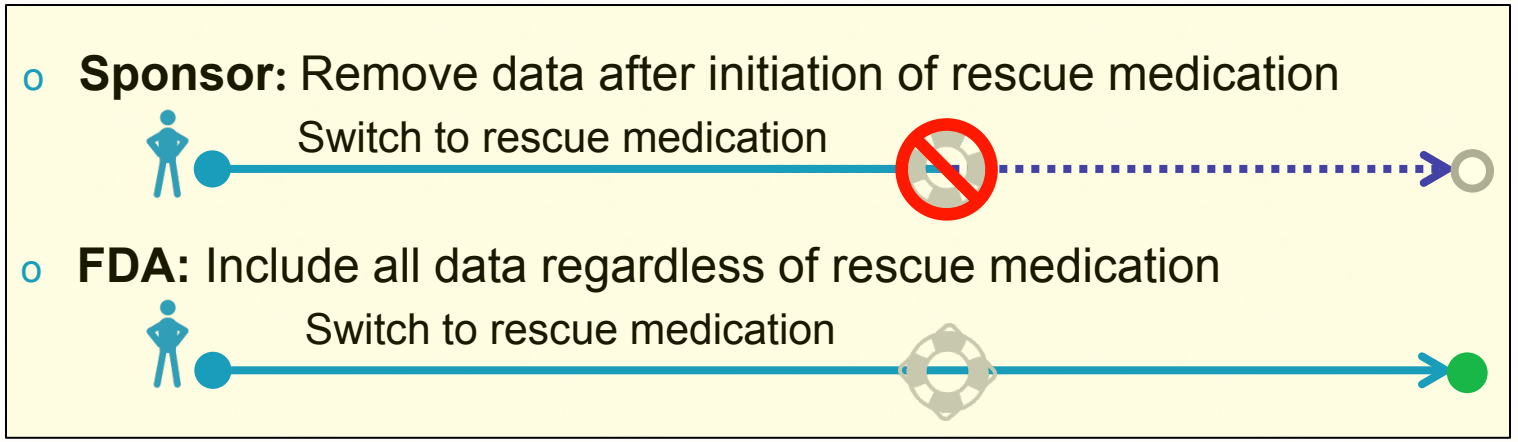
Outline

- ICH E9/R1: why and what?
- Continuous endpoint
  - Case study (diabetes)
  - Case study (depression)
- Time-to-event endpoint
  - Hazard ratio estimands
  - Example
- Conclusions

ICH E9/R1: Why?

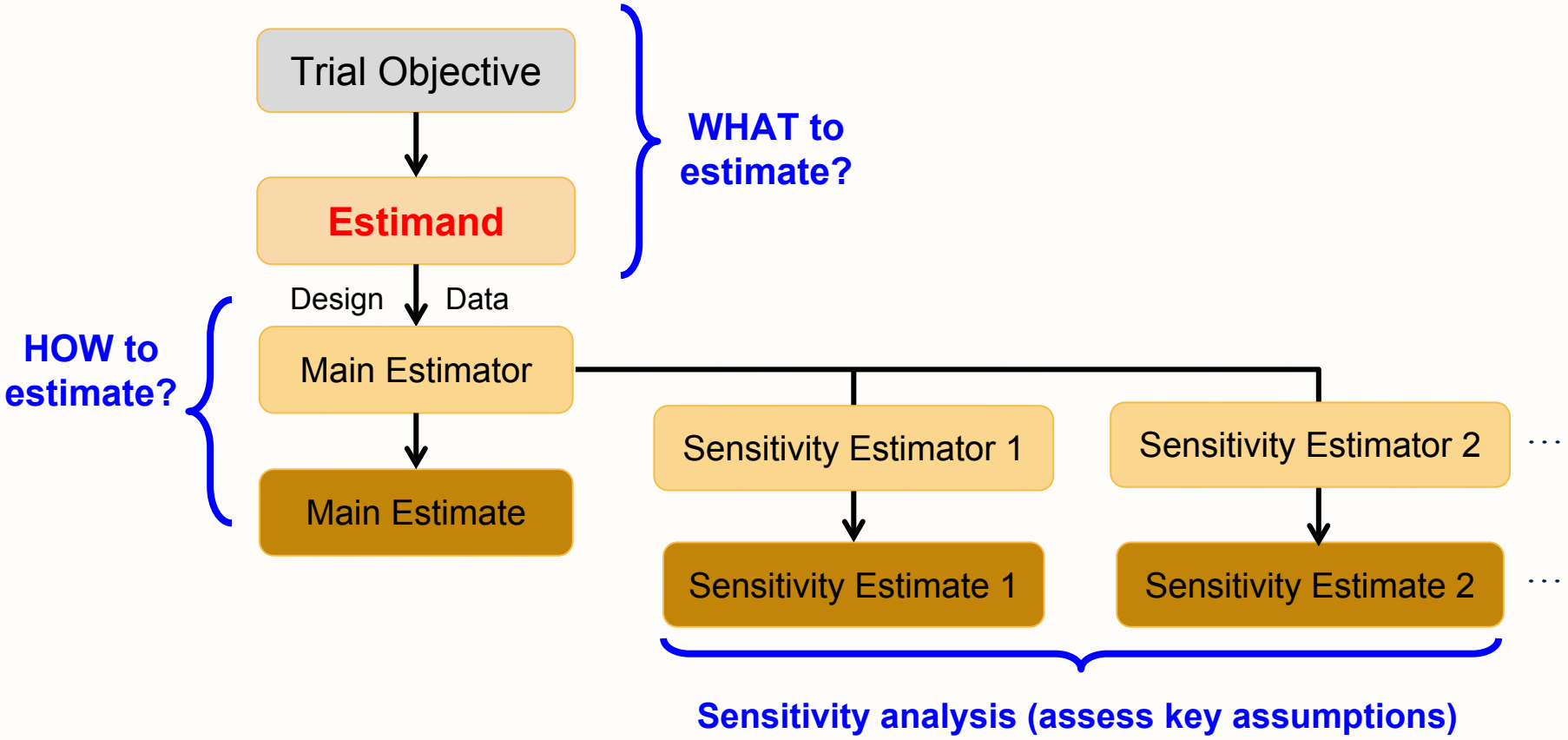
- **Feedback from regulatory statisticians** on several clinical trial protocols & new drug applications:
  - Insufficient clarity in objectives and related treatment effect parameters (i.e., estimands) of interest
  - Lack of logical connectivity between trial objectives, design, conduct, analysis and interpretation
  - Misalignment between "missing data" analysis methods and estimands of interest

2011 FDA advisory committee for dapagliflozin



ICH E9/R1: A Structured Framework

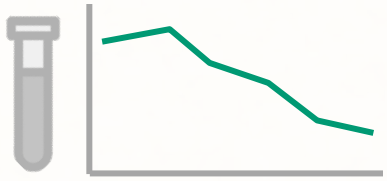
For a given trial objective: aligning target of estimation, design, method of estimation and sensitivity analysis



