Answering Old Questions with New Tools: Application of the ICH E9 Addendum in Oncology

Evgeny Degtyarev & Kaspar Rufibach
Novartis and Roche, Basel
76th Annual Deming Conference on Applied Statistics
9th December 2020

Acknowledgments

We borrowed from slides by

- **Hans-Jochen Weber** & **Renaud Capdeville**,
- **Björn Bornkamp**.

All our colleagues of the **industry working group on estimands in oncology**.

Keaven Anderson (Merck) and **Frank Bretz** (Novartis).

Regulatory colleagues around the world for regular discussion, their input, and feedback.

The intellectual illness of clinical drug evaluation that I have discussed here can be cured, and it will be cured when we restore intellectual primacy to the questions we ask, not the methods by which we answer them.

Lew Sheiner
American Clinical Pharmacologist

Sheiner (1991)

Agenda

1. ICH E9(R1) addendum: Why? And what's new?
2. Case study: hematology
3. Case study: CAR-T
4. Hypothetical strategy to address ICEs: application to Covid-19
5. Case study: treatment switching
6. Subgroups by post-randomization event - principal stratification
7. Impact and conclusions
8. Industry working group *Estimands in oncology*

Agenda

ICH E9 draft addendum

ICH E9: "Statistical principles for Clinical Trials."

1998.

Why amend E9?

Lack of alignment between trial objectives and reported effect quantification.

Example: Dapagliflozin

ICH E9 working group toy example, Hemmings (2015).

Dapagliflozin:

- Anti-diabetic therapy to treat hyperglycemia.
- Discussed in 2011 in a public advisory committee at **FDA**.

Trial objective: Assess whether drug works compared to placebo.

Example: Dapagliflozin

	Sponsor	FDA
Proposed analysis	Remove data after rescue.	Use all data, irrespective of rescue.
Implied scientific question	Treatment effect of the initially randomized treatments **had no patient received rescue medication**.	Compare treatment **policies** "dapagliflozin + rescue" vs. "control + rescue".

What is going on?

- Implied objectives / scientific questions of interest **differ for sponsor and regulator**.
- Discussion only at time of **filing**, while this is actually a **design** question!
- Estimand hidden behind the method of estimation / handling of missing data $\Rightarrow$ statistics section defines trial objective!

"How should we handle missing data?" becomes
"What question are we really interested to answer?"

What is a “treatment effect”?

Treatment effect

Not defined in original E9!

How outcome compares to what would have happened to same subject under alternative treatment, e.g. had they

- **not** received treatment,
- received a **different** treatment.

Potential outcome $\Rightarrow$ causal inference!

Estimate average treatment effect from **randomized clinical trial**.

Understanding treatment effects

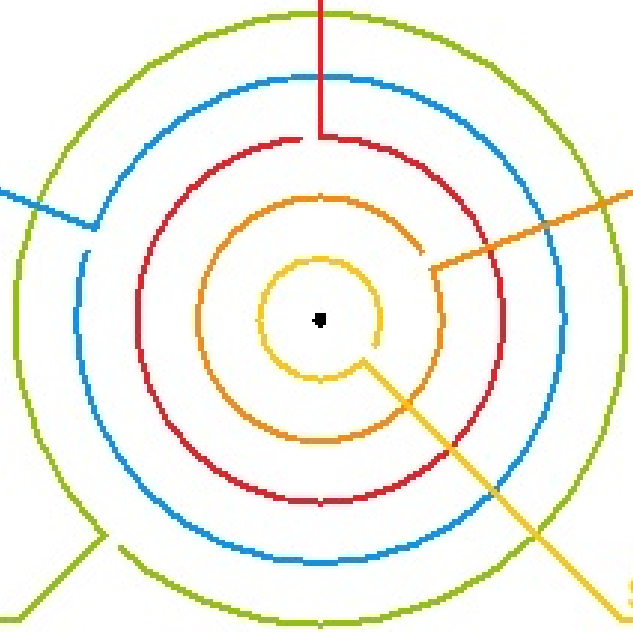
- Multiple definitions of **treatment effect**.
- Different definitions addressing **different scientific questions**.
- Not all equally acceptable for **regulatory decision making**.
- Not all alternatives can be reliably estimated! **Iterative** process of estimand - estimator definition.
- Stakeholders: regulators, HTA / payers, phyisicians, patients $\Rightarrow$ all need to **make decisions**.

How does the addendum fix this?

More precise definition of trial objective
$\Rightarrow$ estimand!

ESTIMAND

TARGET OF ESTIMATION



VARIABLE

The variable (or endpoint) to be obtained for each patient

POPULATION

The population of patients targeted by the clinical question

INTERCURRENT EVENTS

Other intercurrent events (not already addressed by treatment, population, and variable) and how they are addressed

TREATMENT

The treatment condition of interest

SUMMARY

A population-level summary for the variable which provides a basis for treatment comparison

Objective pre- and post-addendum

Pre:

Treatment difference between Gazyva and Rituximab on PFS.

Post:

The trial will compare 6 or 8 21-day cycles obinutuzumab D1 + C1D8, C1D15: 1000mg/m2 flat + site-specific choice of CT (CVP, Benda, CHOP) in induction followed in responding patients by 1000mg flat every 2 months until PD or up to 2y with 6 or 8 21-day cycles rituximab 375mg/m2 D1 + site-specific choice of CT (CVP, Benda, CHOP) in induction followed in responding patients by 375mg/m2 every 2 months until PD or up to 2y in first-line follicular lymphoma patients.

The primary comparison of interest is the hazard ratio of progression-free survival. The primary trial objective is to demonstrate superiority of the experimental over the control treatment.

The primary comparison of progression-free survival will be made regardless of whether patients withdraw from treatment or receive new-anti lymphoma therapy prior to disease progression.

Objective pre- and post-addendum

Pre:

Treatment difference between Gazyva and Rituximab on PFS.

Post:

The trial will compare 6 or 8 21-day cycles obinutuzumab D1 + C1D8, C1D15: 1000mg/m2 flat + site-specific choice of CT (CVP, Benda, CHOP) in induction followed in responding patients by 1000mg flat every 2 months until PD or up to 2y with 6 or 8 21-day cycles rituximab 375mg/m2 D1 + site-specific choice of CT (CVP, Benda, CHOP) in induction followed in responding patients by 375mg/m2 every 2 months until PD or up to 2y in first-line follicular lymphoma patients.

The primary comparison of interest is the hazard ratio of progression-free survival. The primary trial objective is to demonstrate superiority of the experimental over the control treatment.

The primary comparison of progression-free survival will be made regardless of whether patients withdraw from treatment or receive new-anti lymphoma therapy prior to disease progression.

Objective pre- and post-addendum

Pre:

Treatment difference between Gazyva and Rituximab on PFS.

Post:

The trial will compare 6 or 8 21-day cycles obinutuzumab D1 + C1D8, C1D15: 1000mg/m2 flat + site-specific choice of CT (CVP, Benda, CHOP) in induction followed in responding patients by 1000mg flat every 2 months until PD or up to 2y with 6 or 8 21-day cycles rituximab 375mg/m2 D1 + site-specific choice of CT (CVP, Benda, CHOP) in induction followed in responding patients by 375mg/m2 every 2 months until PD or up to 2y in first-line follicular lymphoma patients.

The primary comparison of interest is the hazard ratio of progression-free survival. The primary trial objective is to demonstrate superiority of the experimental over the control treatment.

The primary comparison of progression-free survival will be made regardless of whether patients withdraw from treatment or receive new-anti lymphoma therapy prior to disease progression.

Objective pre- and post-addendum

Pre:

Treatment difference between Gazyva and Rituximab on PFS.

Post:

The trial will compare 6 or 8 21-day cycles obinutuzumab D1 + C1D8, C1D15: 1000mg/m2 flat + site-specific choice of CT (CVP, Benda, CHOP) in induction followed in responding patients by 1000mg flat every 2 months until PD or up to 2y with 6 or 8 21-day cycles rituximab 375mg/m2 D1 + site-specific choice of CT (CVP, Benda, CHOP) in induction followed in responding patients by 375mg/m2 every 2 months until PD or up to 2y in first-line follicular lymphoma patients.

