Industry Perspective on Subsequent Therapy: An Estimands Approach

Jonathan Siegel, Bayer

Viktoriya Stalbovskaya, Novartis

Yue Shentu, Merck

Juliane Manitz, EMD Serono

Evgeny Degtyarev, Novartis

Kaspar Rufibach, Roche

Hannes Buchner, Staburo

Rui Tang, Servier

Marie-Laure Casadebaig, Celgene

Natalia Kan-Dobrosky, PPDI

Vincent Haddad, AstraZeneca

Science For A Better Life

The Pharmaceutical Industry Working Group on Estimands in Oncology

ASA Biopharmaceutical Section Regulatory-Industry Statistics Workshop September 23, 2020

Disclaimer

The views presented in this presentation are solely those of the authors on behalf of the Pharmaceutical Industry Working Group on Estimands in Oncology, Treatment Switching Subteam, and do not represent the views of their respective employers or any other organization.

Background

- This presentation addresses cases where
  - Patients in a clinical trial choose subsequent therapy because of the special conditions of the clinical trial environment
  - They would not have chosen the same subsequent therapy if they had taken the study drug in real-world clinical practice
- It does not address cases where the subsequent therapy given is SOC therapy that ordinary patients would receive outside a trial
- In the session abstract example, hematopoietic stem cell transplantation and kidney transplantation are both SOC subsequent therapies for patients receiving dialysis.
  - They are not artifacts of clinical trial conditions.
  - This presentation does not cover this case.

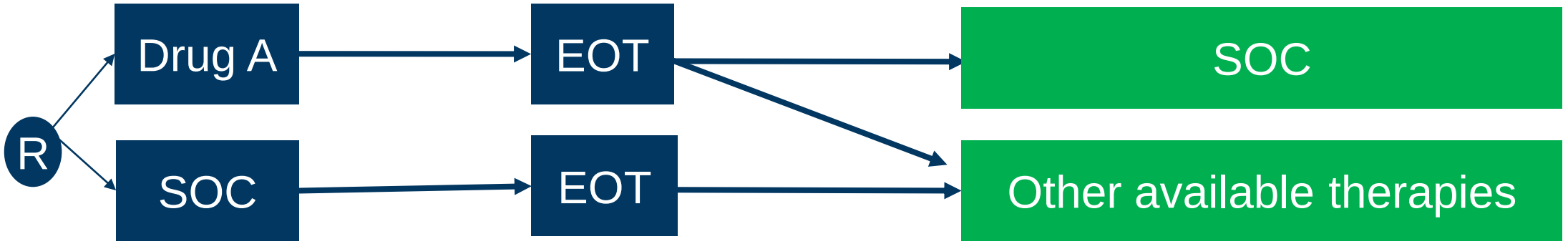
Trial subsequent therapy vs. real-world practice

- Examples of cases where the subsequent therapy selected in a clinical trial does not reflect real-world practice include:
  - Subsequent therapy is itself experimental and unavailable as SOC.
  - Patients assigned to placebo or control withdraw study treatment and receive subsequent therapy immediately or early.

An artifact of being in a clinical trial

- Non-SOC subsequent therapy situations can be regarded as an artifact of the clinical trial environment.
  - SOC may be context-specific
- Experimental subsequent therapy is not available in the real world.
- There is no placebo in the real world.
  - Patients do not normally stop therapy immediately after being prescribed it.
  - Behavior incident to being assigned placebo is an artifact of a clinical trial environment and not the environment of inference.

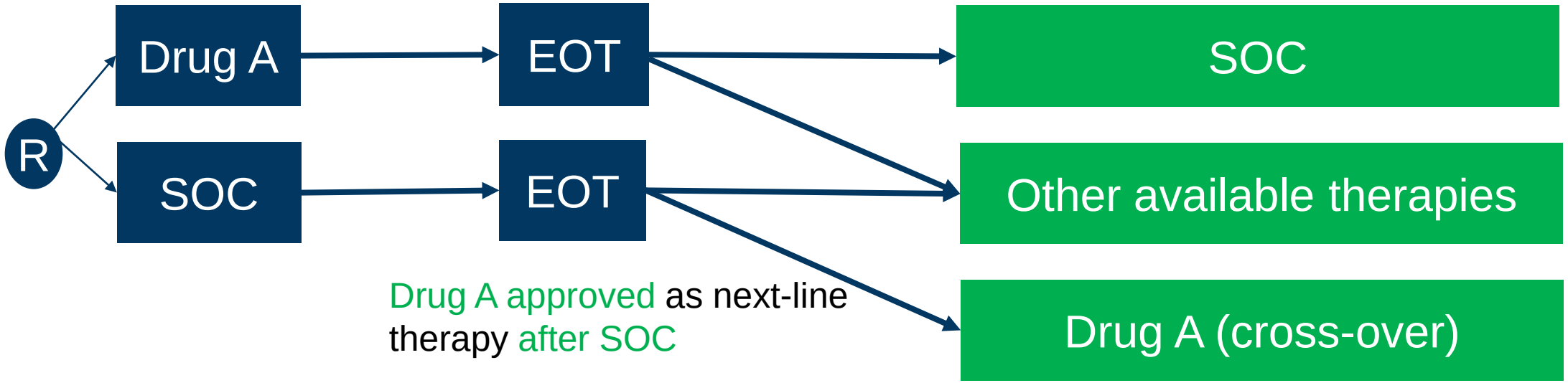
Subsequent therapy is SOC



☺ choice of subsequent therapies after EOT reflects clinical practice

SOC: Standard of Care; EOT: End of Treatment

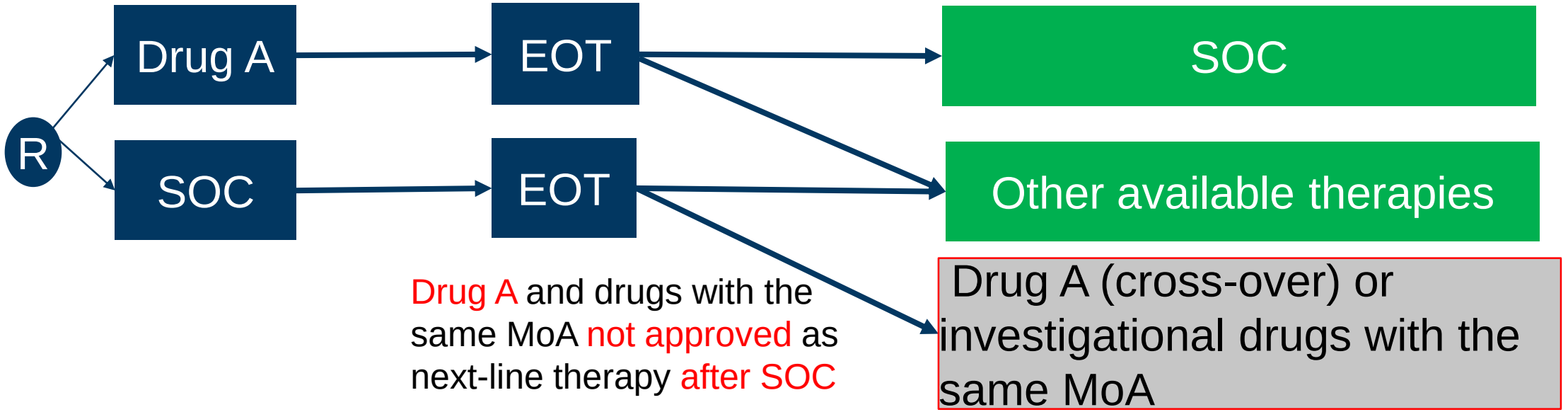
Approved 2nd line therapy tested in first line