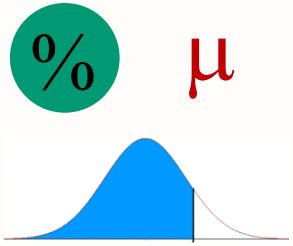
ICH E9/R1: Inputs for Defining an Estimand

A
Population

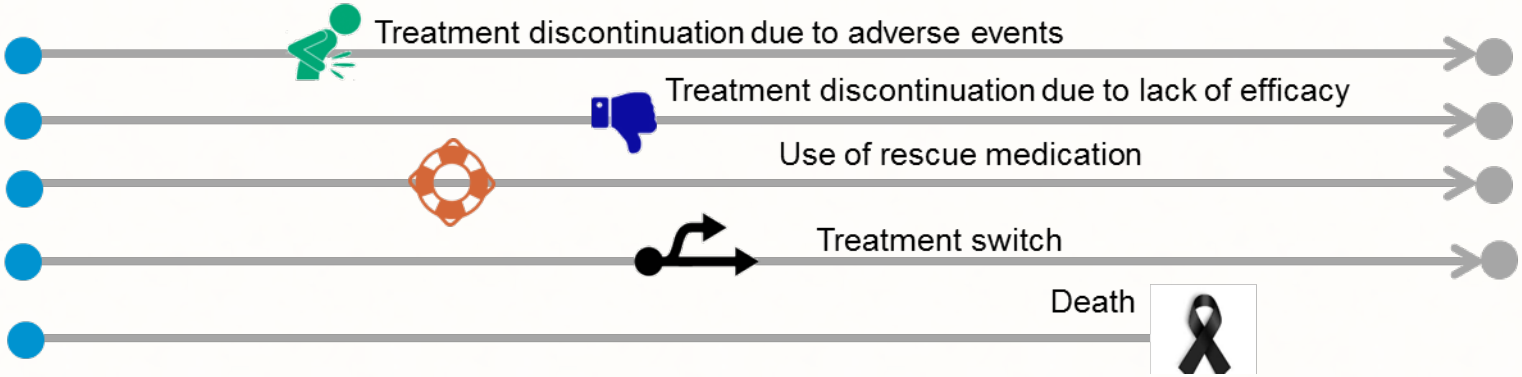
B
Variable (Endpoint)



D
Population-level
Endpoint Summary



C
Intercurrent Event(s)



Case Study: Diabetes (HbA1c; asymptomatic)

Primary Objective: assess whether drug is more effective than placebo in lowering HbA1c without rescue medication (the latter was allowed, but to address a different objective)

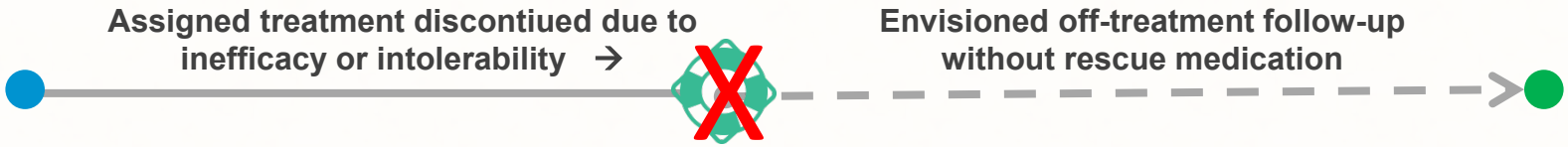
Construction of Primary Estimand

Population: adults with type II diabetes (per intended label)

Endpoint: HbA1c change from baseline at 24 weeks

Intercurrent event: rescue medication

Treatment effect of interest is based on envisioned endpoint under complete follow-up even if the assigned treatment is discontinued, but without rescue medication at any time



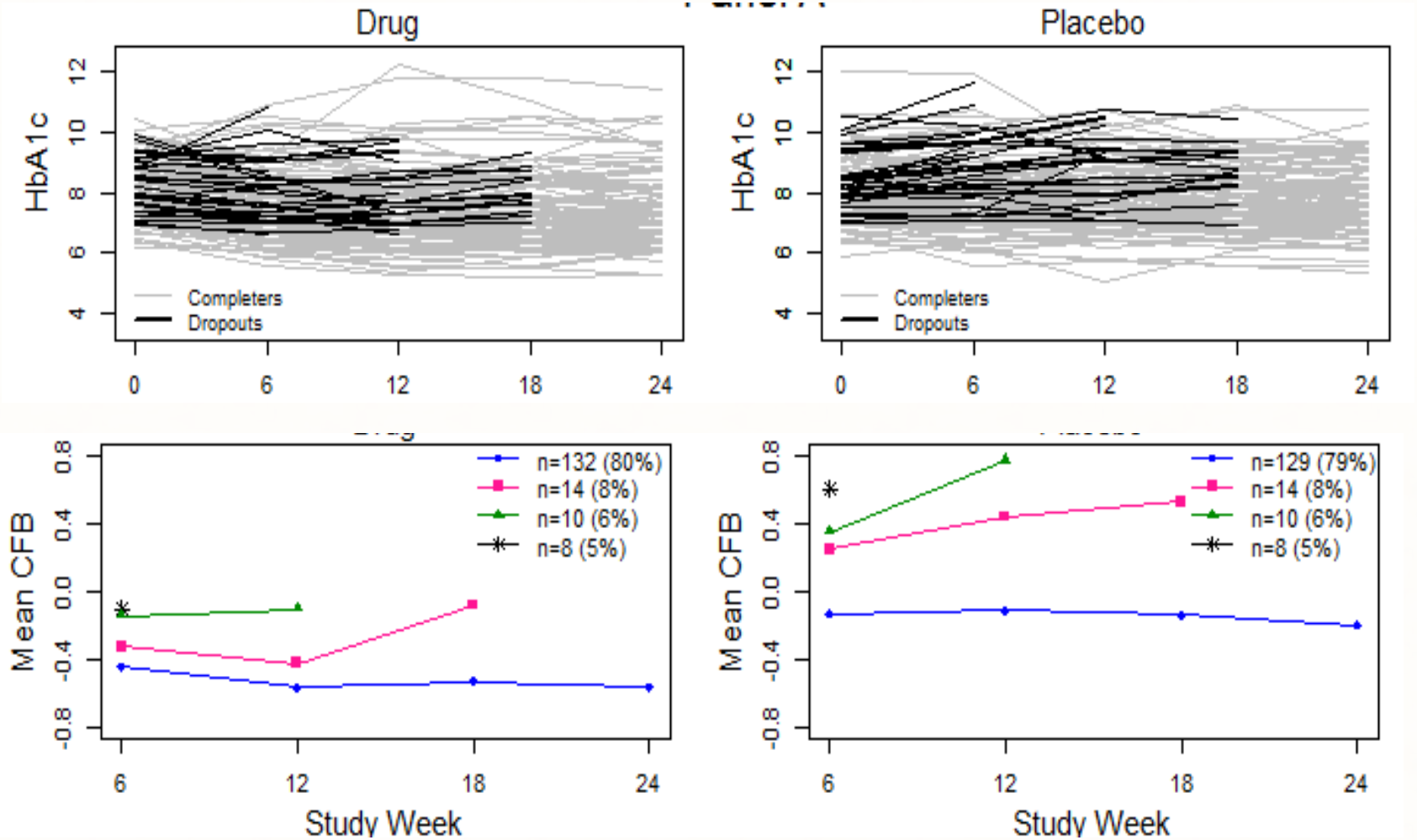
Population-level summary: mean of the endpoint