The primary comparison of interest is the hazard ratio of progression-free survival. The primary trial objective is to demonstrate superiority of the experimental over the control treatment.

The primary comparison of progression-free survival will be made regardless of whether patients withdraw from treatment or receive new-anti lymphoma therapy prior to disease progression.

Objective pre- and post-addendum

Pre:

Treatment difference between Gazyva and Rituximab on PFS.

Post:

The trial will compare 6 or 8 21-day cycles obinutuzumab D1 + C1D8, C1D15: 1000mg/m2 flat + site-specific choice of CT (CVP, Benda, CHOP) in induction followed in responding patients by 1000mg flat every 2 months until PD or up to 2y with 6 or 8 21-day cycles rituximab 375mg/m2 D1 + site-specific choice of CT (CVP, Benda, CHOP) in induction followed in responding patients by 375mg/m2 every 2 months until PD or up to 2y in first-line follicular lymphoma patients.

The primary comparison of interest is the hazard ratio of progression-free survival. The primary trial objective is to demonstrate superiority of the experimental over the control treatment.

The primary comparison of progression-free survival will be made regardless of whether patients withdraw from treatment or receive new-anti lymphoma therapy prior to disease progression.

Objective pre- and post-addendum

Pre:

Treatment difference between Gazyva and Rituximab on PFS.

Post:

The trial will compare 6 or 8 21-day cycles obinutuzumab D1 + C1D8, C1D15: 1000mg/m2 flat + site-specific choice of CT (CVP, Benda, CHOP) in induction followed in responding patients by 1000mg flat every 2 months until PD or up to 2y with 6 or 8 21-day cycles rituximab 375mg/m2 D1 + site-specific choice of CT (CVP, Benda, CHOP) in induction followed in responding patients by 375mg/m2 every 2 months until PD or up to 2y in first-line follicular lymphoma patients.

The primary comparison of interest is the hazard ratio of progression-free survival. The primary trial objective is to demonstrate superiority of the experimental over the control treatment.

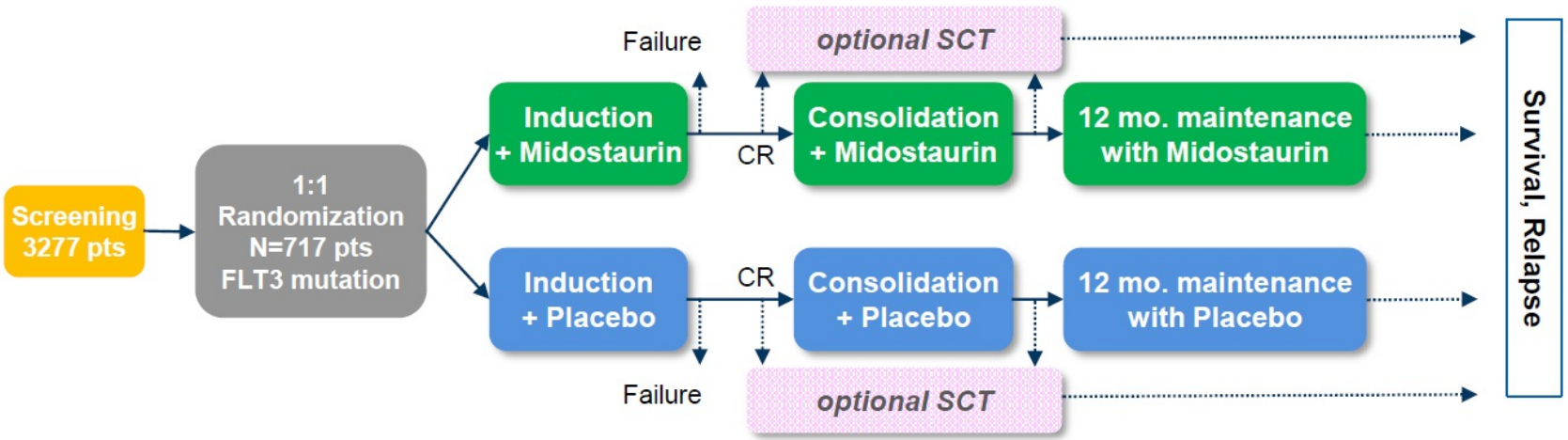
The primary comparison of progression-free survival will be made regardless of whether patients withdraw from treatment or receive new-anti lymphoma therapy prior to disease progression.

Estimand follows from precise trial objective (or vice-versa).

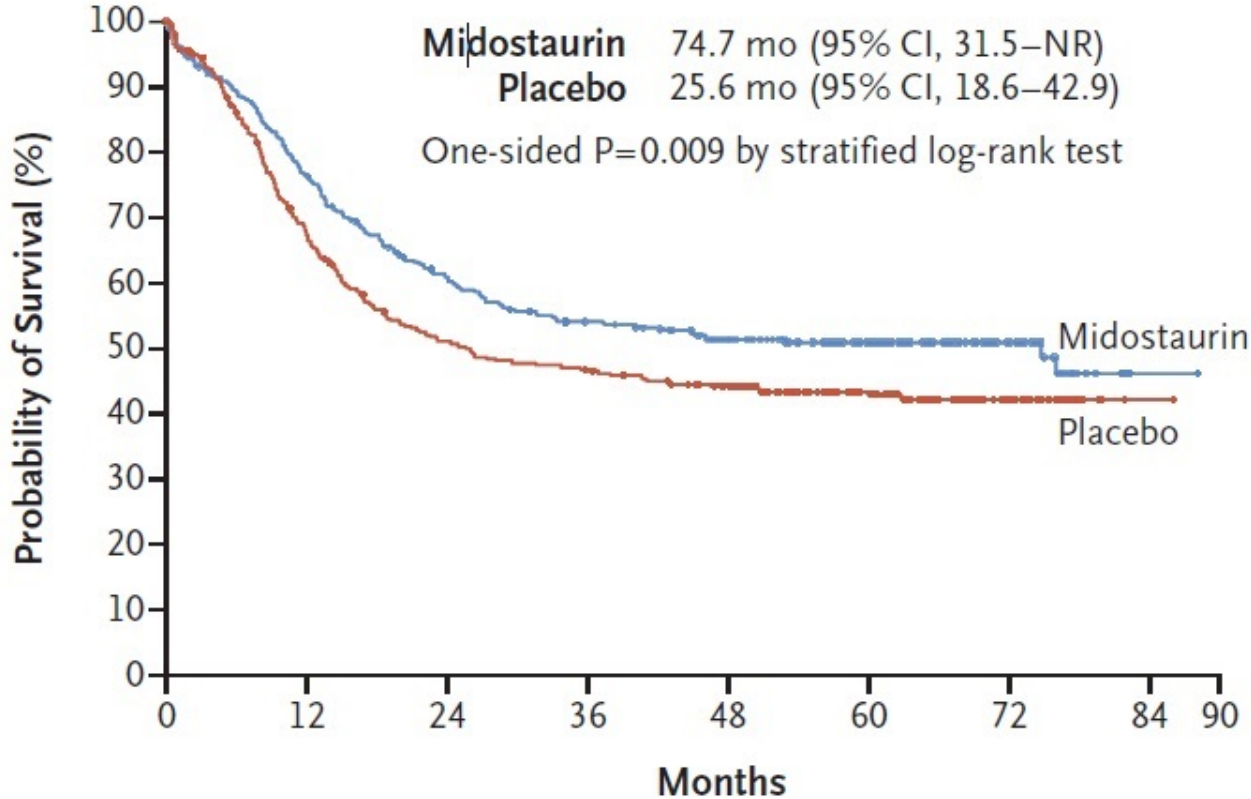
Agenda

Complex treatment strategies in hematology

Ratify trial, Stone *et al.* (2017).



- **Randomized, phase III**, open-label, double-blind clinical trial.
- **Population**: newly diagnosed AML with a FLT 3 mutation.
- **Comparison**: after completion of primary therapy: Midostaurin vs. placebo.
- **Primary endpoint**: OS.
- **Key secondary endpoint**: EFS.



