Estimands and Sensitivity Analyses in Clinical Trials – Strengthening Alignment with ICH E9 (R1)

Devan V. Mehrotra*
Biostatistics and Research Decision Sciences
Merck & Co., Inc., Kenilworth, NJ, USA

* devan_mehrotra@merck.com

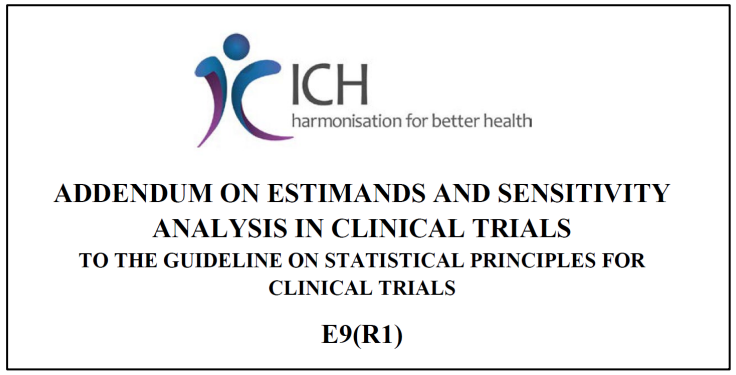
ASA BIOP Regulatory-Industry Statistics Workshop (virtual)
Sep 22, 2021

Outline

- ICH E9(R1)
  - Introduction
  - Key content
  - Impact (thus far)
- Opportunities for stronger alignment with ICH E9(R1)
  - Role of the principal stratum strategy
  - Tipping point sensitivity analyses
  - Estimands/estimation for time-to-event endpoints
  - Estimands/estimation for personalized medicine
- Conclusions

ICH E9(R1): Introduction

- ICH = International Council for Harmonization
- ICH E9 (Statistical Principles for Clinical Trials; 1998)
  - Articulated foundational principles (randomization, double blind, interim analysis, non-inferiority, etc.)
  - Served as a bedrock of regulatory guidance on major statistical aspects of confirmatory clinical trials
- 3Q 2014: ICH expert working group (regulatory and industry statisticians) began developing E9 addendum on **estimands** and **sensitivity analyses**
- 3Q 2017: draft addendum [E9(R1)] released for public comment
- 4Q 2019: first regulatory adoption of final E9(R1) [FDA adoption in 2Q 2021]