Case Study (continued)

- **Estimand**: between-treatment difference in target population endpoint means for the treatment effect of interest (δ)
- **Statistical objectives**
  - Deliver acceptable point estimate and 95% CI for δ
  - Test H_{null}: $\delta=0$ vs. H_{alt}: $\delta<0$ (with type 1 error rate $\leq \alpha$)
- **Tackling rescue medication in the analysis**
  - Given estimand of interest, HbA1c values after initiation of rescue medication can be discarded, resulting in "missing" endpoint data for such patients (and dropouts)

- **Analysis challenge**: all randomized patients need to be included in the analysis (per the estimand), so how do we tackle the missing endpoint data problem?

Case Study (continued)



% missing endpoint: **20%** (33/165) Drug **21%** (34/164) Placebo

1 patient assigned to drug and 2 patients assigned to placebo were dropouts before week 6

Case Study (continued)

Control-based mean imputation is one of many options that can be considered in the pre-specified analysis plan

- Obs = endpoint observed, miss = endpoint missing
- π_i^{miss} = true Pr(endpoint missing under trt i) = $1 - \pi_i^{obs}$

Placebo	***Drug***
$\mu_P = \pi_P^{obs}\,\mu_P^{obs} + \pi_P^{miss}\,\mu_P^{miss}$	$\mu_D = \pi_D^{obs}\,\mu_D^{obs} + \pi_D^{miss}\,\mu_D^{miss}$
$\hat{\mu}_P = \hat{\pi}_P^{obs}\,\hat{\mu}_P^{obs} + \hat{\pi}_P^{miss}\,\hat{\mu}_P^{miss}$	$\hat{\mu}_D[c] = \hat{\pi}_D^{obs}\,\hat{\mu}_D^{obs} + \hat{\pi}_D^{miss}\,(\boldsymbol{\hat{\mu}_P + c})$

$\hat{\mu}_P^{miss}$ = estimate of μ_P assuming missing endpoints are MAR for placebo

Estimand: $\delta = \mu_D - \mu_P$

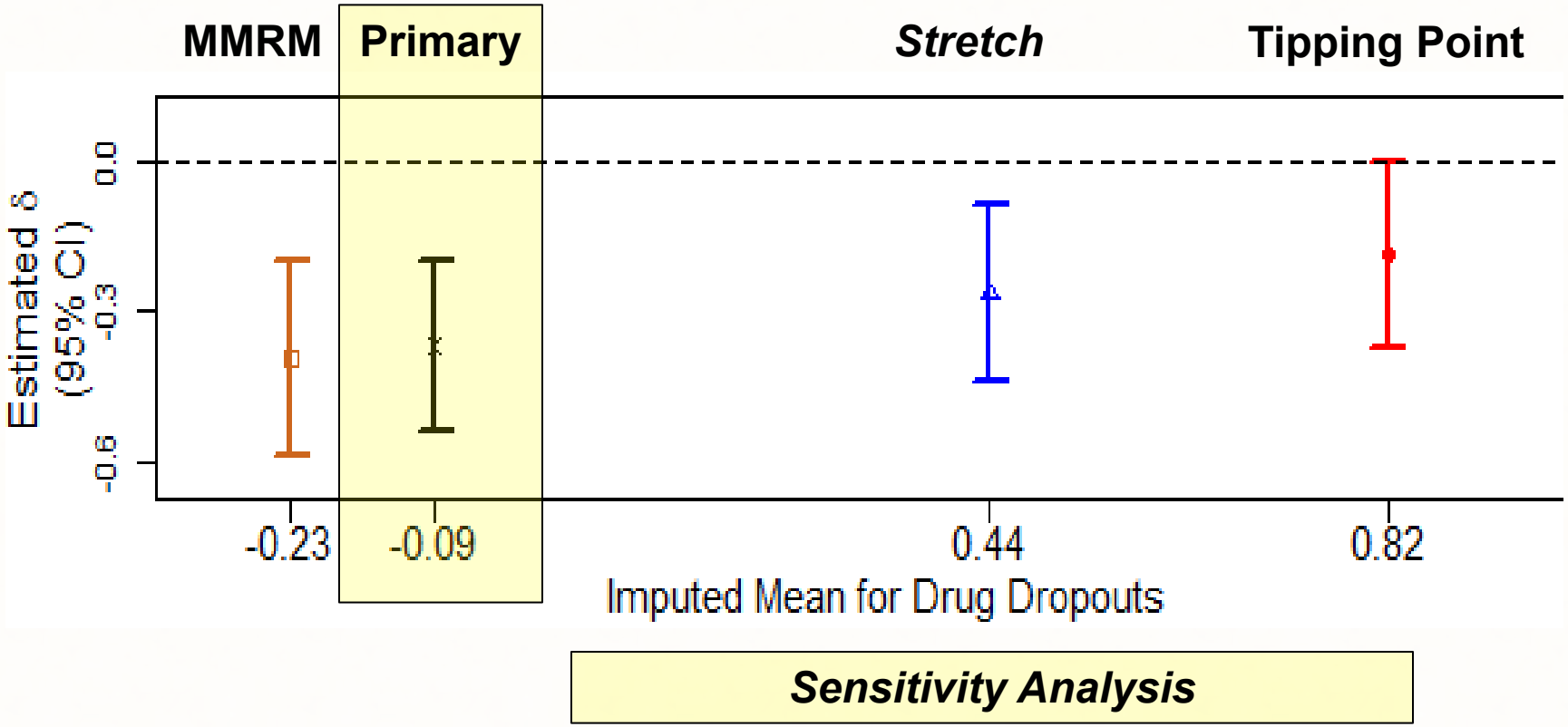
Estimation: $\hat{\delta}[c] = \hat{\mu}_D[c] - \hat{\mu}_P$

Primary Analysis: use $c = 0$

***Sensitivity Analysis*:** increase c until **Tipping Point** is reached

Details: Mehrotra, Liu, Permutt (2017; Pharmaceutical Statistics)