☺ choice of subsequent therapies after EOT reflects clinical practice

SOC: Standard of Care; EOT: End of Treatment

Subsequent therapy is not SOC



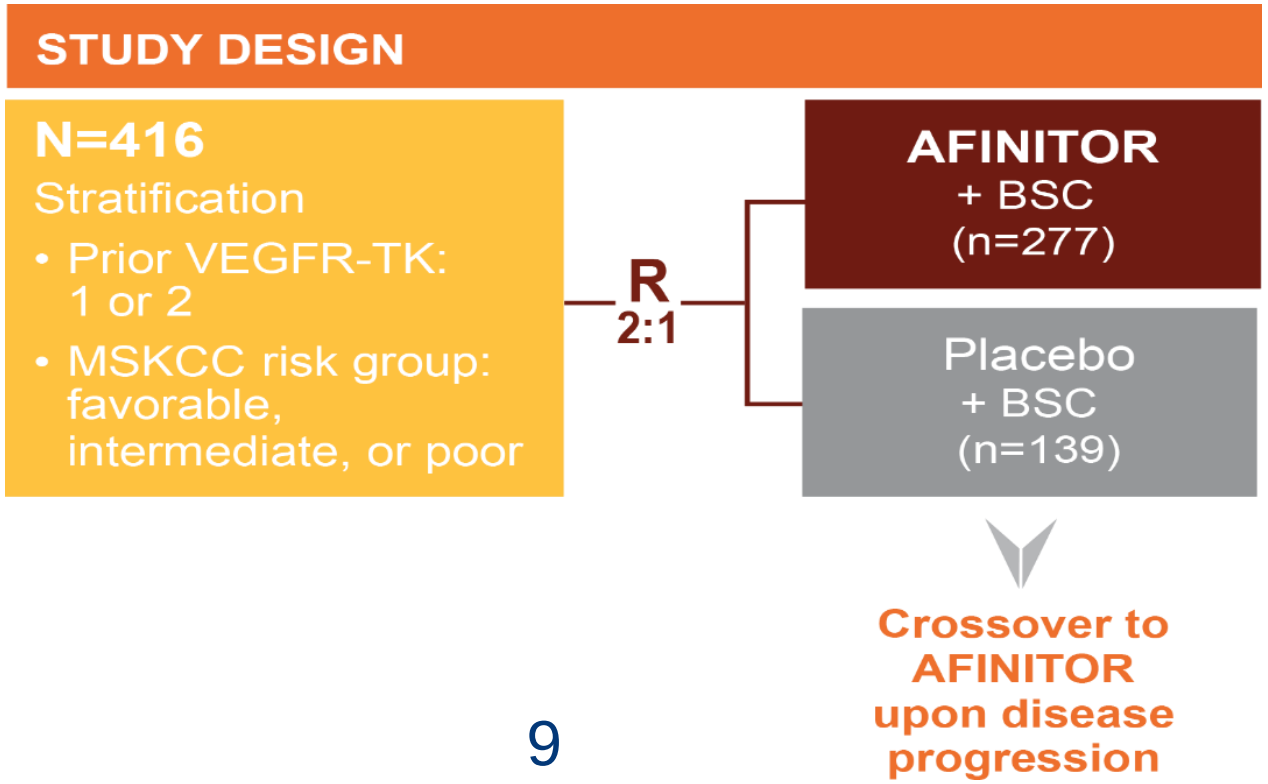
☹ choice of subsequent therapies after EOT does **not** reflect clinical practice

SOC: Standard of Care; EOT: End of Treatment

NOVARTIS | Reimagining Medicine

Detailed Example: study RECORD-1

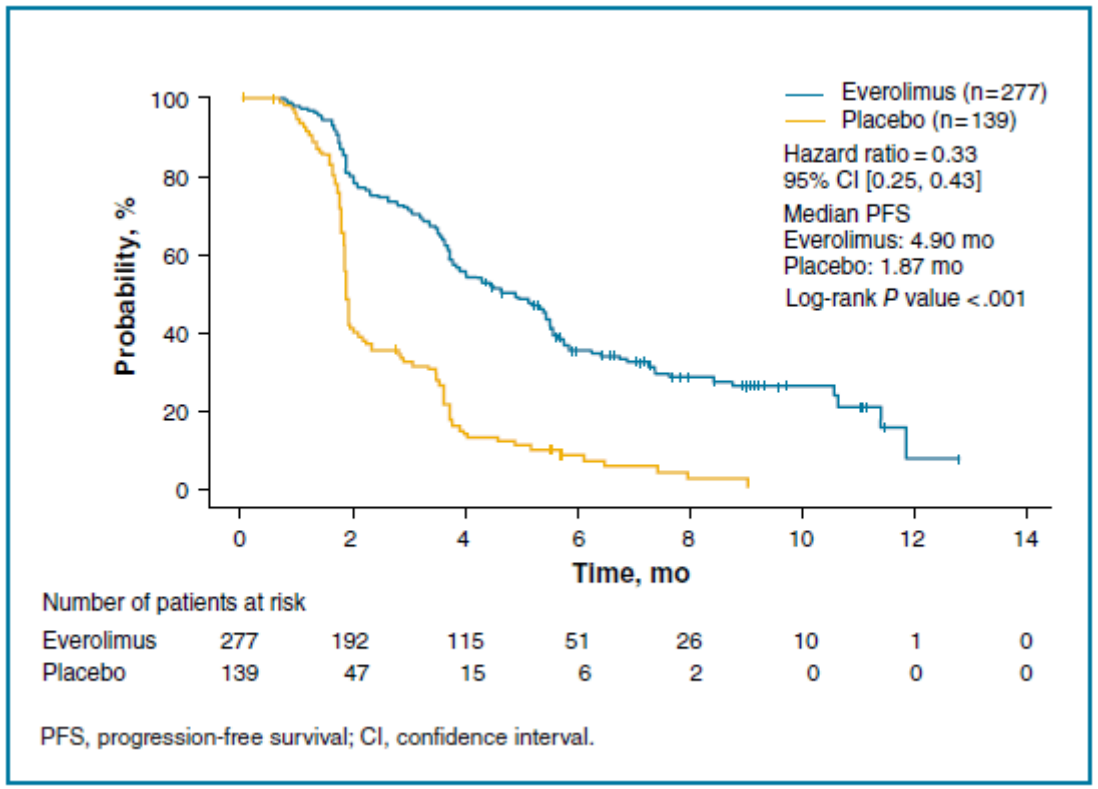
- Phase III study of everolimus in metastatic renal cell carcinoma Motzer et al (2008, 2010)
  - Double-blind, multicenter study with patients randomized to receive either everolimus (n = 277) or placebo (n = 139)
  - Primary endpoint – Progression-free survival defined as time from randomization until disease progression or death