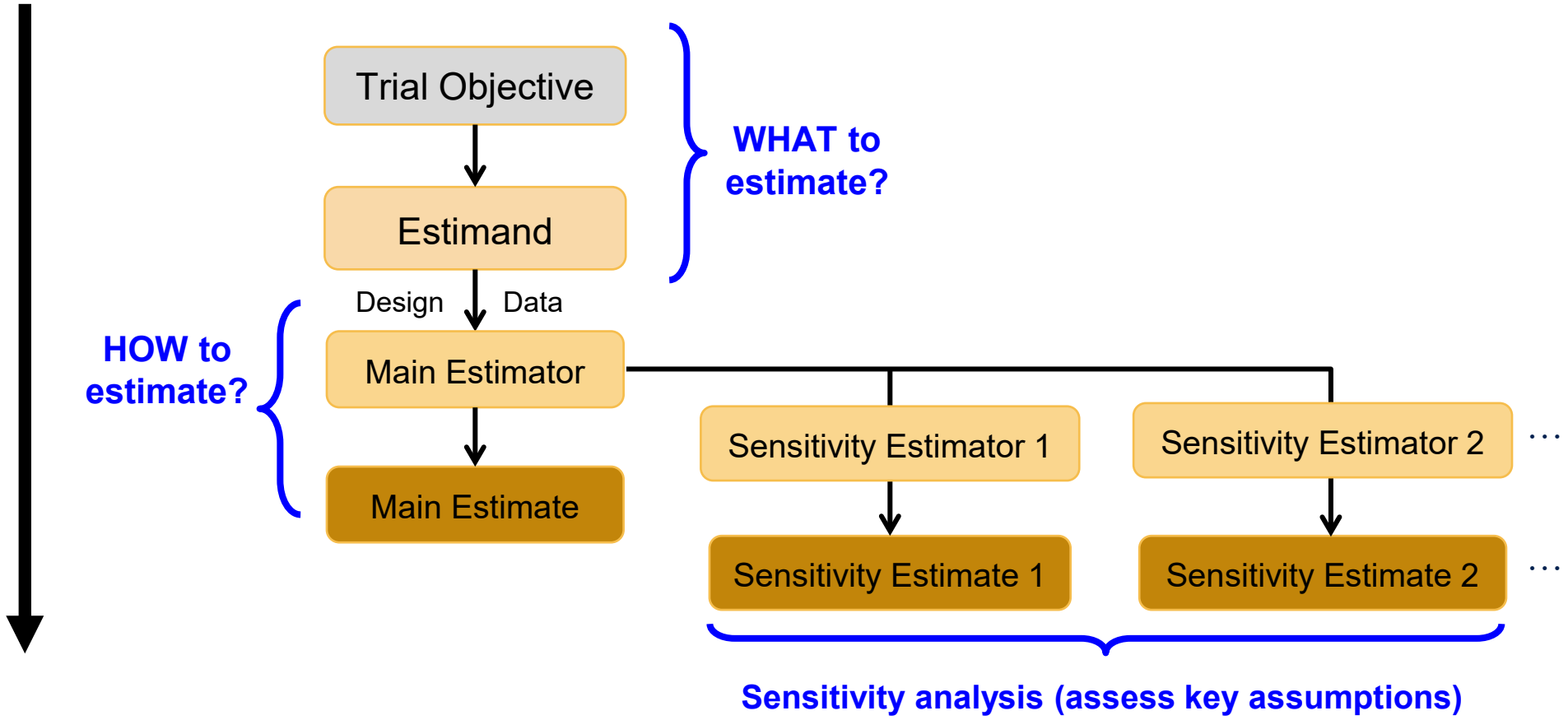


E9(R1) intent: to improve alignment between objectives, design, analysis, interpretation and reporting of clinical trials in publications, regulatory submissions and product labels

ICH E9(R1): Key Content

A Structured Framework

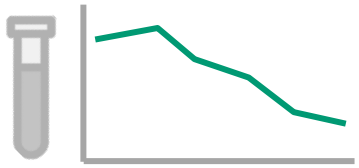
For a given trial objective: aligning target of estimation, design, method of estimation and sensitivity analysis

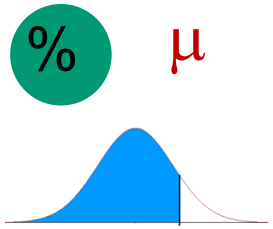


ICH E9(R1): Key Content [2]

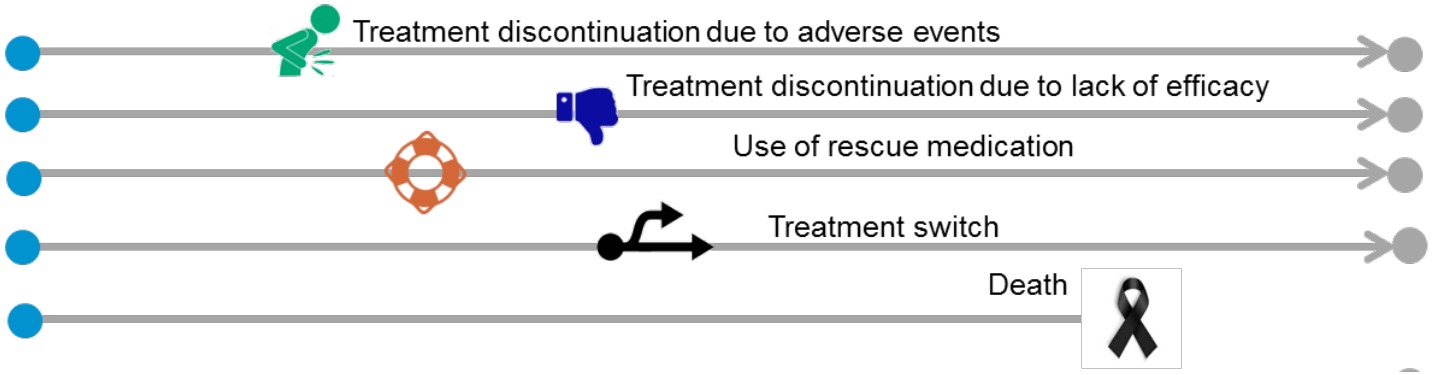
Inputs for Describing an Estimand

Target Population	Variable (Endpoint)	Treatments being compared	Population-level Endpoint Summary