No. at Risk

Midostaurin	360	269	208	181	151	97	37	1
Placebo	357	221	163	147	129	80	30	1

OS was significantly longer in the midostaurin group than in the placebo group, as was EFS. [...] In both the primary analysis and an analysis in which **data for patients who underwent transplantation were censored**, *the benefit of midostaurin was consistent across all FLT3 subtypes.*

What question are we asking?

Protocol objective: To determine if the addition of midostaurin to induction, consolidation, and maintenance therapy improves OS in mutant AML patients.

- **Primary analysis**: survival regardless of receiving SCT or maintenance
  $\Rightarrow$ treatment effect = if SCT is part of treatment strategy.
- **Sensitivity analysis**: censoring at transplant $\Rightarrow$ treatment effect = **hypothetical** estimand strategy, if no SCT was given. Estimand is **implicit**!

Completely different clinical questions!

What question are we asking?

Protocol objective: To determine if the addition of midostaurin to induction, consolidation, and maintenance therapy improves OS in mutant AML patients.

What ended up in the label?

- **SmPC**: In combination with **induction** and **consolidation**, and for patients in complete response followed by single agent **maintenance** therapy.
- **USPI**: In combination with standard **induction** and **consolidation**.

AML: treatment strategy based on sequence of

- multiple decision points and
- treatment modalities.

RATIFY:

- Despite detailed description of objectives and treatment in protocol
  $\Rightarrow$ **insufficient alignment** on underlying question of interest.
- SCT:
  - Component of treatment strategy with potential major impact on B/R.
  - Impact not clearly outlined in trial objective.
- Maintenance: Despite explicit inclusion in trial objective $\Rightarrow$ **inconsistently included in approved labels EMA and FDA**.

How would we define the estimand today?

Clinical trial objective: To determine if the addition of midostaurin to induction, consolidation, and maintenance therapy with the option to receive SCT in CR improves OS in mutant AML patients.

Treatment strategy:

- Experimental: DNR AraC + midostaurin induction, AraC + midostaurin consolidation in pts with a CR, midostaurin maintenance, option to receive SCT in CR.
- Control: DNR AraC induction, AraC consolidation in pts with a CR, option to receive SCT in CR.

Population: newly diagnosed AML with a FLT 3 mutation eligible for intensive chemotherapy.

Variable: OS.

Intercurrent events: none left for OS - all integrated in treatment strategy attribute.

Summary measure: hazard ratio.

Complex (multiphase) strategies:

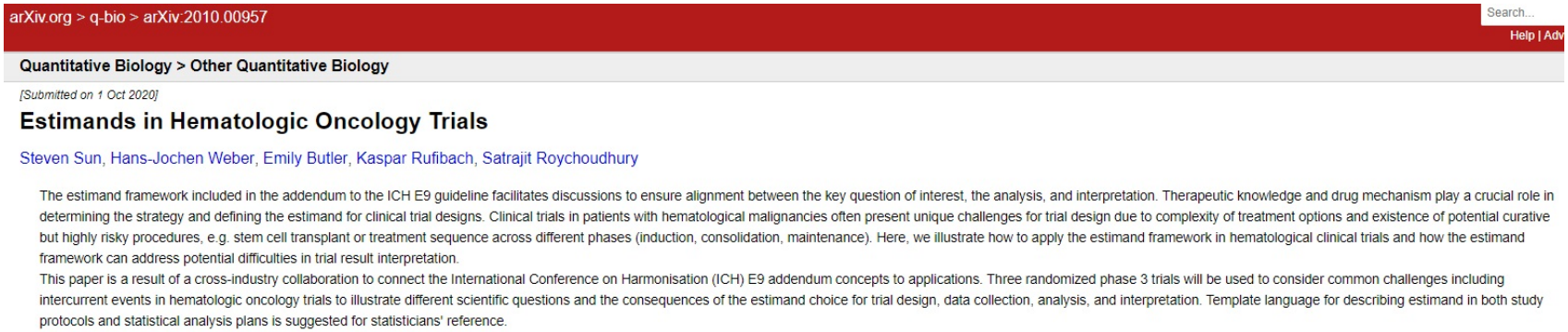
Non-proportional hazards?

Cure?

What do these findings have in common?

They can all be anticipated!

Clear formulation of
clinical trial objective is key.



arXiv.org > q-bio > arXiv:2010.00957

Search...

Help | Adv

Quantitative Biology > Other Quantitative Biology

[Submitted on 1 Oct 2020]

Estimands in Hematologic Oncology Trials

Steven Sun, Hans-Jochen Weber, Emily Butler, Kaspar Rufibach, Satrajit Roychoudhury

The estimand framework included in the addendum to the ICH E9 guideline facilitates discussions to ensure alignment between the key question of interest, the analysis, and interpretation. Therapeutic knowledge and drug mechanism play a crucial role in determining the strategy and defining the estimand for clinical trial designs. Clinical trials in patients with hematological malignancies often present unique challenges for trial design due to complexity of treatment options and existence of potential curative but highly risky procedures, e.g. stem cell transplant or treatment sequence across different phases (induction, consolidation, maintenance). Here, we illustrate how to apply the estimand framework in hematological clinical trials and how the estimand framework can address potential difficulties in trial result interpretation.

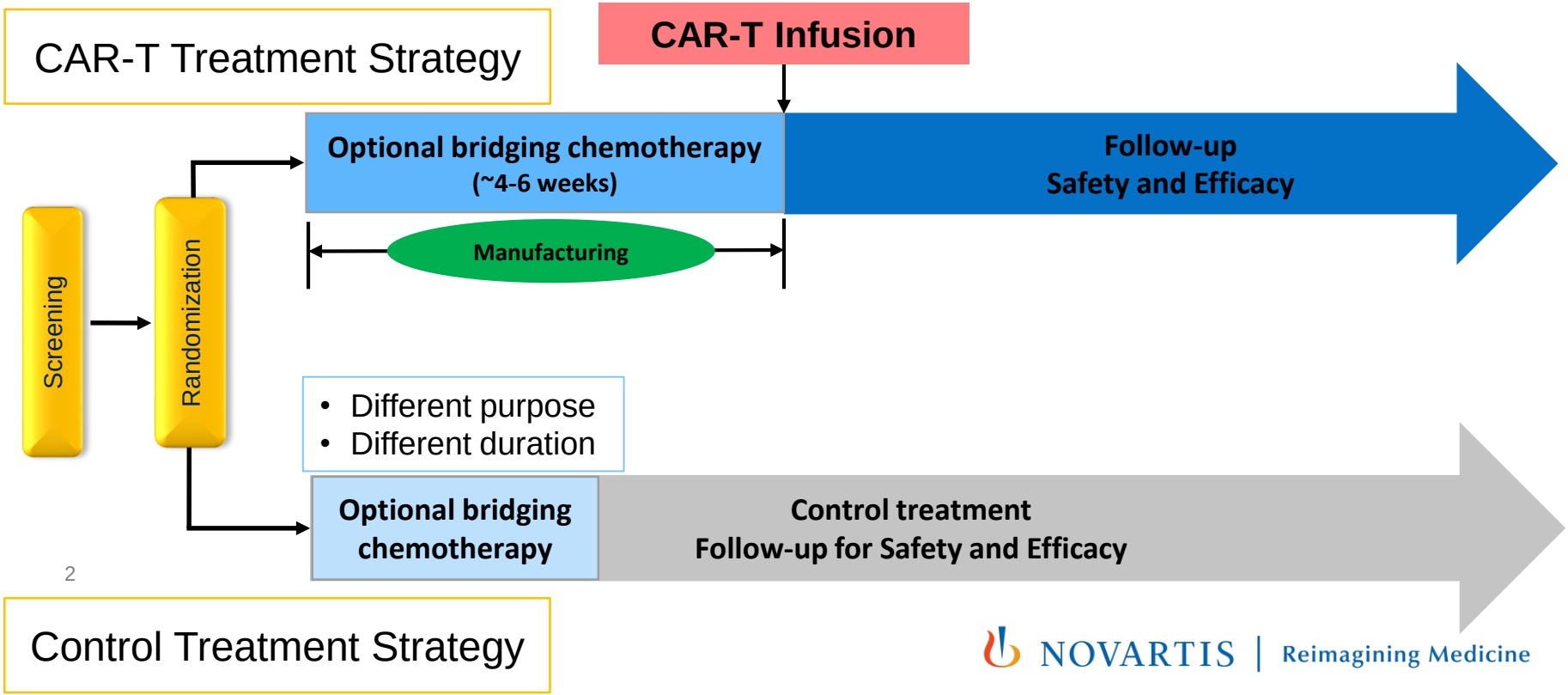
This paper is a result of a cross-industry collaboration to connect the International Conference on Harmonisation (ICH) E9 addendum concepts to applications. Three randomized phase 3 trials will be used to consider common challenges including intercurrent events in hematologic oncology trials to illustrate different scientific questions and the consequences of the estimand choice for trial design, data collection, analysis, and interpretation. Template language for describing estimand in both study protocols and statistical analysis plans is suggested for statisticians' reference.