Motzer et al (2010)

Example: study RECORD-1

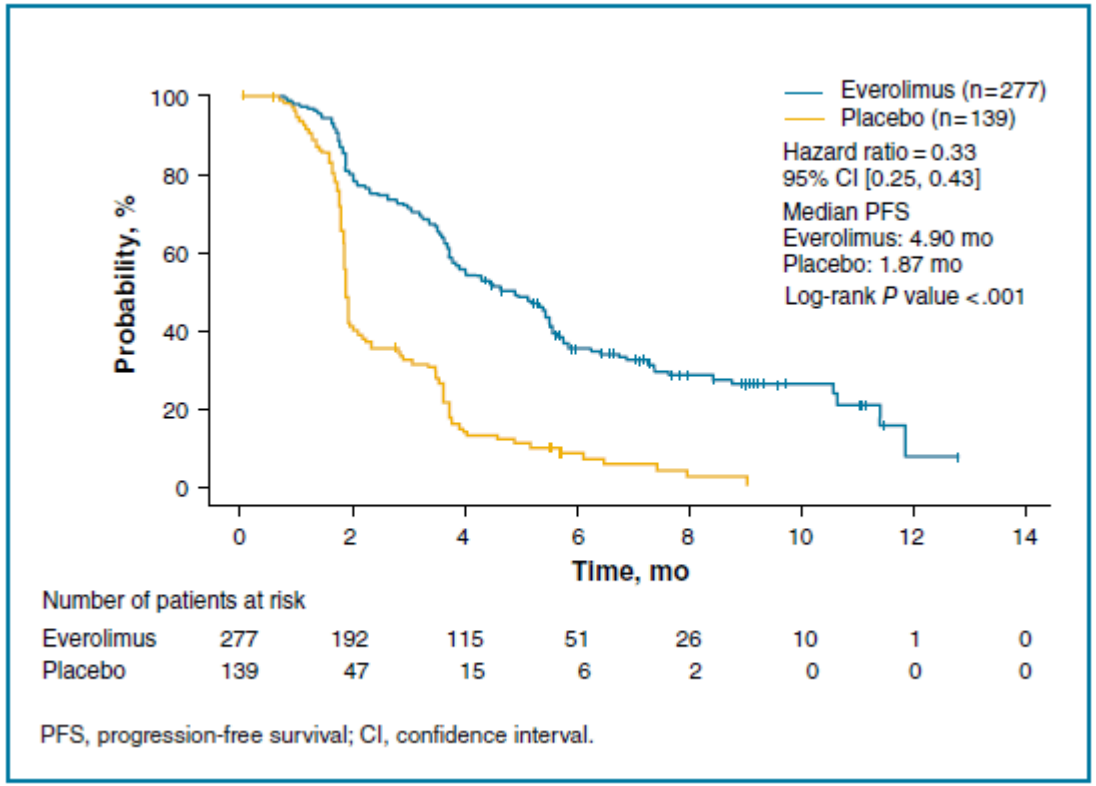
- Positive study with clinically meaningful improvement in PFS (HR=0.33, 95% CI: 0.25, 0.43, p-value < 0.001)

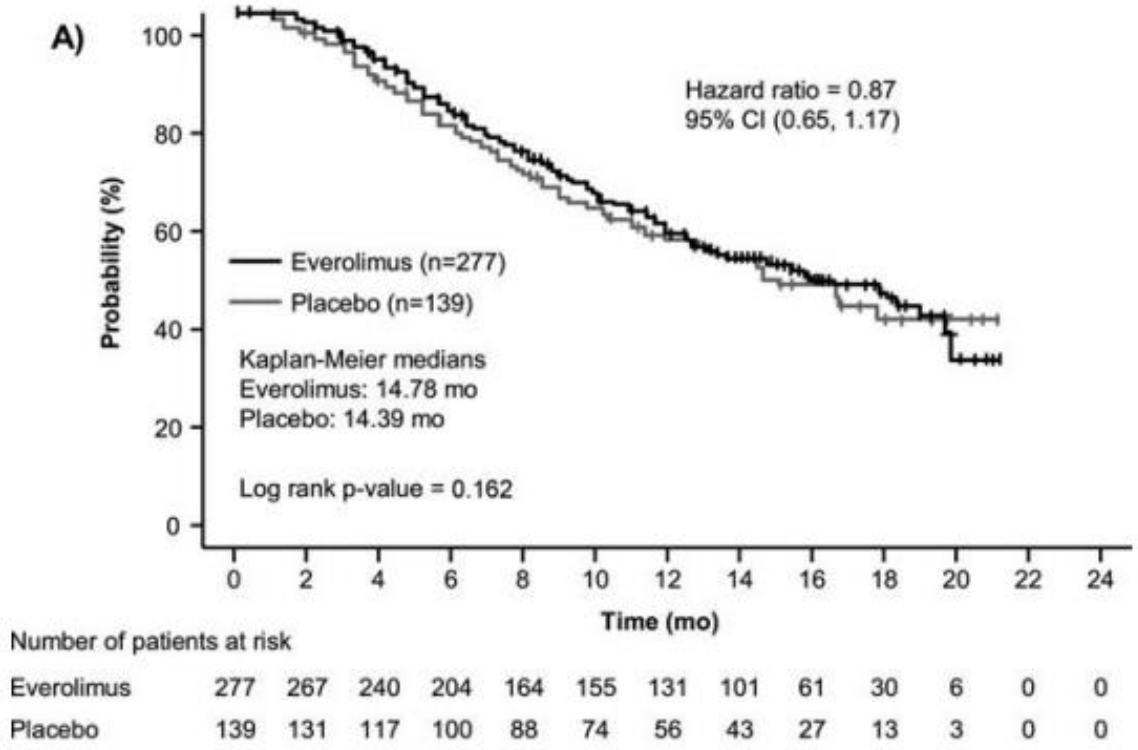


Motzer et al (2010)