Intercurrent Event(s) that can impact the above input(s)



ICH E9(R1): Key Content [3]

Strategies for Addressing Intercurrent Events

Strategy	***Example of Endpoint or Effect of Interest***
Treatment Policy	Overall survival regardless of whether or when **treatment switching** happens 
Composite	Heart attack **or treatment discontinuation due to AE** 
Hypothetical	Change in HbA1c if **rescue medication** is not used 
Principal Stratum	Infection severity in subpopulation that will **become infected** despite preventive treatment 
While on Treatment	QoL under palliative treatment until **death** in terminal illness 

ICH E9(R1) Impact

Increased use/reporting of estimands in randomized clinical trials

- Reporting of estimands for RCTs in leading medical journals [since 2019]
  - 3 trials in JAMA (1 diabetes, 2 obesity)
  - 4 trials in NEJM (2 diabetes, 1 obesity, 1 COPD)
  - 4 trials in Lancet (4 diabetes)
- Creative naming of estimands in above 11 examples, sometimes with insufficient clarity
  - "Treatment regimen" estimand
  - "Efficacy" estimand
  - "Trial product" estimand
  - "Attributable" estimand

accounted for intercurrent events differently. The efficacy estimand is the treatment effect between tirzepatide and insulin degludec among all randomised participants who continued to receive the study drug without rescue medication. The treatment-regimen

- Almost all analyses of continuous endpoints still done using mixed model repeated measures (MMRM) with hard-to-justify missing at random (MAR) assumption

- Sensitivity analysis not reported (and presumably not done?) in 10 of the above 11 RCTs

ICH E9(R1) Impact [2]

Thoughtful discussions/new ideas in the literature (example below)

Research Article

Statistics in Medicine

Received 5 May 2016, Accepted 10 June 2016 Published online 19 July 2016 in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.7033

Estimands in clinical trials – broadening the perspective

Mouna Akacha,[a*†] Frank Bretz[a] and Stephen Ruberg[b]

perhaps more) as noted earlier. If this treatment is taken as labeled

(1) *Safety*: What percentage of patients was unable to tolerate the treatment and stopped taking it?

  (a) What side effects did they have?
  (b) When did those side effects become apparent, how long did they last?
  (c) Did the side effects reverse themselves and how long did it take?

(2) *Lack of efficacy*: What percentage of patients stopped taking the treatment due to lack of efficacy?

  (a) How long should one continue treatment before giving up on the treatment or adjusting the dose?

(3) *Effects in adherers*: What are the expected beneficial effects and safety concerns of the treatment for patients that complete/continue the course of treatment?

ICH E9(R1) Impact [3]

Formation of Estimand Working Groups (example below)

International Oncology Estimands Working Group

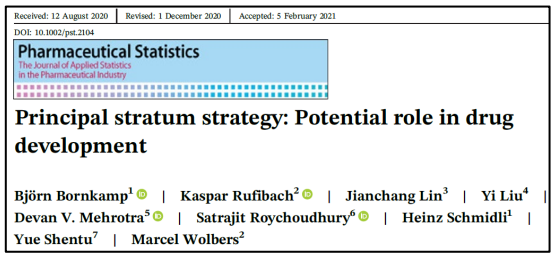
- Goal: A common understanding across industry
- As of 13 April 2021, the working group has 61 members (from Europe, US, and Asia) representing 33 companies
- EFSPI SIG (Nov 2018) and ASA Biopharm Section SWG (Apr 2019)
- In dialogue with eight health authorities globally
- Weblink www.oncoestimand.org

Opportunities for Stronger Alignment with ICH E9(E1)