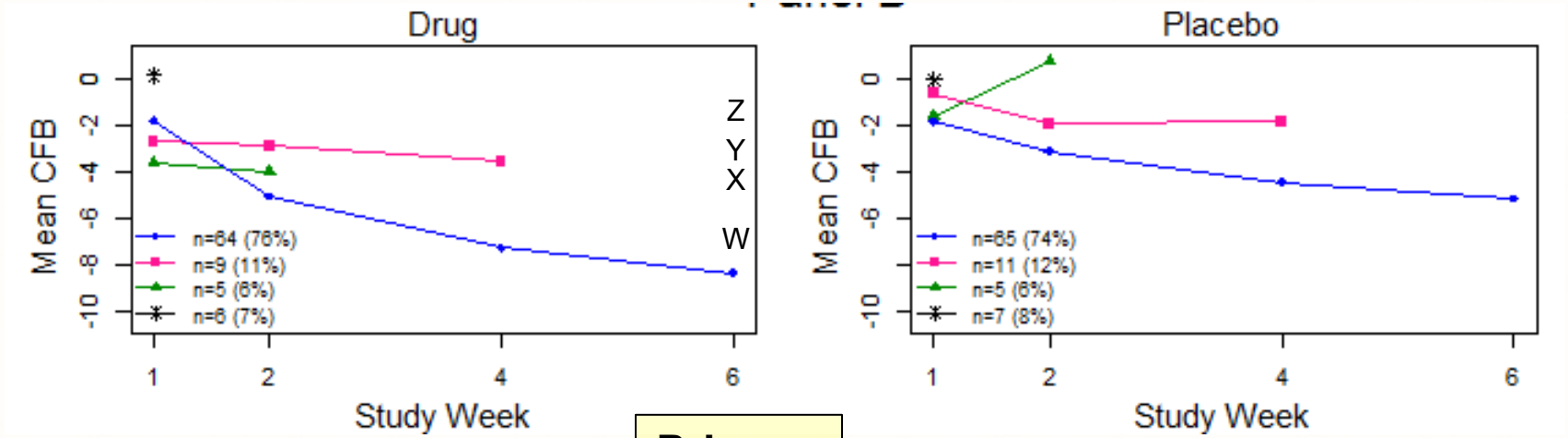
Case Study (continued)

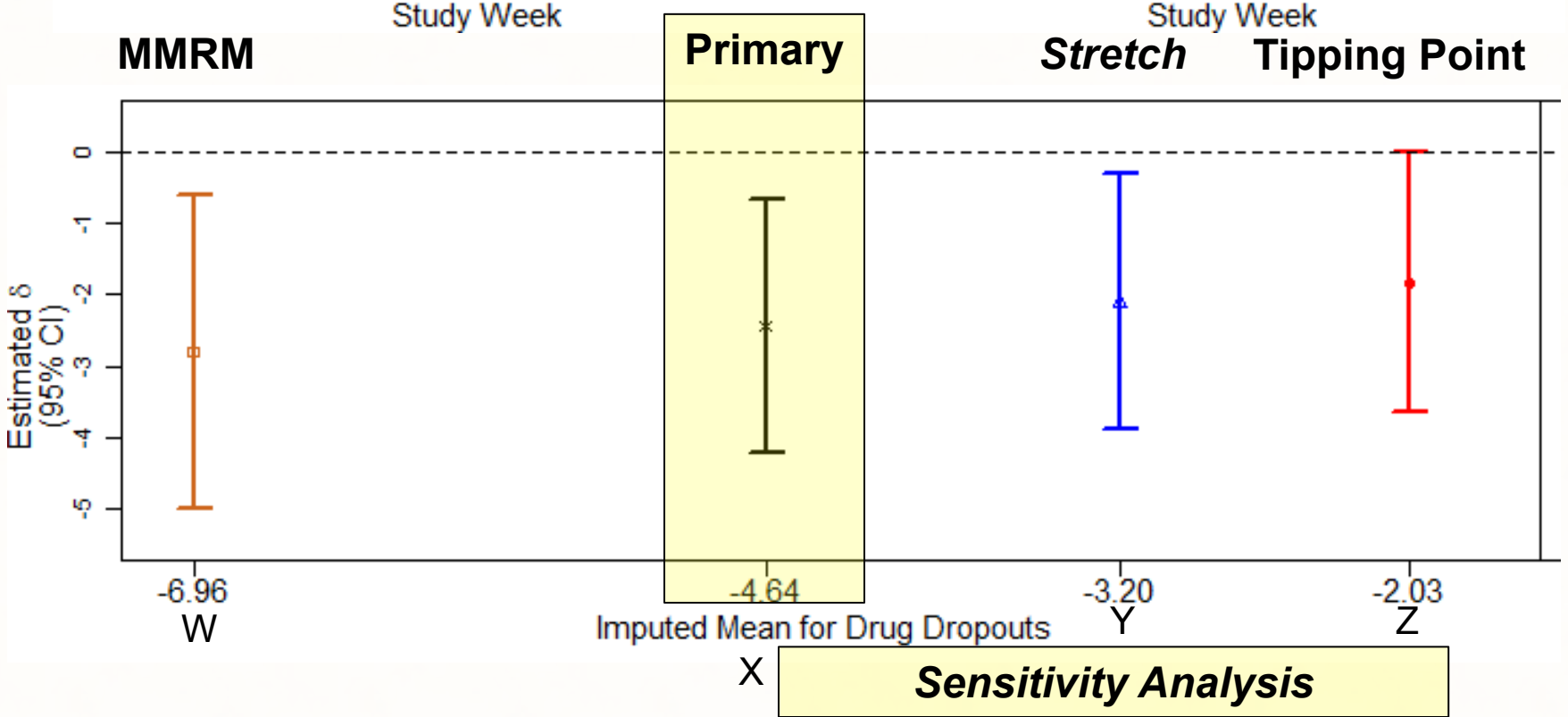


- MMRM: mixed model repeated measures assuming MAR dropout for drug and placebo [shown for historical reference only]
- *Stretch*: imputed mean for drug dropouts matches estimated mean for placebo dropouts assuming MAR dropout for placebo
- ***Note***: tipping point after *stretch* imputation ⇒ **robust** evidence of drug effect

Case Study #2: Depression (HAMD17; symptomatic)

% missing endpoint: **24%** (20/84) Drug, **26%** (23/88) Placebo





HR Estimands for Time-to-Event Endpoints

Estimand for a homogeneous population

- $S_j(t) =$ true proportion event-free at time t under trt j

  Under proportional hazards: $\theta(t) = \frac{\log\{S_A(t)\}}{\log\{S_B(t)\}} = \theta$ for all t