Example: study RECORD-1

- Positive study with clinically meaningful improvement in PFS (HR=0.33, 95% CI: 0.25, 0.43, p-value < 0.001)
- Protocol allowed crossover from placebo to everolimus upon progression (106 out of 139 patients, 76%)
  - ITT analysis of OS showed trend in OS benefit (HR=0.87, 95% CI: 0.65-1.15, p-value=0.162)

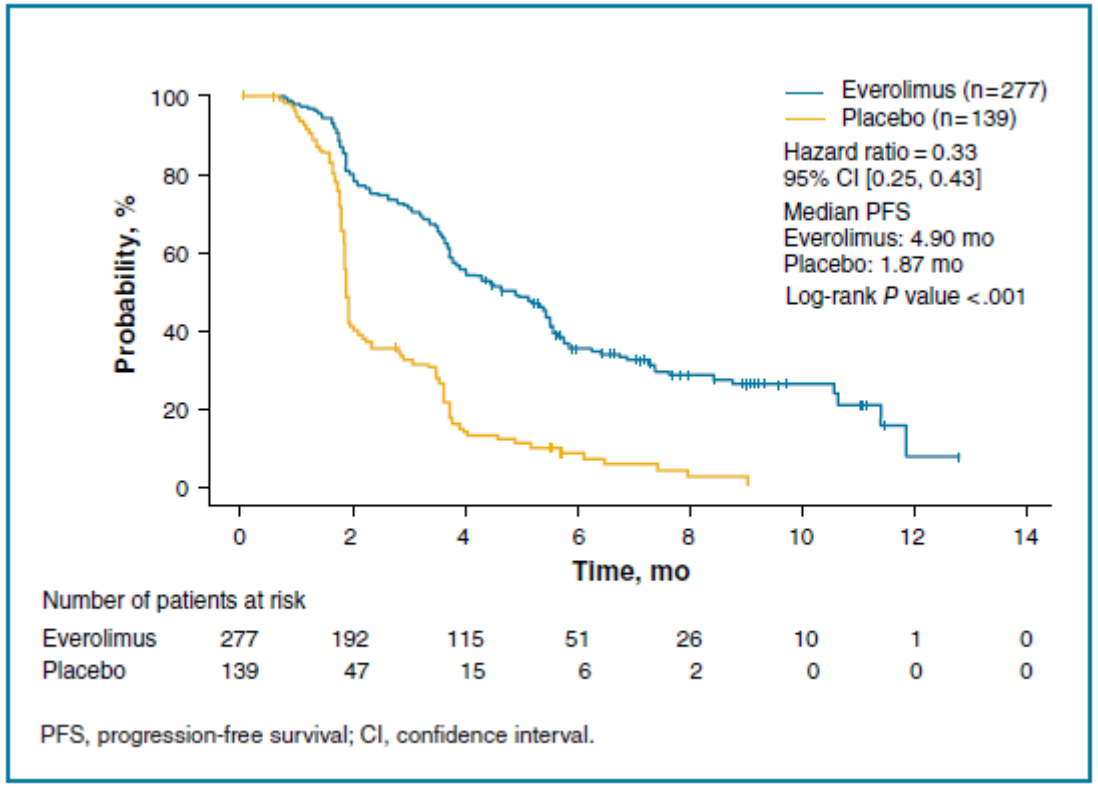


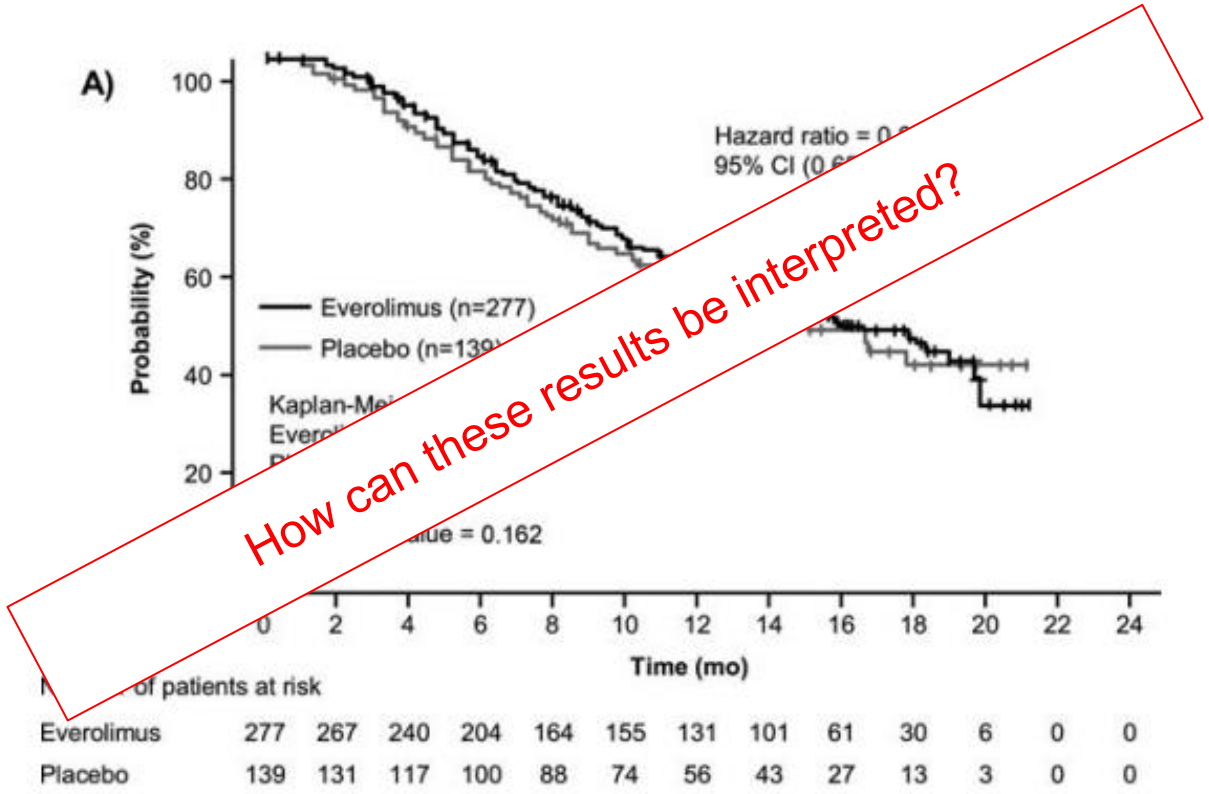


Motzer et al (2010)

Example: study RECORD-1

- Positive study with clinically meaningful improvement in PFS (HR=0.33, 95% CI: 0.25, 0.43, p-value < 0.001)
- Protocol allowed crossover from placebo to everolimus upon progression (106 out of 139 patients, 76%)
  - ITT analysis of OS showed trend in OS benefit (HR=0.87, 95% CI: 0.65-1.15, p-value=0.162)





Motzer et al (2010)

Further Examples

Example 1:

The placebo-controlled GRID trial with a high rate of crossover of placebo patients to regorafenib (85%) at progression **where crossover was allowed per protocol**

→regorafenib significantly improved PFS and OS showed a positive trend, although not reaching statistical significance in the ITT

Demetri G.D., Reichardt P., Kang Y-K., Blay J-Y., Joensuu H., Wagner A., Kappeler C., Casali P.G., Final overall survival (OS) analysis with modeling of crossover impact in phase III GRID trial of regorafenib vs placebo in advanced gastrointestinal stromal tumors (GIST). Journal of Clinical Oncology 34 (4_suppl): 156-156, DOI: 10.1200/jco.2016.34.4_suppl.156.