Sun *et al.* (2020):

- Three case studies.
- Categorization and discussion of sensitivity and supplementary analyses.
- Templates for protocol and SAP.

Agenda

Randomized study comparing two treatment strategies



NOVARTIS | Reimagining Medicine

FDA Comment on the protocol

FDA Comment

Subjects in the CAR-T arm may receive extensive bridging chemotherapy while awaiting CAR-T manufacture, and some, especially those experiencing extended delays in product manufacture, could achieve a CR/CRi [...] status in response to aggressive bridging chemotherapy even before initiation of CAR-T treatment. Since these response cannot be directly attributed to CAR-T treatment, the statistical assessment plan should should prospectively create rules for appropriately censoring CR [...] subjects.

Censoring implying hypothetical estimand

- FDA proposal for supplementary EFS analysis: add specific rule for CAR-T arm to censor patients who are responding to bridging chemotherapy

- Targeting hypothetical scenario in which no patient would respond to bridging chemotherapy in CAR-T arm

- Is this estimand relevant for patients, physicians and regulators?

Getting the questions right

- Sponsor realized that requested analysis does not address a relevant question of interest
- Sponsor suggested that principal stratum estimand would address FDA's actual question of interest
- Question of interest: What is the effect of the CAR-T treatment strategy relative to control treatment strategy on EFS in patients who would not respond to bridging chemotherapy if they were given bridging chemotherapy for CAR-T?
- FDA agreed to use the principal stratum strategy as supplementary analysis instead of censoring

Using communication as a guide

- Sometimes useful to think ahead to labeling
  - Picture yourself telling a patient the effect in the label is what they should care about
- "This is the mean effect among people who wouldn't need rescue" vs.
- "This is the mean effect in everyone if rescue medication didn't exist"

- Effect among patients who would not respond to bridging chemotherapy?
- Effect if no patient would respond to bridging chemotherapy?

John Scott at ASA&EFSPI&PSI Webinar on Estimands, https://www.psiweb.org/vod/item/joint-psi-efspi-asa-biop-webinar-estimands

Principal stratum: Opportunities

- Improved HA interactions discussing questions of interest and not censoring rules resulting in more meaningful analyses
  - Estimand framework provides common language to discuss questions of interest and to do more meaningful analyses
- Opportunity for regulators and sponsors to learn together and to collaborate with academia addressing important questions
  - many examples of practical relevance in drug development
- Further examples and more details on the analysis in the second part of the talk

Agenda

Hypothetical estimands

- ICH E9(R1) addendum acknowledges that some hypothetical scenarios are likely to be of more clinical or regulatory interest than others
  - previously shown CAR-T example: less relevant hypothetical scenario

- Hypothetical estimands often implicitly targeted by primary analysis in pivotal trials
  - PFS analysis censoring new anticancer therapies per FDA guideline
  - proposed in EMA guidelines for Alzheimer or Diabetes
- Two other relevant examples for hypothetical estimand follow

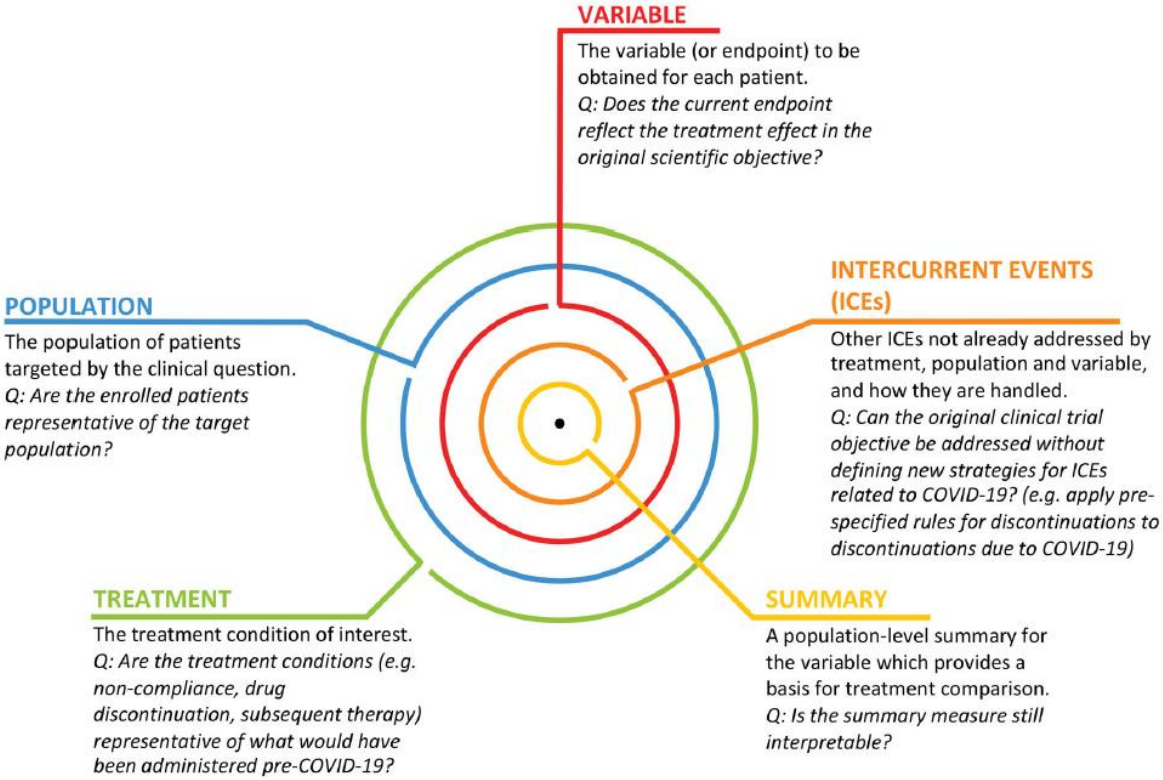
COVID-19 and estimands

- primary intention of the ICH E9 addendum: alignment between clinical trial objectives and treatment effect estimation *prior to* the start of a trial
- ICH E9 addendum also specific for unforeseen events during the trial:

"Addressing intercurrent events that were not foreseen at the design stage, and are identified during the conduct of the trial, should discuss not only the choices made for the analysis, but the effect on the estimand, that is, on the description of the treatment effect that is being estimated, and the interpretation of the trial results. "

- Framework useful to discuss the impact of COVID-19 on ongoing and future trials

Assessing impact of COVID-19 on estimand



NOVARTIS | Reimagining Medicine

COVID-19 and hypothetical estimand

- Ongoing trials designed implicitly assuming no major disruption of healthcare systems and absence of a highly infectious disease with severe complications and for which no effective therapy is available

→ Trial objectives should relate to a world without COVID-19 pandemic

→e.g. hypothetical strategy reasonable for intercurrent events primarily caused by the disruption of healthcare systems or patients' desire to minimize traveling independently of disease or treatment