  θ is the time-invariant hazard ratio under PH

Estimand for a mixture of homogeneous subpopulations

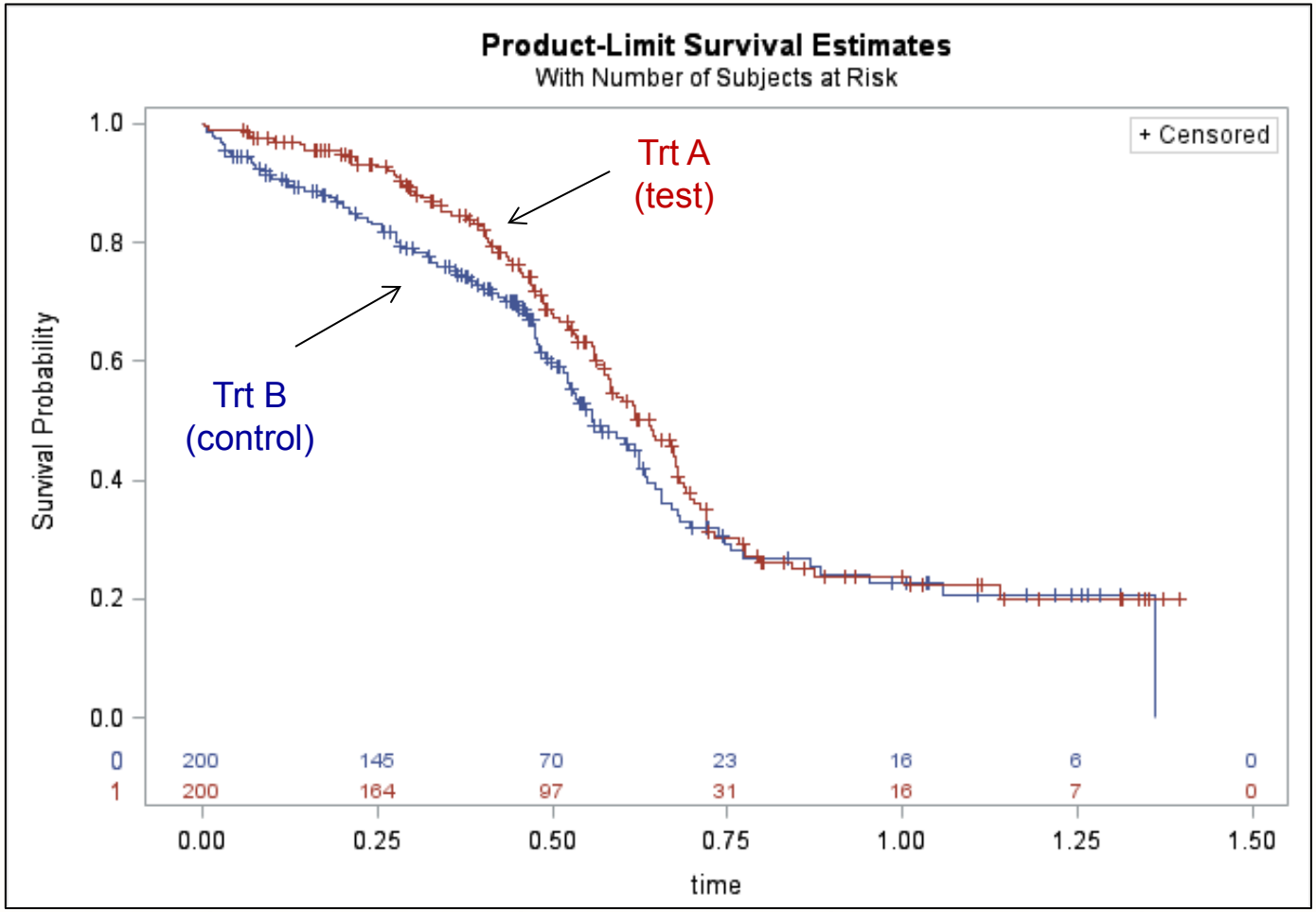
- Assume PH holds within each subpopulation ("stratum")

  Define $\bar{\beta} = \sum_j f_j \beta_j$

  f_j = prevalence and $\beta_j = \log(\theta_j)$ for stratum j

  $\bar{\theta} = \exp(\bar{\beta})$ is the time-invariant average hazard ratio

Example: Simulated Dataset (N=400)