Example 2:

The GLARIUS trial which compared standard temozolomide (TMZ) versus bevacizumab plus irinotecan (BEV+IRI) in patients with newly diagnosed glioblastoma

- Crossover to BEV+IRI therapy was given to 81.8% of all patients who received any sort of second-line therapy in the TMZ arm, affecting OS

➢ Within such settings it can even happen that, on average, patients in the control arm have a similar exposure to the investigational treatment as the patients in the investigational arm

Herrlinger U, Schaefer N, Stainbach JP et al. Bevacizumab Plus Irinotecan Versus Temozolomide in Newly Diagnosed O 6-Methylguanine-DNA Methyltransferase Nonmethylated Glioblastoma: The Randomized GLARIUS Trial. Journal of Clinical Oncology. 2016; 34 (14): 1611-1619. doi: 10.1200/JCO.2015.63.4691

Immediate treatment switching in open label trials

Open-label studies have the risk that patients stop randomized treatment after randomization in the control arm and seek the opportunity to receive an investigational therapy in another clinical trial, possibly even from the same class as the investigational drug in the previous trial.

Examples:

- Checkmate-37, comparing Nivolumab vs chemotherapy where 20% of the patients from the control arm withdrew consent immediately after they learned that they were randomized into the control arm (vs. 1.5% on investigational arm)
- Quantum-R trial (2019): 23% in placebo withdrew immediately vs 1.6% on investigational arm

Larkin J., Minor D., D'Angelo S., Neyns B., Smylie M., Miller W.H. Jr., Gutzmer R., Linette G., Chmielowski B., Lao C.D., Lorigan P., Grossmann K., Hassel J.C., Sznol M., Daud A., Sosman J., Knushalani N., Schadendorf D., Hoeller C., Walker D., Kong G., Horak C., Weber J., Overall Survival in Patients With Advanced Melanoma Who Received Nivolumab Versus Investigator's Choice Chemotherapy in CheckMate 037: A Randomized, Controlled, Open-Label Phase III Trial. J Clin Oncol. 2018; 36(4):383–390. doi:10.1200/JCO.2016.71.8023

Position

The ICH E9 R1 Addendum, "Addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials," can serve as a basis for addressing design and interpretation of trials involving heavy treatment switching.

*The particular concern is to ask the right scientific question, defining the **treatment** element of the estimand appropriately and inferring to a **population** of patients in a real-world clinical practice environment.*

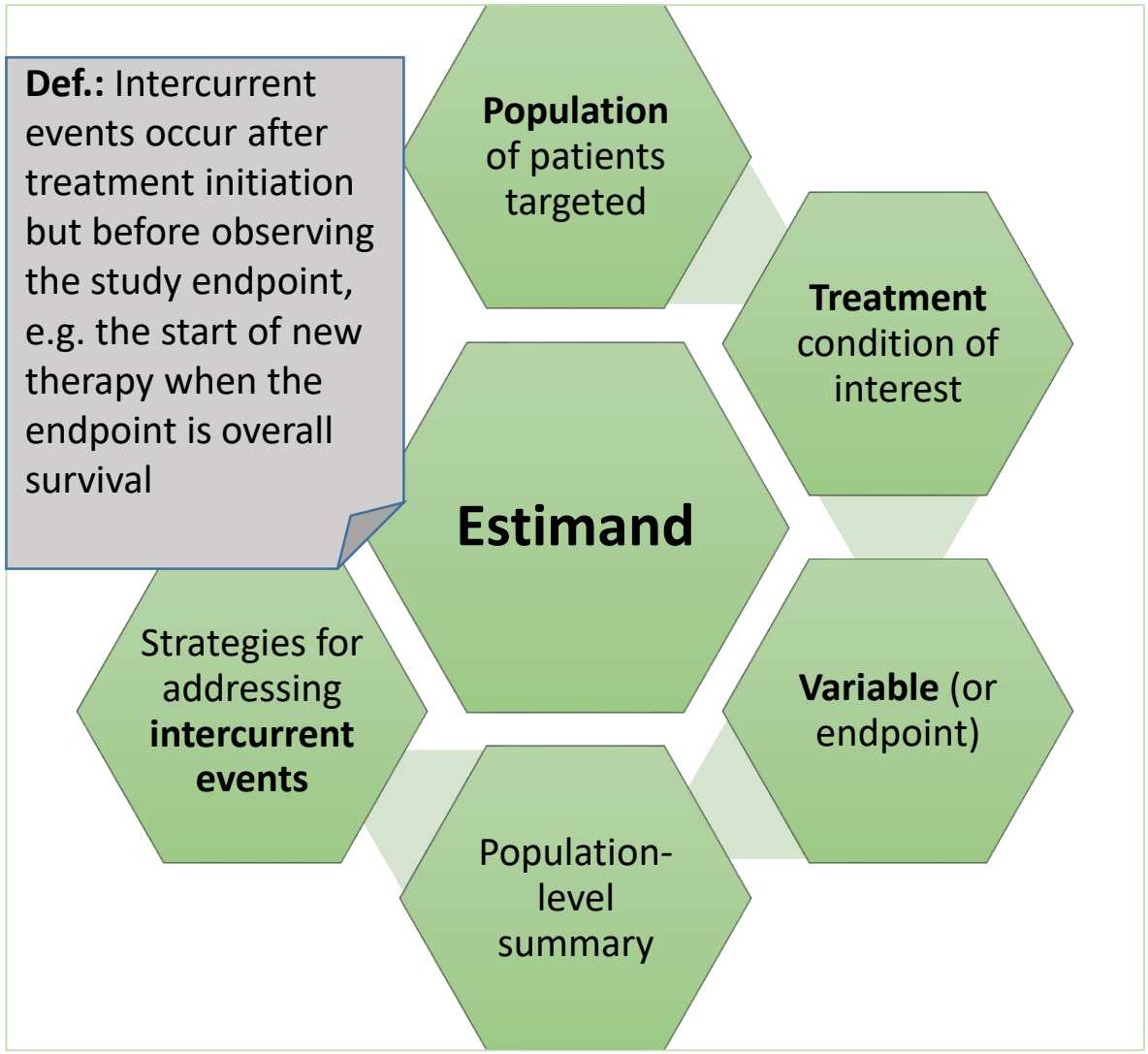
Who We Are - Oncology Estimands WG

- initiated and led by Evgeny Degtyarev (Novartis) and Kaspar Rufibach (Roche), first TC Feb 2018
- Goal: **A common understanding across industry**
- 39 members (14 from Europe and 24 from US) representing **24** companies
  - 5 subteams: causal, treatment switching, censoring mechanisms, case studies in solid tumors, case studies in hematology
- **EFSPI SIG** (Nov 2018) and **ASA Biopharm Section SWG** (Apr 2019)
- in dialogue with regulators from EMA, FDA, Japan, China, Taiwan, Canada, MHRA
- **Weblink** www.oncoestimand.org.