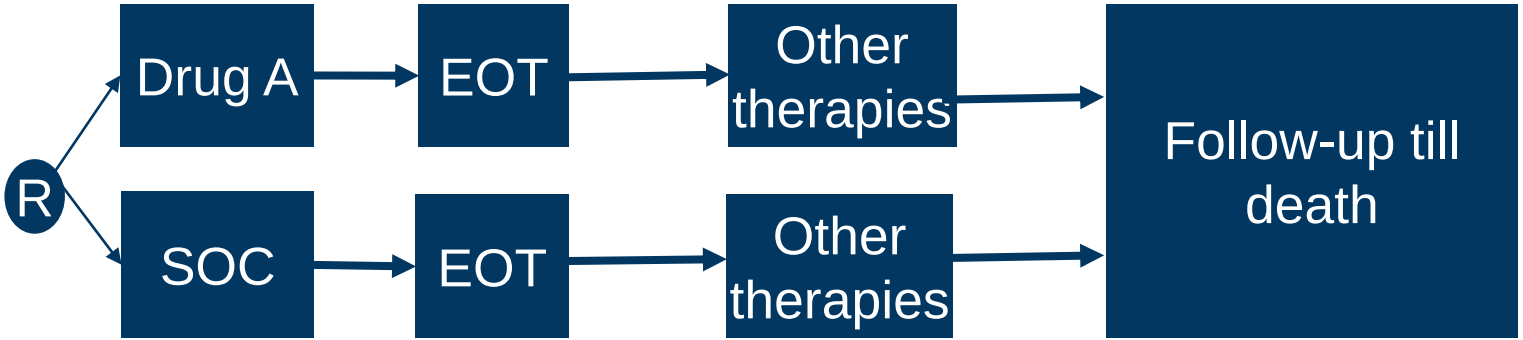
Implications on analysis?

- Change in estimand not always requires change in analysis
- Estimates from initially planned analysis may still be sufficiently precise to address the objective to assess effect in a world without COVID-19 pandemic
- Focus on questions of interest results in more clarity in interpretation regardless of whether there is a change in analysis

Agenda

Overall survival (OS) in clinical trials

Treatment Switching



OS usually analyzed using treatment policy strategy

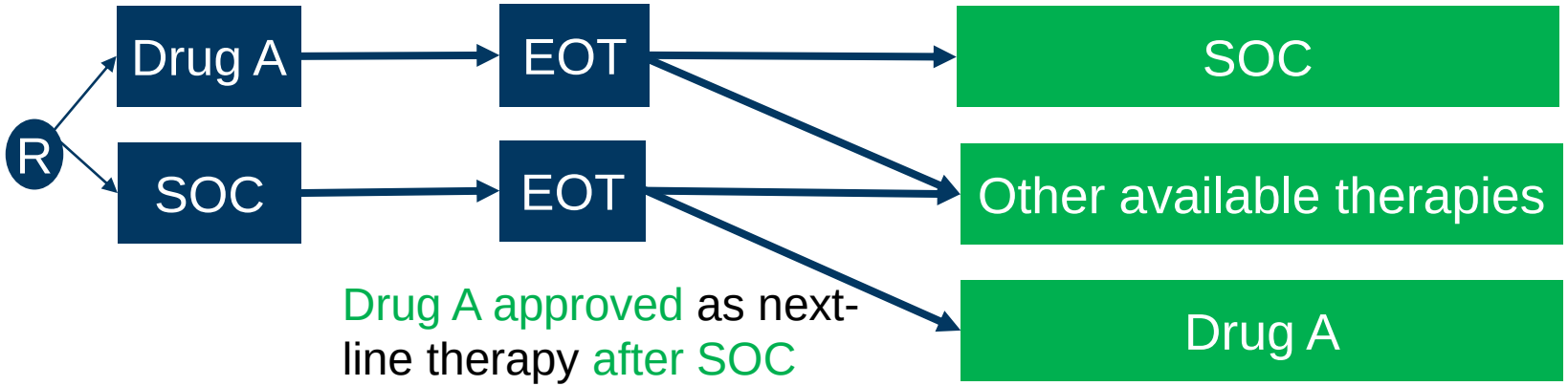
- using time from randomization to death regardless of patient's journey
- captures effect on the choice and impact of subsequent therapies
- balance in subsequent therapies generally not expected as physician choose subsequent therapy in light of previously administered therapies
- clinically meaningful if choice of subsequent therapies after EOT reflects clinical practice

NOVARTIS | Reimagining Medicine

SOC: Standard of Care; EOT: End of treatment

Overall survival (OS) in clinical trials

Treatment Switching



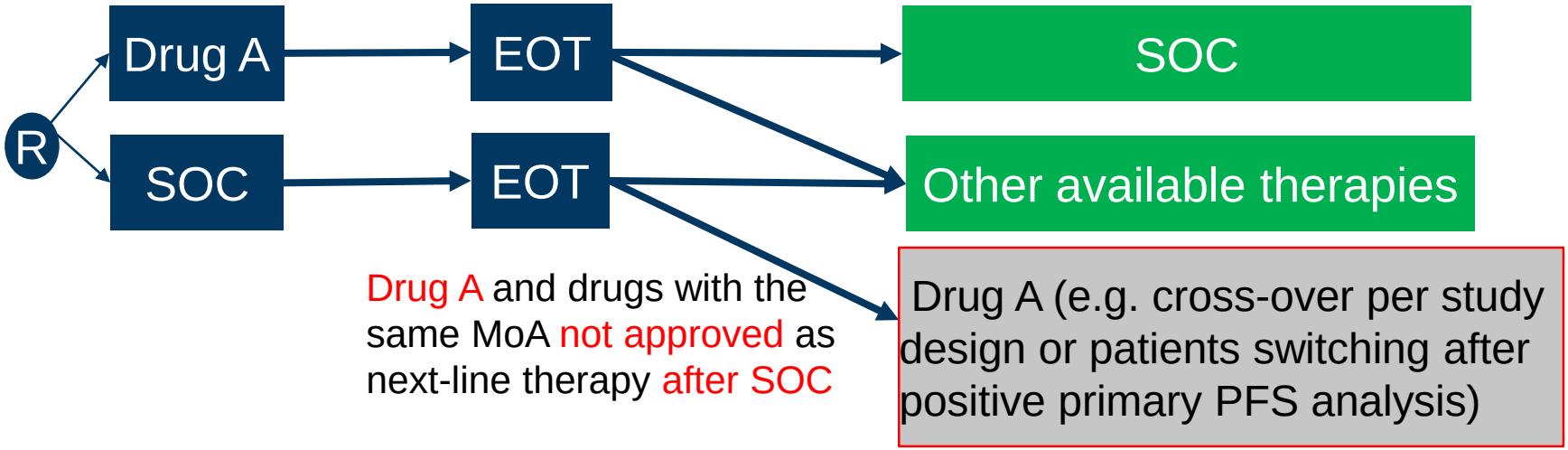
☺ choice of subsequent therapies after EOT reflects clinical practice

→ Treatment policy OS estimand interpretable at the time of the readout

NOVARTIS | Reimagining Medicine

Overall survival (OS) in clinical trials

Treatment Switching



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

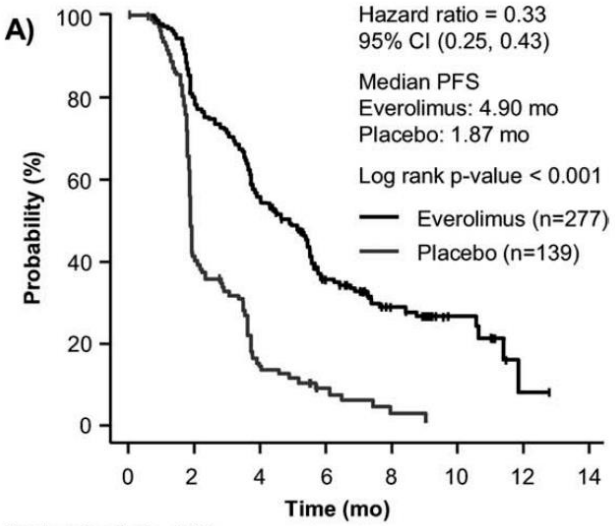
→ Treatment policy estimand comparing Drug A followed by SOC or other available therapies vs SOC followed by Drug A or other available therapies relevant?