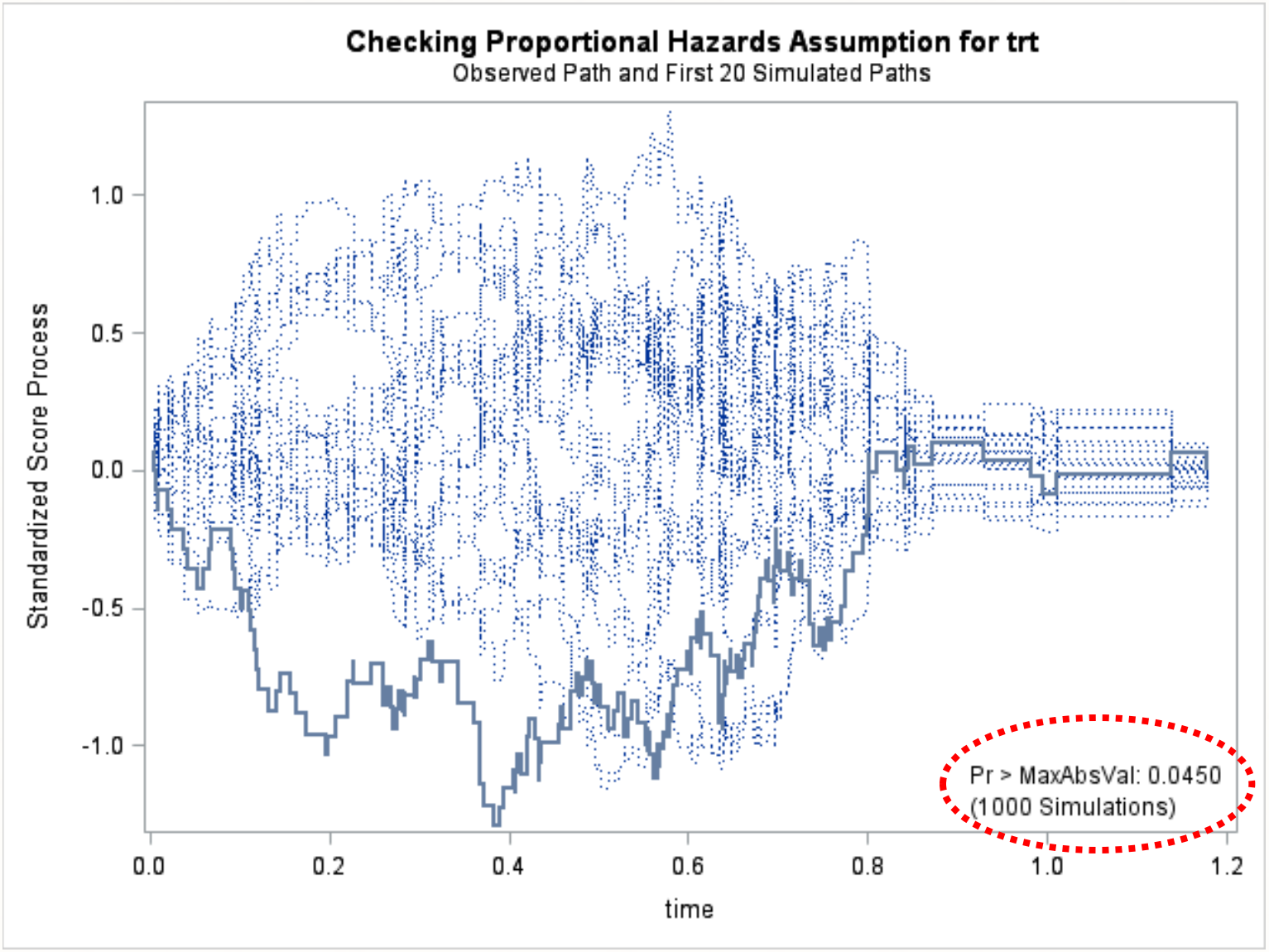


Method	Hazard Ratio (HR)		2-tailed p-value
	Estimate	95% CI	
Cox PH model	0.82	(0.62, 1.07)	.1398

Are results reliable?

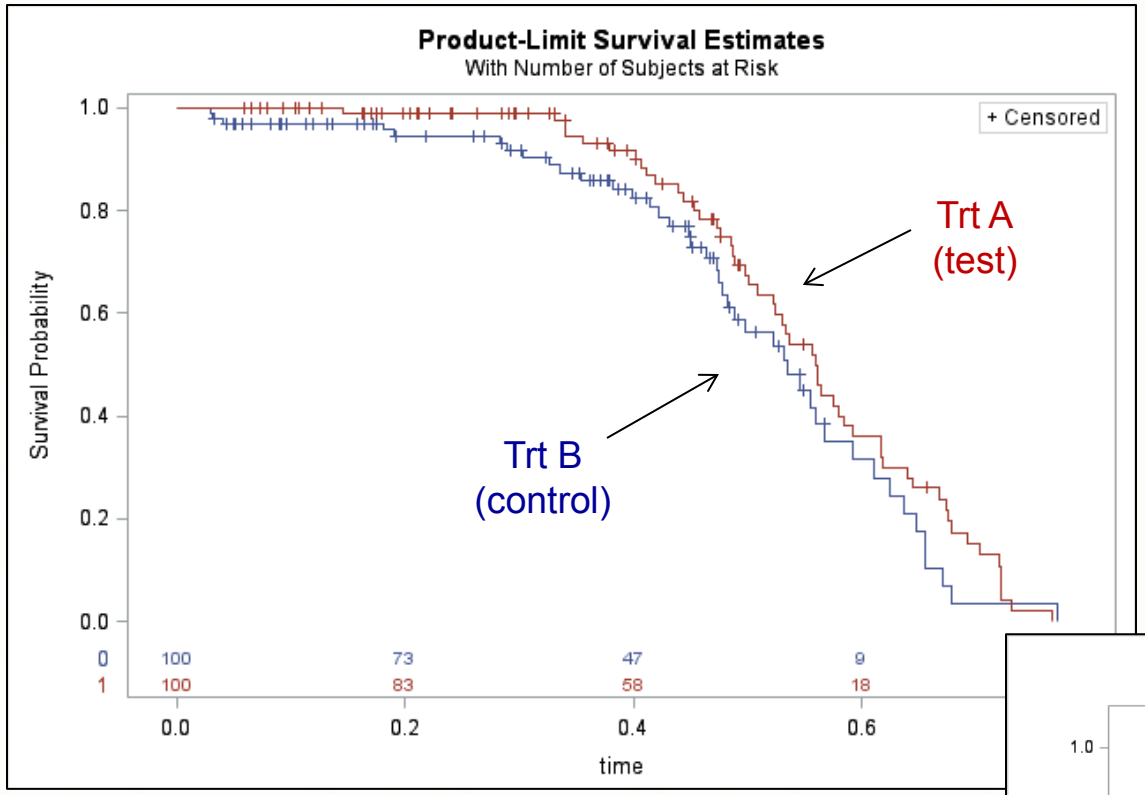
Example (continued): Check PH assumption

Using Cumulative Sums of Martingale Residuals (Lin et al 1993; PHREG/ASSESS)

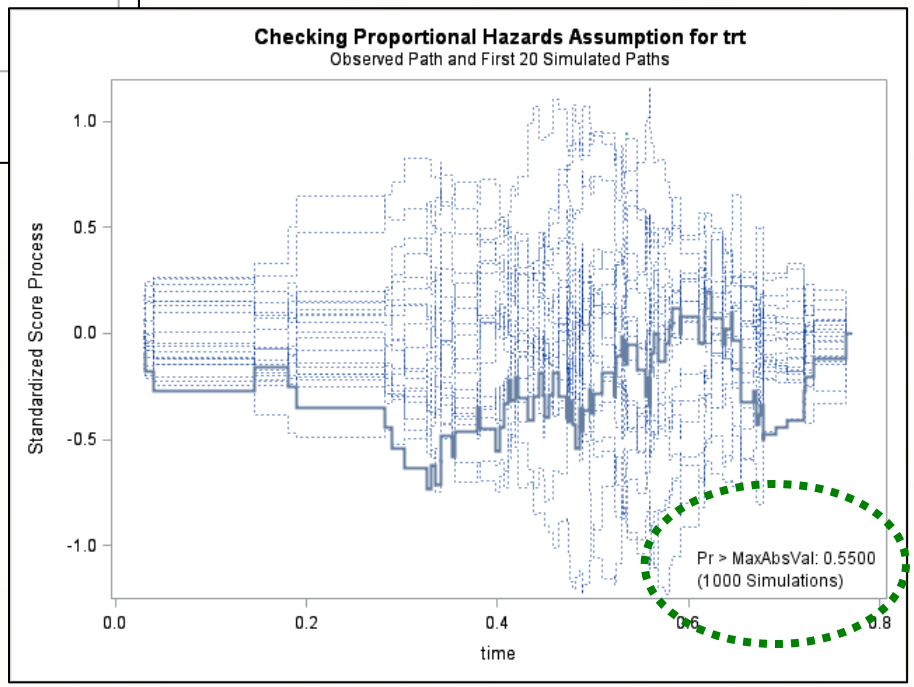


Not PH

Example (continued): Patients in Baseline Risk "Stratum 1"