Treatment switching subteam

- Viktoriya Stalbovskaya
- Juliane Manitz
- Marie-Laure Casadebaig
- Emily Martin
- Rui (Sammi)Tang
- Godwin Yung
- Vincent Haddad
- Fei Jie
- Christelle Lorenzato
- Jiangxiu Zhou
- Evgeny Degtyarev
- Hannes Buchner
- Jonathan Siegel
- Natalia Kan-Dobrosky

What is an Estimand?



- **Estimand** is the target of estimation to address the scientific question of interest posed by the study objective.
- An estimand is described by five attributes, defining together the treatment effect of interest.

➢ Increase transparency with respect to data analysis and inference

➢ Align trial objectives and statistical analyses by requiring a precise definition of the population quantity of interest

➢ Strengthen the dialogues between disciplines involved in the formulation of clinical study objectives, design, conduct, analysis and interpretation

TYPES OF INTERCURRENT EVENTS, AND STRATEGIES

Uninformative

- Scientific question is what happened regardless of the intercurrent event.
- Uninformative events do not introduce bias or alter the estimand.
- Outcome after event is still of interest
-

Primary Strategy:

- Treatment Policy

Positively Informative
provide qualitative information about the event of interest

- Scientific question is what actually happened, including the intercurrent event
- Intercurrent event is informative for effect of interest
- Goal of methodological improvement is to better incorporate the intercurrent event into the analysis

Primary Strategy:

- Composite

Counterfactual
confound the event of interest

- Scientific question is what would have happened if intercurrent event had not occurred.
- Intercurrent events rendered uninformative conditioned on a model

Primary Strategies:

- Hypothetical
- Principal Stratum

Irrelevant

- Scientific question is about what happened prior to the intercurrent event
- Outcome after intercurrent event is considered irrelevant.

Primary Strategy

- While on Treatment

No way to determine informative censoring definitely

Purpose of treatment policy estimand

- The estimand closest to the conventional ITT approach is a treatment policy estimand.
- In a treatment policy estimand, the scientific question addresses the effect of the complete regimen including assigned study treatment and all subsequent therapy. Subsequent therapy is part of the treatment element.
- The approach rests on the following assumptions:
  - Subsequent therapy reflects clinical practice in the particular decision context
  - Patients receiving subsequent treatments and dose intensity as expected (as SOC) between investigational and control arm

Treatment Policy Estimand - Details

- ***Conventional ITT methods correspond to a treatment policy estimand in the estimands framework.***
- ***Objective:*** Evaluate OS benefit including subsequent therapies representing clinical practice
- ***Estimand:***
  - **Population:** Defined through appropriate I/E criteria to reflect the target patient population for approval
  - **Variable:** Overall survival: Time from randomization to death
  - **Treatment:**
    - Sequence of investigational drug + any subsequent therapies vs. sequence of control + any subsequent therapies (including Investigational drug)
  - **Handling of intercurrent events:**
    - Intercurrent events including subsequent therapy and crossover are ignored
  - **Population-level Summary:** Hazard ratio and confidence interval
- ***Estimate:*** Log-Rank test, Cox model, KM estimates etc. using ITT approach

What happens when the clinical trial environment does not predict real-world practice?

- As we've discussed at the beginning of this presentation, in a number of trial contexts subsequent therapy choice is an artifact of the clinical trial context and does not reflect real-world practice.
- In these cases, we suggest a treatment policy estimand may not be the most appropriate approach.
- It will achieve its methological goal of reliably predicting results of a future repeat clinical trial under the same conditions.
- But it will not achieve its clinical goal of reliably predicting what will happen in real world practice if the drug were approved.

Position

Where the clinical trial environment results in events which do not reflect real-world clinical practice, a hypothetical estimand should be considered.

Hypothetical Estimand

- Scientific question is what would have happened if the intercurrent event had not occurred.
- Trial events which do not reflect real-world clinical practice represent intercurrent events, confounding inferability of the trial results.
- Hypothetical strategies in regulatory submission trials are not novel
  - Censoring for subsequent therapy, previously mandated by the 2007 FDA Cancer Endpoint Guidance, implements a hypothetical estimand
  - Addresses scientific question of what would have happened if patients (a) had not received subsequent therapy AND (b) had same future hazards as patients who did not
  - Estimand is NOT recommended. The underlying assumptions (e.g. therapy change provides no information about future hazards) are generally unreasonable. Example merely demonstrates that hypothetical strategies are part of the traditional repertoire.

Hypothetical Estimand - Details

- ***Objective:*** Evaluate OS benefit adjusted for treatment switching
- ***Estimand:***
  - **Population:** Defined through appropriate I/E criteria to reflect the target patient population for approval
  - **Variable:** Overall survival: Time from randomization to death
  - **Treatment:**
    - Investigational drug vs control (what would have happened if there were no subsequent therapies)
  - **Handling of intercurrent events:**
    - Analysis method addresses counterfactual (what would have happened if intercurrent event had not occurred)
  - **Population-level Summary:** Hazard ratio and confidence interval