NOVARTIS | Reimagining Medicine

SOC: Standard of Care; EOT: End of Treatment; PFS: Progression-Free Survival

If subsequent therapies do not reflect clinical practice… Trial results are difficult to interpret

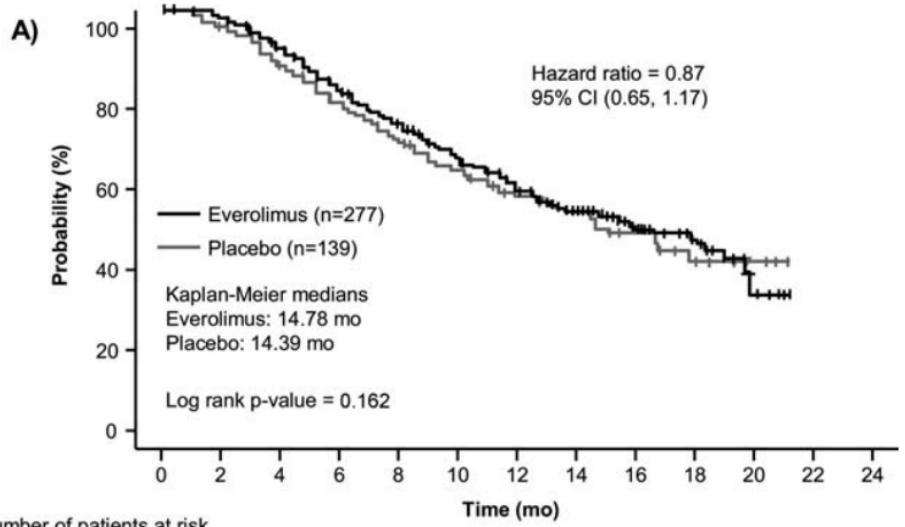
PFS



Number of patients at risk

Everolimus	277	192	115	51	26	10	1	0
Placebo	139	47	15	6	2	0	0	0

OS



Number of patients at risk

Everolimus	277	267	240	204	164	155	131	101	61	30	6	0	0
Placebo	139	131	117	100	88	74	56	43	27	13	3	0	0

PD=Progression

NOVARTIS | Reimagining Medicine

If subsequent therapies do not reflect clinical practice… OS description in labels is ambiguous

Regorafenib US Prescribing Information

A statistically significant improvement in PFS was demonstrated among patients treated with STIVARGA compared to placebo (see Table 8 and Figure 2).

There was no statistically significant difference in overall survival at the final OS analysis, conducted at 162 OS events (Table 8). Cross-over to open label STIVARGA occurred in 58 (88%) placebo-treated patients after disease progression.

Nivolumab Summary of Product Characteristics:

There was no statistically significant difference between nivolumab and chemotherapy in the final OS analysis. The primary OS analysis was not adjusted to account for subsequent therapies, with 54 (40.6%) patients in the chemotherapy arm subsequently receiving an anti-PD1 treatment. OS may be confounded by dropout, imbalance of subsequent therapies and differences in baseline factors.

NOVARTIS | Reimagining Medicine

If subsequent therapies do not reflect clinical practice… Drugs are perceived as not improving survival

Over half of new cancer drugs 'show no benefits' for survival or wellbeing

Of 48 cancer drugs approved between 2009-2013, 57% of uses showed no benefits and some benefits were 'clinically meaningless', says BMJ study

PHARMALOT

STAT+

Flawed trials supported half of recent approvals of cancer drugs in Europe, study says

By ED SILVERMAN @Pharmalot / SEPTEMBER 18, 2019

6:36pm, Sep 19, 2019

LIFE • WELLBEING •

Poorly designed cancer drug trials may be exaggerating benefits

NOVARTIS | Reimagining Medicine

If subsequent therapies do not reflect clinical practice…
Regulatory standards are perceived to be low

THE MILBANK QUARTERLY
A MULTIDISCIPLINARY JOURNAL OF POPULATION HEALTH AND HEALTH POLICY

Original Scholarship | Open Access

Approval of Cancer Drugs With Uncertain Therapeutic Value: A Comparison of Regulatory Decisions in Europe and the United States

MAXIMILIAN SALCHER-KONRAD, HUSEYIN NACI, COURTNEY DAVIS

First published: 06 October 2020 | https://doi.org/10.1111/1468-0009.12476

Conclusions: US and European regulators often deemed early and less complete evidence on benefit-risk profiles of cancer drugs sufficient to grant regular approval, raising questions over regulatory standards for the approval of new medicines. Even when imposing confirmatory studies in the postmarket-

ELSEVIER
European Journal of Cancer
Volume 136, September 2020, Pages 176-185

Original Research

Progression-free survival is a suboptimal predictor for overall survival among metastatic solid tumour clinical trials

If subsequent therapies do not reflect clinical practice… Hypothetical strategy represents key question of interest!

A statistically significant improvement in PFS was demonstrated among patients treated with STIVARGA compared to placebo (see Table 8 and Figure 2).

There was no statistically significant difference in overall survival at the final OS analysis, conducted at 162 OS events (Table 8). Cross-over to open label STIVARGA occurred in 58 (88%) placebo-treated patients after disease progression.

- Would it not be more relevant for patients and prescribers to see in the label the effect of STIVARGA on OS if placebo-treated patients did not have the possibility to cross-over to STIVARGA after disease progression?
  - hypothetical strategy for cross-over

NOVARTIS | Reimagining Medicine

Hypothetical strategy: analysis

- Statistical methods such as IPCW can answer this question if properly planned (incl. data collection)

- Facing some headwinds as the methods rely on assumptions and many of us are not experienced with this methodology

- Opportunity for sponsors and regulators to learn together and to collaborate with academia to address important questions for patients!
  - need to develop best practices for various aspects from implementation to data collection

NOVARTIS | Reimagining Medicine

Conclusions

- Treatment policy estimand will be the main question of interest for patients and physicians with regard to OS in vast majority of the situations
- In some settings hypothetical strategy appears to be more meaningful
- Estimand framework provides us the opportunity
  - discuss alternatives to main OS analysis addressing relevant questions for patients and prescribers
  - to improve communication between physician and patient by improving OS description in the labels and publications
  - to communicate added value of our drugs better
- Opportunity for regulators and sponsors to learn together and to collaborate with academia addressing important questions

Agenda

"... The target population might be taken to be the "principal stratum" in which an **intercurrent event would occur**. Alternatively, the target population might be taken to be the principal stratum in which an intercurrent event **would not occur**. The clinical question of interest relates to the treatment effect only within the principal stratum..."

ICH E9 working group (2019)

Principal stratification:

- Originates in causal inference: Frangakis and Rubin (2002).
- Framework for comparing treatments adjusting for **posttreatment** variables.
- Yields principal effects which are **causal** effects within a principal stratum.

First, let us summarize what does not work.

2-arm RCT experimental (E) vs. control (C)

Do patients that are ADA+
in E have lower treatment effect?

"Subgroup" built by post-randomization event!

How can we make valid causal statements?