Role of the Principal Stratum Strategy

Scientific questions of interest often involve definable but non-identifiable subpopulations

Example	Scientific question	Primary endpoint	Intercurrent event	Stratum of interest
Multiple sclerosis	Treatment effect on confirmed disability progression in the subpopulation of relapse-free patients	Time to confirmed disability progression	Post-randomization relapse	Patients who would be relapse-free under both treatments
Treatment effect in early responders	Predict treatment effect on long-term primary endpoint based on early biomarker-type readout	Time-to-event	Biomarker value above or below a pre-specified threshold	Patients who would respond early under treatment vs. those that would not
Antidrug antibodies (ADA) for targeted oncology drugs	Do patients that develop ADAs on either arm still benefit from the drug?	Time-to-event	Development of antidrug antibodies because of receiving treatment	Patients who would be ADA+ under treatment
Impact of exposure on OS	Do patients with insufficient exposure have lower treatment effect?	Time-to-event	Exposure below a pre-specified threshold	Patients with low vs. non-low exposure under treatment
Prostate cancer prevention	Assess effect of treatment to prevent prostate cancer on severity of prostate cancer among those men who would be diagnosed with prostate cancer regardless of their treatment assignment	Time-to-event	Getting prostate cancer	Patients who get prostate cancer irrespective of treatment


Received: 12 August 2020 | Revised: 1 December 2020 | Accepted: 5 February 2021
DOI: 10.1002/pst.2104
Pharmaceutical Statistics
The Journal of Applied Statistics in the Pharmaceutical Industry

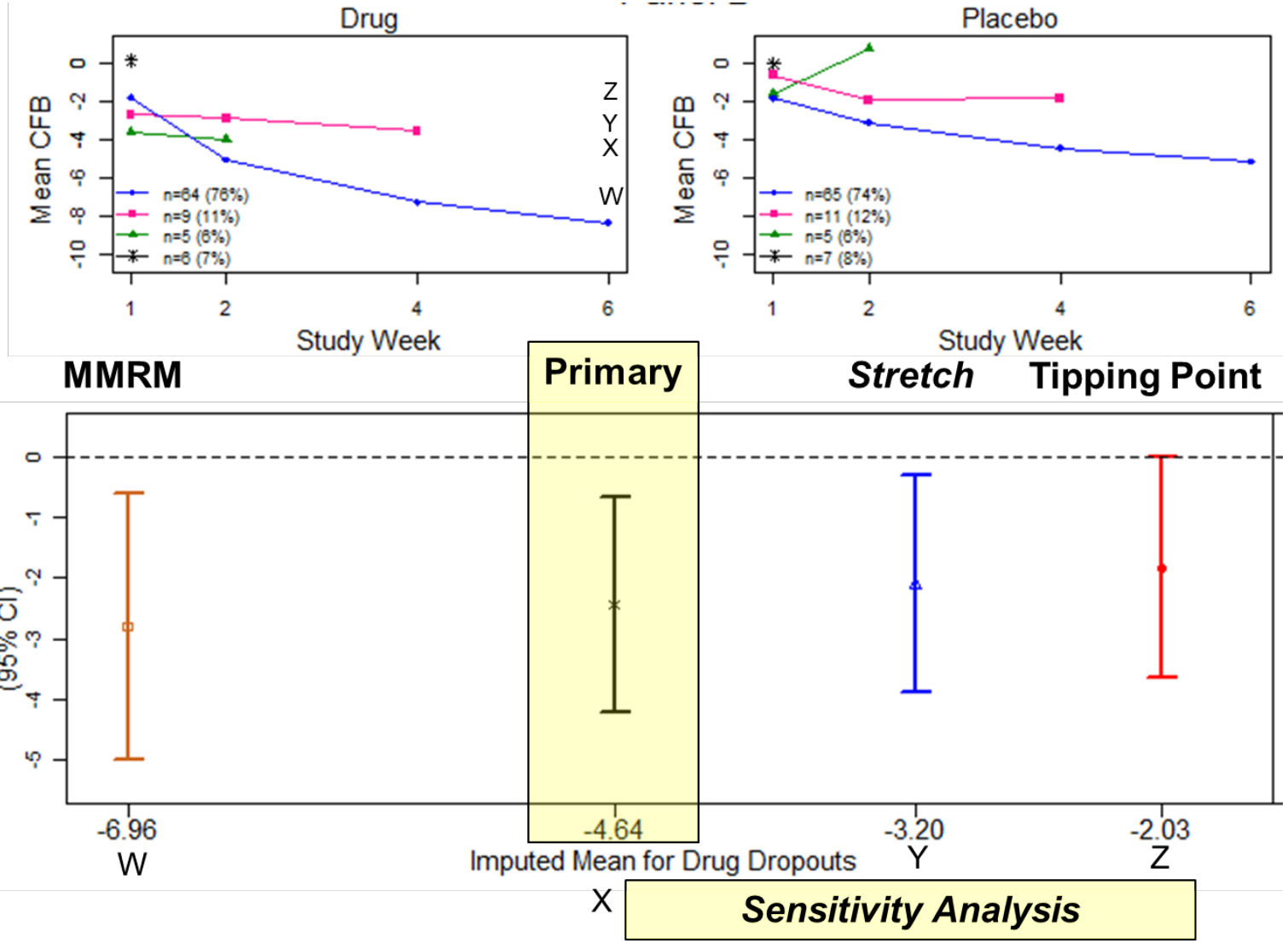
Principal stratum strategy: Potential role in drug development