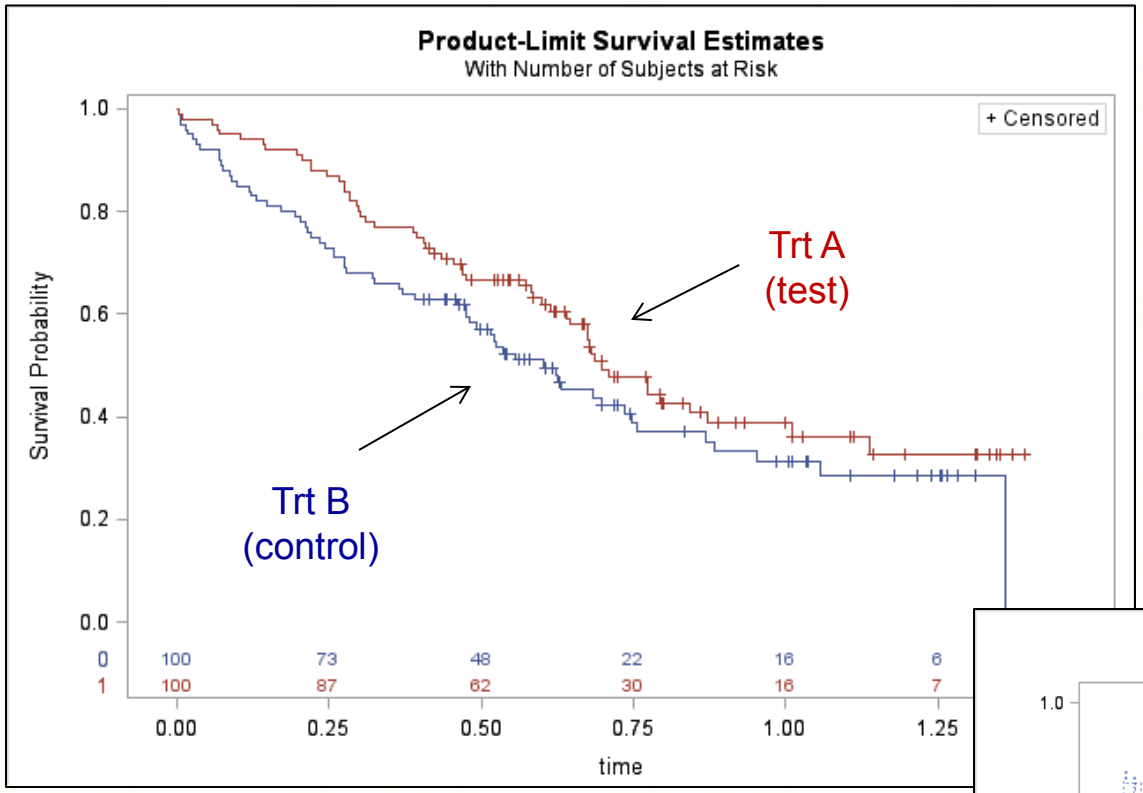


PH

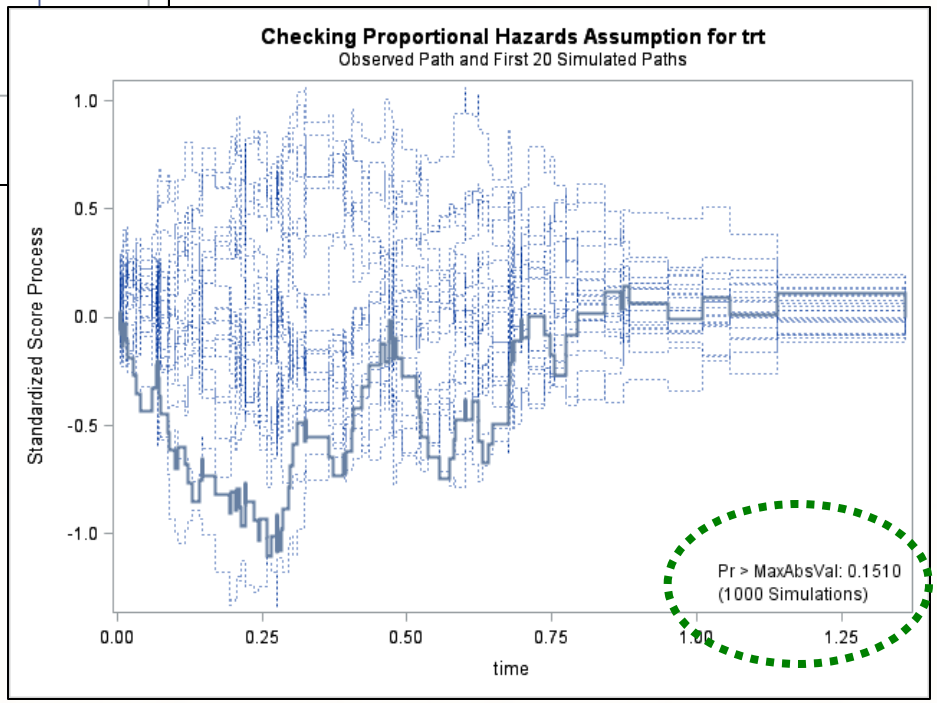


No. of patients	**HR in Stratum 1** Estimate	95% CI
200	0.74	(0.49, 1.13)

Example (continued): Patients in Baseline Risk "Stratum 2"