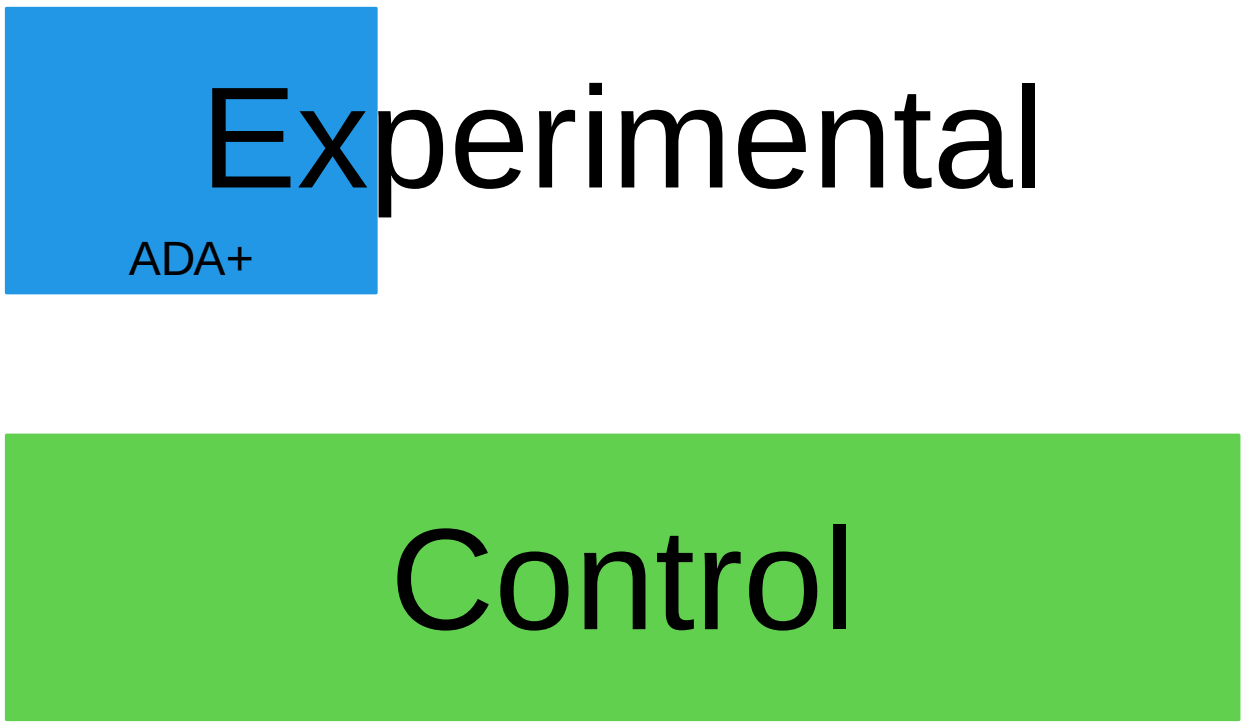
Need "matched control patients"!

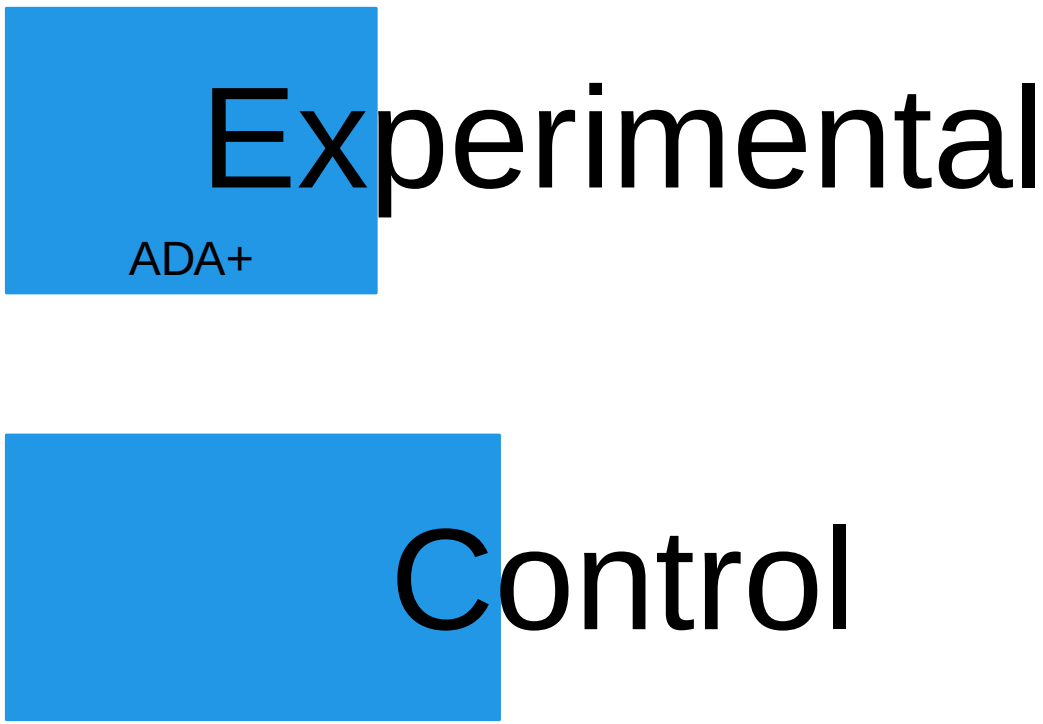
Experimental

Control


Experimental
ADA+
ADA−

Patients randomized to E
experiencing ADA+
had they received control


Experimental
ADA+
Control


Experimental
ADA+
Control


Experimental
ADA+
ADA−

For every complex problem, there is a solution that is simple, neat, and wrong.

H.L. Mencken, American Journalist

Naive analyses are misleading and do not answer causal question

Principal stratification: "subgroup analysis for post-baseline subgroups"

randomization + assumptions

Are such questions relevant?

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 prespecified 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 experimental drug	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 prespecified 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

Bornkamp *et al.* (2020).

Potential outcomes and principal stratification

$$Z := \begin{cases} 1 & \text{test treatment} \\ 0 & \text{control treatment.} \end{cases}$$

Y: outcome (binary, continuous, time-to-event).

Ideal world: treating physician decides on treatment based on outcome if given

- control treatment: $Y(Z = 0) = Y(0)$,
- test treatment, $Y(Z = 1) = Y(1)$.

Neither $Y(0)$ nor $Y(1)$ known when assigning treatment!

Only one observed at all $\Rightarrow$ **individual causal effects** $Y(1) - Y(0)$ not observed.

Population level: targets average causal effect $E(Y(1) - Y(0))$.

Estimation of average causal effect

RCT:

- **Exchangeability**: treatment assignment independent of patient characteristic.
- $Y(1)$ and $Y(0)$ independent of Z, implying that:

$$\begin{aligned} E(Y(1) - Y(0)) &= E(Y(1)) - E(Y(0)) \\ &= E(Y(1)|Z = 1) - E(Y(0)|Z = 0) \\ &= E(Y|Z = 1) - E(Y|Z = 0). \end{aligned}$$

Observational study:

- Decision between $Z = 0$ and $Z = 1$ might depend on X (measured or unmeasured).
- Patients who receive $Z = 1$ (for whom we observe $Y(1)$) might be **systematically different** from those who receive $Z = 0$ (for whom we observe $Y(0)$).
- $Y(1)$ and $Y(0)$ **not** independent of Z.
- $E(Y(1)) \neq E(Y(1)|Z = 1)$ and $E(Y(0)) \neq E(Y(0)|Z = 0)$
- Patients receiving $Z = 0$ **not representative** of overall population.

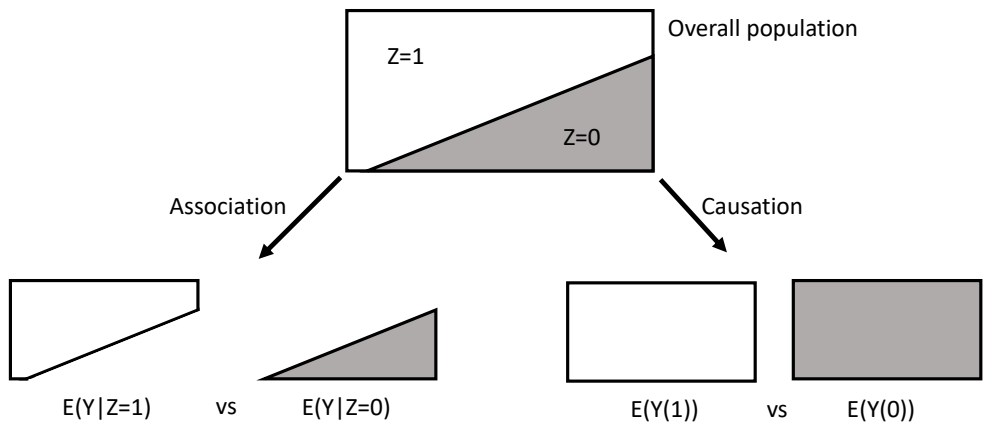
What are causal effects?