Björn Bornkamp[1] | Kaspar Rufibach[2] | Jianchang Lin[3] | Yi Liu[4] | Devan V. Mehrotra[5] | Satrajit Roychoudhury[6] | Heinz Schmidli[1] | Yue Shentu[7] | Marcel Wolbers[2]

← **Bornkamp et al (2021)**

Opportunities for Stronger Alignment with ICH E9(E1)

Tipping Point Sensitivity Analysis



Example: tackling **missing data due to dropouts** when estimating the true mean treatment difference for the endpoint of interest (CFB at wk 6)

- comp = completer, drop = dropout
- π_i^{drop} = true Pr(dropout under trt i) = $1 - \pi_i^{comp}$

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

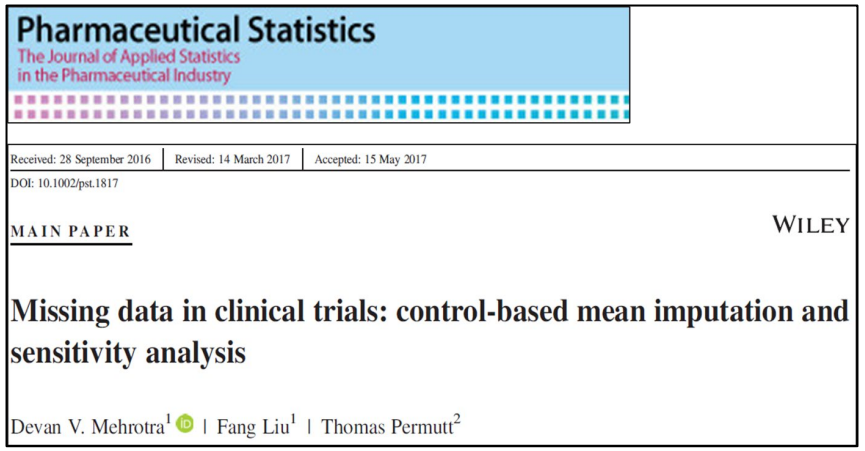
$\hat{\mu}_P$ = estimate of μ_P assuming missingness is MAR for placebo

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

- **Primary Analysis:** use $c = 0$

  ***Stretch* Analysis:** use c such that $\hat{\mu}_P + c = \hat{\mu}_P^{drop}$

  Tipping Point: find c such that p-value = α


Pharmaceutical Statistics
The Journal of Applied Statistics in the Pharmaceutical Industry

Received: 28 September 2016 | Revised: 14 March 2017 | Accepted: 15 May 2017

DOI: 10.1002/pst.1817

MAIN PAPER

WILEY

Missing data in clinical trials: control-based mean imputation and sensitivity analysis