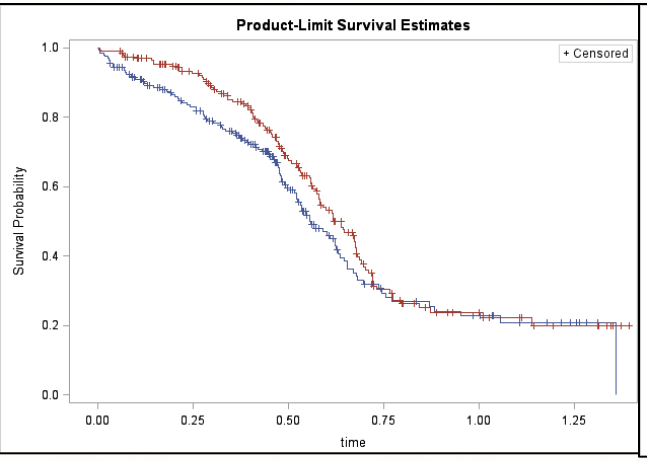
PH

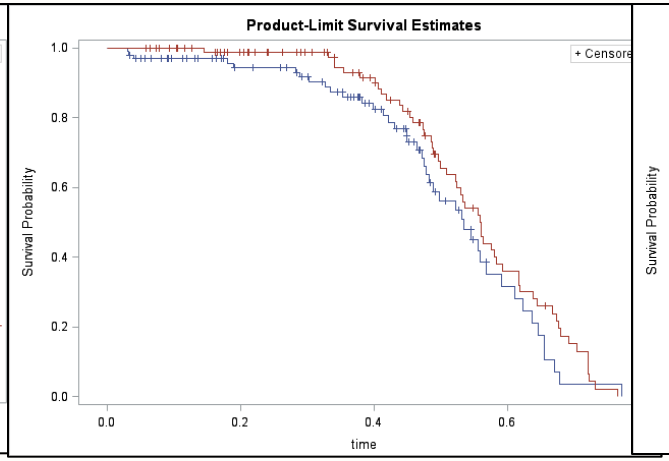


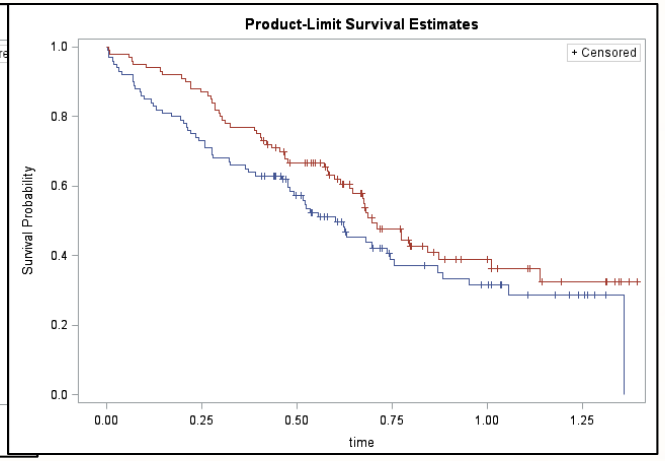
No. of patients	**HR in Stratum 2** Estimate	95% CI
200	0.74	(0.52, 1.07)

Example (continued): Summary of Results

Pooled (not PH) | Stratum 1 (PH) | Stratum 2 (PH)







HR=0.82 (interpretation?) | HR=0.74 | HR=0.74

Analysis Method	Hazard Ratio (HR) Estimate	95% CI	2-tailed p-value
Unstratified Cox PH (pooled)	0.82	(0.62, 1.07)	.1398
Stratified analysis (SSIZE wts)*	0.74	(0.56, 0.98)	.0362

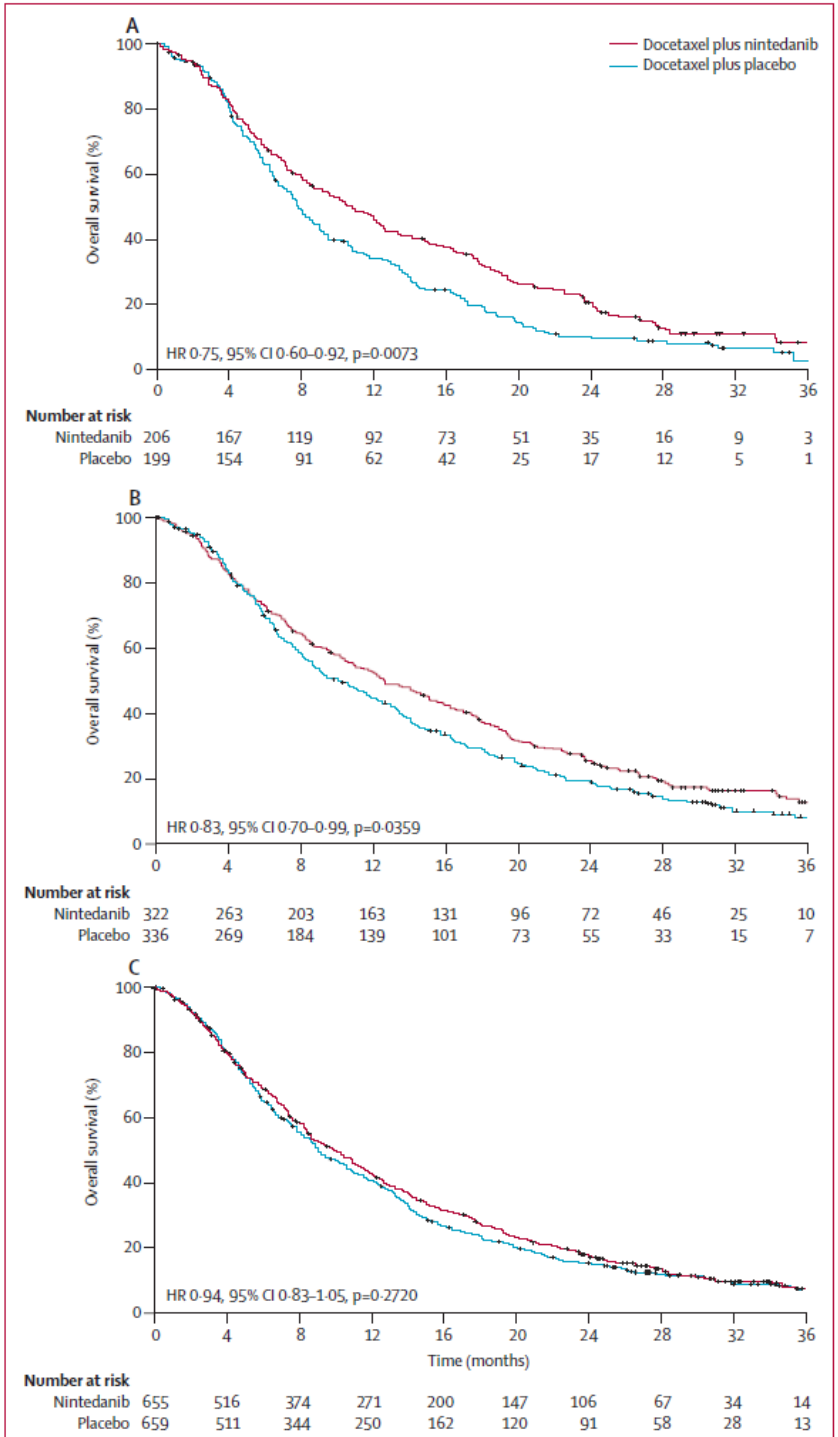
* Mehrotra et al (2012; Statistics in Medicine)

Conclusions

- **ICH E9/R1 estimand & analysis framework is intended to:**
  - Enable better planning of clinical trials and application dossiers for new drugs/vaccines/biologics
  - Strengthen understanding of decision-making by regulatory authorities and advisory committees

- **Open items for estimands and sensitivity analyses:**
  - Time-to-event endpoints under PH/non-PH settings
  - Non-inferiority trials: different from superiority?
  - Treatment effects in well-defined but non-identifiable populations (principal stratification)
  - Others

Back-Up Slide



Docetaxel plus nintedanib versus docetaxel plus placebo in patients with previously treated non-small-cell lung cancer (LUME-Lung 1): a phase 3, double-blind, randomised controlled trial

Lancet Oncol 2014; 15: 143–55

A: subgrp 1
(N=405, HR=0.75)

B: subgrp 1 + subgrp 2
(N=658, HR=0.83)

C: subgrp 1 + subgrp 2 + subgrp 3
(N=1314, HR=0.94)