Estimate: Adjusted HR and CI from IPCW-weighted Cox model

Recommended treatment policy vs. hypothetical approaches

OBJECTIVE	*Evaluate OS benefit assuming subsequent therapies represent clinical practice*	*Evaluate OS benefit adjusted for treatment switching*	*Evaluate OS benefit adjusted for treatment crossover*	*Evaluate OS benefit adjusted for treatment crossover at disease-related time-point*
ESTIMAND				
Population	Defined through appropriate I/E criteria to reflect the target patient population for approval			
Variable / Endpoint	Overall survival: Time from randomization to death			
Treatment condition of interest	Sequence of investigational drug + any subsequent therapies vs. sequence of control + any subsequent therapies (including Investigational drug)	Investigational drug vs control (if there were no subsequent therapies)	Sequence of investigational drug + any subsequent therapies vs. sequence of control + any subsequent therapy (excluding investigational drug)	Sequence of Investigational drug + any subsequent therapies vs. sequence of control + any subsequent therapy (excluding investigational drug)
Handling of intercurrent events (IEs)				
IE: Start of subsequent therapy at any time	Treatment policy	Hypothetical	Treatment policy	Treatment policy
IE: Crossover to investigational drug at any time	Treatment policy	Hypothetical	Hypothetical	Treatment policy
IE: Crossover to investigational drug at disease – related time point	Treatment policy	Hypothetical	Hypothetical	Hypothetical
Population - level Summary	Kaplan – Meier estimates; Hazard ratio (HR) with confidence interval (CI)			
ESTIMATION	Cox model and KM estimates using ITT approach	Adjusted HR and CI from IPCW – weighted Cox model; weighted KM estimates	HR from RSPFT model using adjusted survival times; bootstrapped CI; KM estimates using adjusted survival times; IPCW methods could also be used	HR from two – stage method using reconstructed survival; modified KM estimates using reconstructed survival times; IPCW and RPSFT methods could be used

A Maximalist Thesis

- ICH E9 R1 requires carefully defining and focusing on the ***clinical*** question of interest as the gravamen of a valid clinical trial.
- Meaningful clinical questions must predict real-world practice
- Where trial behavior differs sharply from clinical practice behavior, the trial fails to answer the clinical question
- A strategy that answers the wrong question is not a valid strategy
  - It simply doesn't' matter how reliably its results can be reproduced or how definitively the wrong question can be answered.*
  - It's like the fable of the Wise Men of Chelm who lose their keys in a dark alley but look under the lamp-post because it's so much easier to look there.
  - We do not compromise on clinical relevance.
- Only a hypothetical strategy provides a meaningful answer to the clinical question of interest
- Only a hypothetical strategy is valid.

A Maximalist Antithesis

- ICH E9 R1 requires carefully defining and focusing on the ***scientific*** question of interest as the gravamen of a valid clinical trial.
  - Questions addressed in individual clinical trials must be answered scientifically.
  - The answers must be (at least conceptually) reproducible.*
- Not every clinical question can be answered in a sufficiently reproducible way.
  - Clinicians often have valid questions that cannot be answered using rigorous science
  - Attempts must be made to find an estimand that approximates their original question and that can be addressed rigorously**
- Hypothetical estimands depend on strong, unproven, and (generally) untestable assumptions
  - Their validity and reproducibility is questionable
  - Counterfactual causal inference assumption of no unobserved confounding may be particularly problematic.
- Accordingly, such methods should be avoided.

**Scott Emerson, personal communication.*

***Professor Emerson does not condone selecting methods solely for reliability. See also Fleming, TR, Rothman, MD, and Lu, HL. Issues in using progression-free survival when evaluating oncology products. J Clinical Oncology 27:2874-2880 (2009). ("The goal of clinical research is not to obtain a statistically significant effect. Rather, 'the primary goal should be to obtain a statistically reliable evaluation regarding whether the experimental intervention is safe and provides clinically meaningful benefit.')*

What Should Industry Do?

- Widespread crossover to study or similar-class experimental therapy, immediate withdrawal from control treatment, etc. can prevent a reasonable expectation of trial success despite efficacious treatments.
  - An estimated 54% of Phase III trials fail.*
  - As discussed in the examples above, multiple recent large-scale trial failures have involved high crossover.
- Likely high crossover situations are often (not always) foreseeable.
- What is industry to do when it forecasts these situations as highly likely?
  - Should industry simply stop developing drugs under these conditions because trials with a reasonable chance of success are simply infeasible?

**Fogel, DB. Factors associated with clinical trials that fail and opportunities for improving the likelihood of success: A review. Contemporary Clinical Trials Communications 11:156-164 (2019)*

Improving the reliability of hypothetical estimands

Data Collection	
Documentation of subsequent therapy by investigator and clarity to ensure accurate coding.	
Collecting data on factors associated with treatment switching as time-varying covariates.	
Sensitivity Analyses	
ICPW	**RPSFT**
Inclusion and exclusion of potential confounders	Stability of treatment effect if treatment continued
Truncation of observations with extreme values	Impact of common treatment effect assumption (RPSFT), reduce relative effect size by a factor for patients who receive the therapy at a later stage.
Exclusion of patients constrained from switching	
The proportion of patients switching should not be close to 0% or 100%	

Best Practices for the Design, Implementation, Analysis, and Reporting of Oncology Trials with High Rates of Treatment Switching. A Guidance Document from the Green Park Collaborative. Center for Medical Technology Policy, October 2016, Version 1.0.
Sullivan TR et al. Adjusting for treatment switching in oncology trials: A systematic review and recommendations for reporting. Value in Health 23:388-396 (2020).

Towards a Synthesis

- W. Edwards Deming distinguished between enumerative studies, involving sampling from a fixed frame, and analytic studies, with no fixed frame.
  - Only enumerative studies can be reliable without unverifiable assumptions
- "Management is prediction":
  - Studies attempting to predict the future are inherently analytic.
  - Human behavior involves interaction
  - Clinical trials involve analytic, interactive elements.
- The estimands framework reflects these decades-old observations

Deming WE. On a classification of the problems of statistical inference. JASA 37: 173–185 (1942); Deming, WE. On probability as a basis for action. American Statistician 29:146-152 (1975)

Can We Go Outside the Box?

- Statisticians play a crucial role in design in the estimands framework
- Good statistical design seeks to prevent bias by changing trial design and conduct up front where possible, using statistical adjustment techniques only as a last resort.
  - A statistician should challenge the status quo and ask if there is a different way.
- Conceived of as a design problem, cross-over to experimental therapy can be regarded as a consequence of a system of interacting trials
  - When a patient crosses from Trial A to Trial B and trial B results are attributed to A, the two trials are not independent. Assuming independence leads to error.
  - The resulting confounding is an example of negative interaction, or interference*
- Deming might have advocated managing such a system of trials cooperatively, as a system, to reduce cross-trial interference, through e.g. a master trial protocol.
- The current environment constrains our ability to do this.
  - We are obliged to work within the environment we have, including its constraints.