Devan V. Mehrotra[1] | Fang Liu[1] | Thomas Permutt[2]

MMRM exaggerates drug benefit if dropouts are mostly due to toxicity or inefficacy (because MAR assumption becomes implausible)

Opportunities for Stronger Alignment with ICH E9(E1)

Estimands/Estimation for Time-to-Event Endpoints

- Hazard Ratio: difficult to interpret under commonly encountered non-proportional hazards (resulting in analysis without clear estimand)
- Restricted Mean Survival Time (RMST) difference: $\delta(\tau) = \int_0^\tau [S_A(t) - S_B(t)]dt$

  Important: τ is supposed to be pre-defined for estimand and estimation; however, in practice, τ is data-depndent, creating misalignment with ICH E9(R1)
- Geometric Mean Survival Time (GMST) ratio [or, simply, "Time Ratio"]

  Treatment difference in mean log survival time: $\delta = E[log(T_A)] - E[log(T_B)]$

  Time Ratio [TR]: $\gamma = exp(\delta)$

  1) TR = 1.25 $\Rightarrow$ patients survive 25% longer on test vs. control trt, on average
  2) TR is a causal estimand without need for a proportional hazards assumption
  3) TR can be estimated using parametric AFT model fits with model averaging*

(*Mehrotra and Marceau West; 2020, *Statistics in Medicine*)

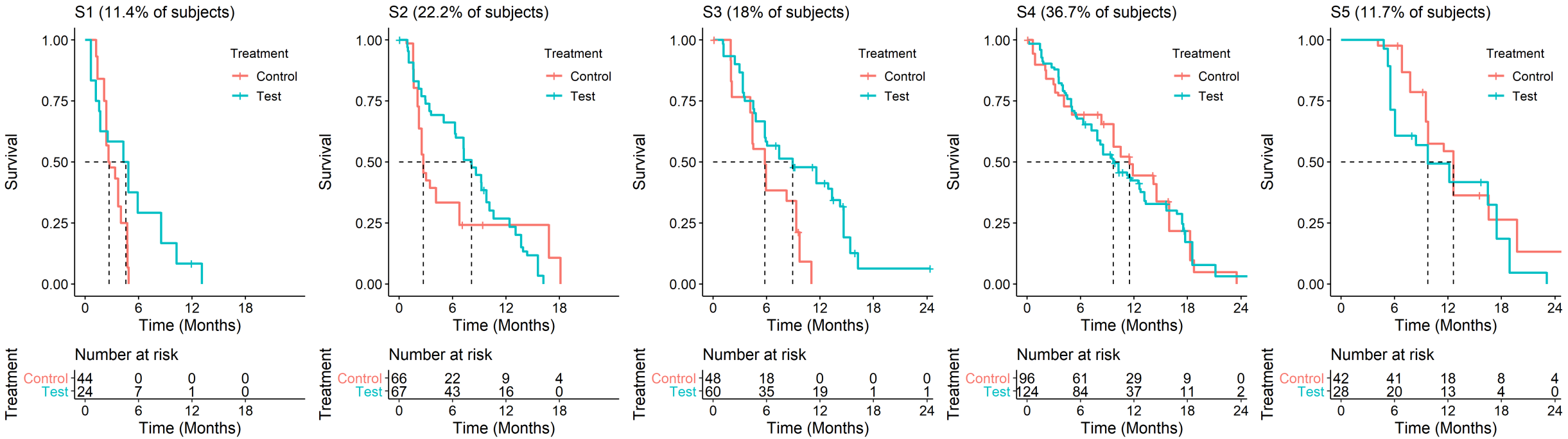
Support for AFT model over PH model: David Cox quote extracted from 1994 interview with Nancy Reid (*Statistical Science*) →

another issue is the physical or substantive basis for the proportional hazards model. I think that's one of its weaknesses, that accelerated life models are in many ways more appealing because of their quite direct physical interpretation, particularly in an en-

5-STAR paper (2020, Stats in Med)
Mehrotra and Marceau West

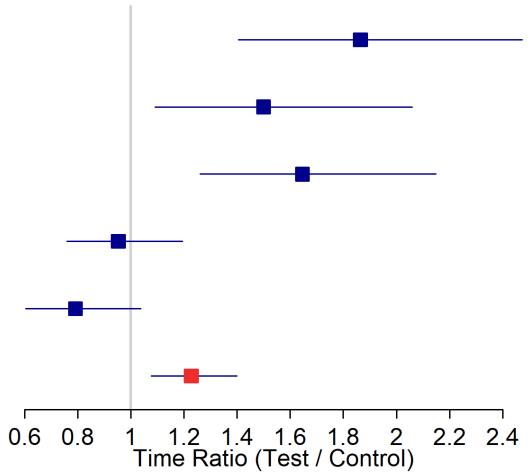
Opportunities for Stronger Alignment with ICH E9(E1)

Estimands/Estimation for Personalized Medicine



Baseline risk-based subgroups identified using treatment-blinded algorithm

Final Strata	No. Subjects (%)	No. Events (%)	Est. TR (95% CI) (Primary)	Pr(TR>1)	Est. HR (95% CI) (Supplemental)	Pr(HR<1)
S1: Cytogencat > 3, Cytogen6 1	68 (11.4)	67 (13.6)	1.86 (1.41, 2.47)	>0.999	0.38 (0.22, 0.64)	0.999
S2: Cytogencat > 3, Cytogen6 0	133 (22.2)	122 (24.7)	1.50 (1.09, 2.06)	0.994	0.94 (0.69, 1.30)	0.617
S3: Cytogencat <= 3, IPSS > 2, Priorout Progress	108 (18.0)	89 (18.0)	1.65 (1.26, 2.15)	>0.999	0.41 (0.28, 0.61)	>0.999
S4: Cytogencat <= 3, IPSS > 2, Priorout in (Failure, Relapse)	220 (36.7)	162 (32.8)	0.95 (0.76, 1.19)	0.336	1.05 (0.80, 1.37)	0.384
S5: Cytogencat <= 3, IPSS <= 2	70 (11.7)	54 (10.9)	0.79 (0.60, 1.04)	0.045	1.59 (1.00, 2.50)	0.049
5-STAR Average	**599 (100)**	**494 (100)**	**1.23 (1.08, 1.40)**		**0.81 (0.66, 0.99)**	



TR: time ratio

HR: hazard ratio

Conclusions

- Uptake/impact of ICH E9(R1) thus far has been [mostly] as anticipated
  - Increased efforts towards aligning trial objectives, design and analysis
  - Reporting of estimands in leading medical journals
  - Some analyses (e.g., MMRM) still misaligned with stated estimands
  - Expanding research literature on estimands (mostly good, some confusing)
  - Abundant enthusiasm across formed estimand working groups
- Examples of areas needing stronger alignment with ICH E9(R1)
  - Sensitivity analyses to evaluate robustness of assumption-driven conclusions
  - Estimands/estimation for time-to-event endpoints (alternatives to HRs)
  - Estimands/estimation for personalized medicine

Back-Up Slides

Principal Stratum Estimand: HIV Vaccine Example

Estimand-aligned primary and sensitivity analysis

Estimand

δ_{VL} = true between-treatment difference (placebo – vaccine) in mean viral load set-point **among those who will become HIV infected regardless of treatment assignment**

	Vaccine	Placebo
Number randomized	N_v	N_p
Number HIV infected	n_v	n_p
Proportion infected	$\frac{n_v}{N_v}$	$\frac{n_p}{N_p}$
Viral load set-points of infected subjects ($\log_{10}$ copies/ml)	$\begin{bmatrix} y_1^{(v)} \\ \vdots \\ y_{n_v}^{(v)} \end{bmatrix}$	$\begin{bmatrix} y_1^{(p)} \\ \vdots \\ y_{n_p}^{(p)} \end{bmatrix}$