**Deming, WE, The New Economics for Industry, Government, Education, 2nd Edition. MIT Press 2000.*

Synthesis

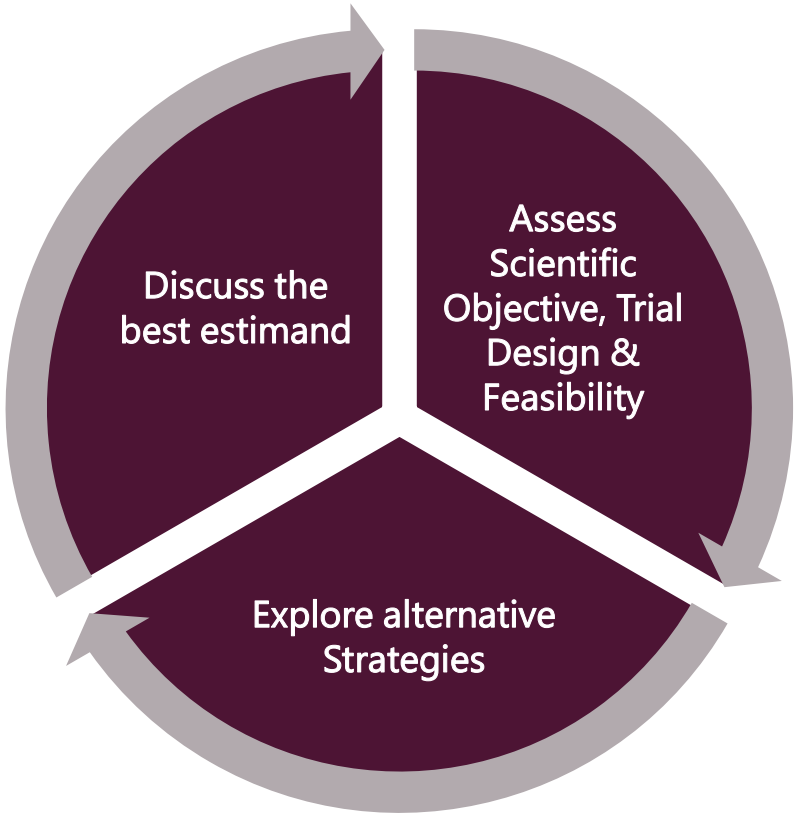
- ”Usually an iterative process will be necessary to reach an estimand that is of clinical relevance for decision making, and for which a reliable estimate can be made.” – ICH E9 R1
- The estimands guidance recognizes a balance between clinical relevance and reliability, with neither holding an absolute veto card.

Synthesis

- ”Whilst an inability to derive a reliable estimate might preclude certain choices of strategy, it is important to proceed sequentially from the trial objective and an understanding of the clinical question of interest, and not for the choice of data collection and method of analysis to determine the estimand.” – ICH E9 R1
- The estimands guidance gives weight to a clinically relevant question.
  - Trial conditions influence the set of possible estimands, but should not dictate them.
- The tradeoff involved is perhaps analogous between the trade-off between bias and variation common in statistical estimation.

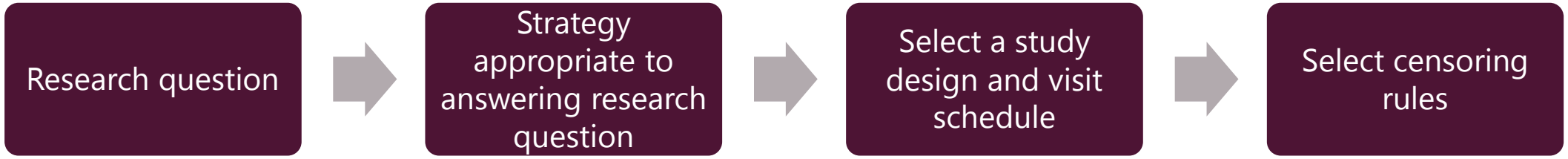
STRATEGY SELECTION AND FEASIBILITY

- A clear scientific objective will help in selecting estimands and strategies
- Team needs to assess the clinical relevance and feasibility of estimating the desired estimand in the proposed setting.
- Should the desired approach have feasibility issues, then alternative estimands and strategies should be considered
- There is often no perfect strategy. The initial or conventional strategy, even with feasibility issues, might still be better than the alternatives.

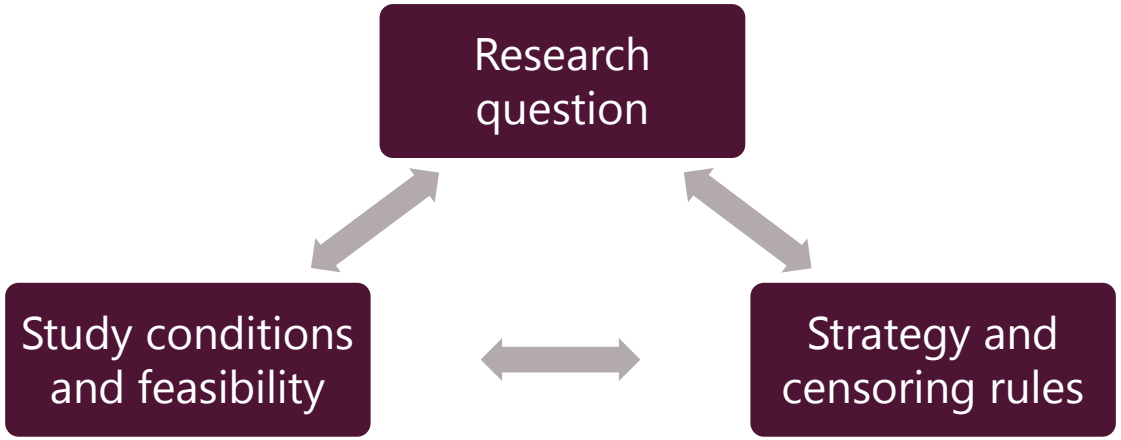


STUDY DESIGN MAY BECOME A LESS LINEAR PROCESS

- The ICH E9 [R1] guidance describes study design as a linear process:



- It may need to become more non-linear in practice.



The procedure needed to address this non-linearity may more closely resemble W. Edward's Deming's "Plan-Do-Study-Act" cycle (Deming, 1986) than a classic linear "waterfall" procedure.

Conclusions

- The estimands guidance emphasizes asking the right questions. Defining the clinically relevant question of interest lies at its core.
- In a regulatory trial, the scientific question involves what will happen in real-world clinical practice, not merely in a repeat of the trial conditions.
- Where subsequent therapy does not reflect real-world clinical practice in the decision context, it can be interpreted as representing an intercurrent event in the estimands framework, confounding inferability of the trial results to real-world clinical practice.
- The scientific question becomes what would have happened had the confounding intercurrent event not occurred

Conclusions (Cont)

- A hypothetical estimand is the most appropriate approach in the estimands framework for addressing what would have happened absent an intercurrent event.
- Accordingly, where subsequent therapy behavior is an artifact of the trial environment and does not reflect real-world practice, hypothetical estimands, despite their known flaws, should be considered for acceptance in registrational clinical trials.
- Choice of estimand will often reflect a tradeoff or iterative balancing between clinical relevance and operational feasibility including reliability.
- Clinical relevance is not the sole factor. But it needs to have more weight than it has in the past.
- Studies requiring hypothetical estimands should be designed as reliably as possible, with careful attention to data collection and sensitivity analyses.

Thank you!