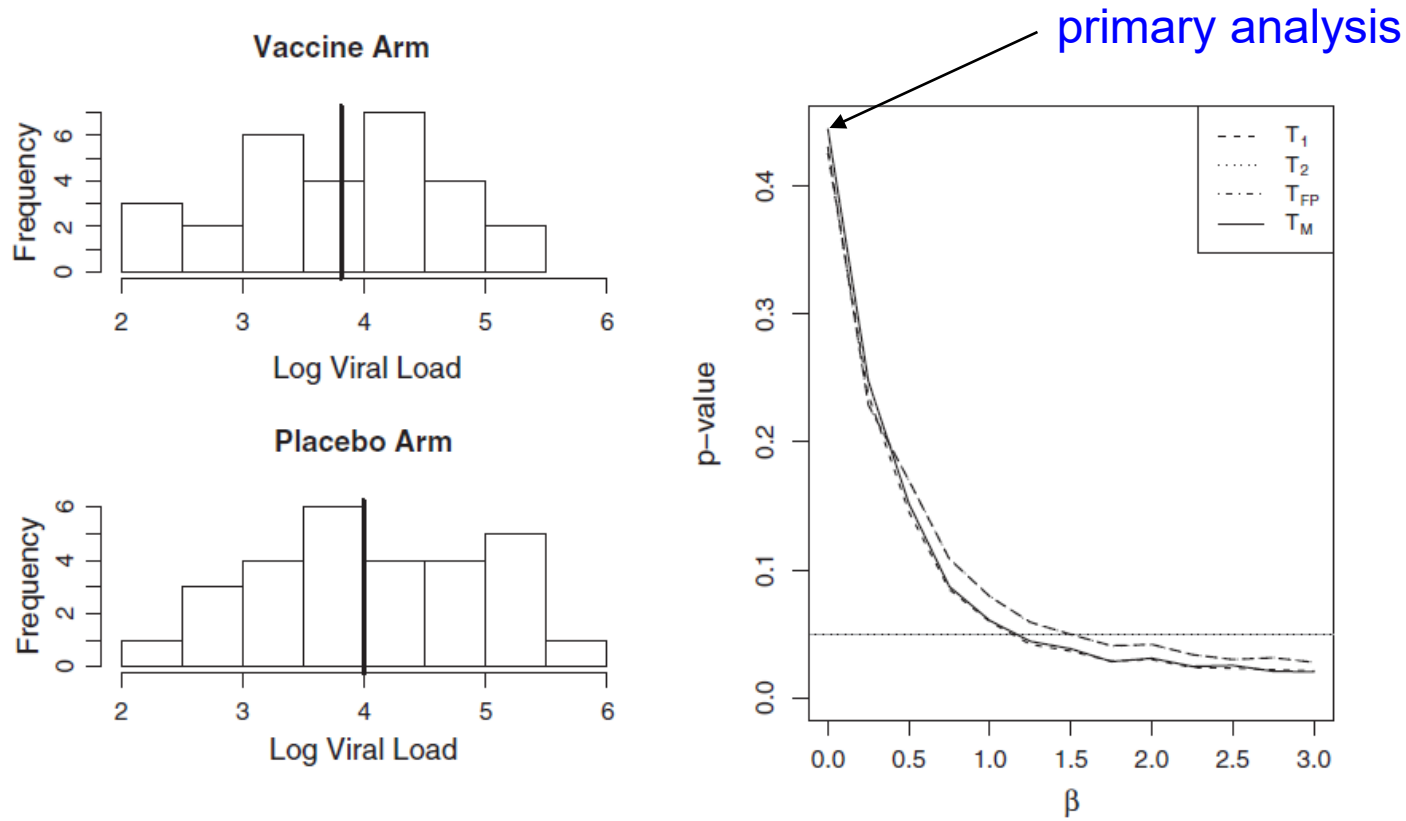


Figure 5. Left plots: Histograms of $\log_{10}$-transformed viral load among non-whites infected in the vaccine and placebo arms of the VaxGen trial. The vertical line represents the mean. Right plot: Sensitivity analysis of the test of association between vaccination and viral load among those who would have been infected regardless of treatment assignment. β represents the log-odds of infection in vaccine given infection in placebo and placebo viral load. All tests were two-sided.

Lu et al (2013, Statistics in Medicine); interpretation of the sensitivity parameter (β) can be challenging