$Y(1)_i$: potential outcome for patient i.

$\mathcal{S}$: population of patients.

Causal treatment effect:

- Comparison of $\{Y(1)_i, i \in \mathcal{S}\}$ vs. $\{Y(0)_i, i \in \mathcal{S}\}$.
- Compare outcomes "had everyone received treatment" vs. outcomes "had everyone received control".



Naive analysis

Not a causal effect: comparison of $\{Y(1)_i, i \in \mathcal{S}_1\}$ vs. $\{Y(0)_i, i \in \mathcal{S}_2\}$ with $\mathcal{S}_1 \neq \mathcal{S}_2$.

Naive analysis: Let S = indicator variable for intercurrent, e.g. ADA+.

- Compare patients with $S = 1$ on both test and control arm.
- RCT: $S(Z)$ **post-randomization** $\Rightarrow$ S **depends on** Z!
- We observe $S(Z = 1)$ on test and $S(Z = 0)$ on control $\Rightarrow$ population of patients with $S(1) = 1$ and $S(0) = 1$ might be **quite different**!
- Breaks randomization $\Rightarrow$ not comparing "like with like" $\Rightarrow$ **not estimating causal effect**.
- Numerically observe a treatment effect in naive analysis $\Rightarrow$ not clear whether
  - due to different treatments or
  - due to difference in compared populations.
- Estimates treatment effect in principal stratum $\{S(1) = 1\} \cap \{S(0) = 1\}$ assuming $S(1) = S(0) \Rightarrow$ post-randomization event not treatment related. Assumption quite strong and **rarely justified**!

Principal stratification

Idea: stratify patients based on potential outcomes $S(0), S(1)$ for **all** treatments.

	$S(0) = 1$	$S(0) = 0$
$S(1) = 1$	$\{S(1) = 1\} \cap \{S(0) = 1\}$	$\{S(1) = 1\} \cap \{S(0) = 0\}$
$S(1) = 0$	$\{S(1) = 0\} \cap \{S(0) = 1\}$	$\{S(1) = 0\} \cap \{S(0) = 0\}$

Causal interpretation:

- Stratify population according to the **same rule on treatment and control arm**.
- Possible since membership to principal stratum **fixed at baseline**, not affected by treatment assignment.

Caveat:

- For patients on test arm we observe $S(1)$, but not $S(0)$, and vice versa for patients on control arm.
- **Identification** of patients in strata of interest generally not possible, not even after observing Y and S in a given trial.

Example: antidrug antibodies in immunotherapies

- Biological drugs: may trigger immune responses $\Rightarrow$ formation of **antidrug antibodies** (ADAs).
- Scientific question: Do patients that develop ADAs still benefit from the drug?
- Y: PFS or OS.
- S: occurrence of ADA at x weeks, say $x = 4$.
- Depending on test and control treatment $\Rightarrow$ ADA only in test arm.

	$S(0) = 1$	$S(0) = 0$
$S(1) = 1$	$\{S(1) = 1\} \cap \{S(0) = 1\}$	$\{S(1) = 1\} \cap \{S(0) = 0\}$
$S(1) = 0$	$\{S(1) = 0\} \cap \{S(0) = 1\}$	$\{S(1) = 0\} \cap \{S(0) = 0\}$

	ADA+ under control	ADA- under control
ADA+ under test	**Stratum of interest**	
ADA- under test		

Effect measures

Primary interest:

- Compare $Y(1)$ vs. $Y(0)$ in stratum $\{S(1) = 1\}$.
- Contrast this to results in $\{S(1) = 0\}$.

Effect measure:

- (Hazard ratio **not causally interpretable**: Aalen *et al.* (2015).)
- Base effect measure on **survival functions**:

$$U_1(t) \;:=\; P(Y(1) > t | S(1) = 1) \quad \text{and} \quad U_0(t) \;:=\; P(Y(0) > t | S(1) = 1).$$

Examples:

- **Milestone** difference at $t^* > \tilde{t}$:

$$\delta(t^*) \quad = \quad U_1(t^*) - U_0(t^*).$$

- Time-averaged version, i.e. difference in **RMST**:

$$\int_0^{t^*} \delta(t)dt \quad = \quad E[\min(Y(1), t^*) - \min(Y(0), t^*)].$$

Potential outcomes, estimands, and PS

All estimand strategies can formulated using potential outcomes:
Lipkovich *et al.* (2020).

Additional complications: Y time-to-event $\Rightarrow$ outcome event = competing risk for intercurrent event. Naive analyses conditioning on observed intercurrent event:

- Compares **non-randomized** populations.
- **Immortal bias**: patients immortal until observation of S.

Estimation of principal effects

Assumptions

Randomization not enough to estimate principal effects.

Need assumptions.

Estimation

SUTVA:

- Underpins virtually all estimation methods.
- POs for any patient do not change with treatment assigned to other patients.
- No multiple versions of treatment.

Monotonicity:

- $S(1) \geq S(0) \Rightarrow$ patients that are ADA+ on control would also be ADA+ on test.
- Patient with $S(0) = 1$ observed $\Rightarrow$ would know that $S(1) = 1 \Rightarrow$ bottom-left stratum in table empty.
- Allows estimation of principal stratum prevalences.

Exclusion-restriction:

- Assume $Y(0) = Y(1)$ (no treatment effect) for patients $\{S(0) = 0\} \cap \{S(1) = 0\}$ and $\{S(0) = 1\} \cap \{S(1) = 1\}$.
- Equivalent to say "randomization has no impact for those subjects for whom treatment has no effect on S", Joffe *et al.* (2007).

Estimation

Joint models, Frangakis and Rubin (2002):

- Model for outcome given PS membership: $Y(0), Y(1)|S(1), S(0)$.
- Model for PS membership $S(0), S(1)$.
- Multiply likelihoods $\Rightarrow$ joint model for Y and S.
- Treat unobserved potential outcomes as missing data $\Rightarrow$ integrate out to define likelihood.
- Can easily include covariates in either model.
- Use (weakly informative) priors to govern "strength" of assumption, e.g. monotonicity.
- Application: Magnusson *et al.* (2019), Public Assessment Report of the European Medicines Agency (EPAR):
  European Medicines Agency, Committee for Medicinal Products for Human Use (2019).

Estimation approaches: principal ignorability

Principal ignorability (PI, or conditional independence):

- Approach very similar to propensity scoring in observational studies.
- Specify **separate models** Y and S.
- Conditional on baseline covariates X: $Y(0)$ and $S(1)$ independent.
- X: all variables that **confound** $Y(0)$ and $S(1)$ $\Rightarrow$ once X are known, $S(1)$ provides no further information on $Y(0)$ (+ vice versa):

$$p(Y(0)|X, S(1)) \quad = \quad p(Y(0)|X)$$

- Allows modeling of $Y(0)$ and $S(1)$ **just based on** X. Unobserved outcome not needed in model.
- Assumption is **across worlds**.

Estimation approaches: principal ignorability

Estimand of interest:

$$P(Y(1) > t|S(1) = 1) - P(Y(0) > t|S(1) = 1).$$

Estimation:

- $P(Y(1) > t|S(1) = 1)$: survival function in ADA+ in treatment arm.
- $P(Y(0) > t|S(1) = 1)$: tricky, because $Y(0)$ and $S(1)$ **never jointly observed**.
- PI allows estimation of second quantity **just based on** X.

Randomization is key:

- Ensures that relationship $X - S$ same in both groups.
- Allows prediction of PS membership in control group using model from treatment group.

Estimation under principal ignorability for ADA example

- Estimate $P(S(1) = 1|X)$ on treatment arm using logistic regression.
- Use predicted probabilities as **weights** for patients in control arm $\Rightarrow$ make samples **comparable**.
- Compute effect measure of interest.
- Alternatives:
  - **Multiple imputation**, i.e. impute $S(1)$ for control patients. Properly accounts for uncertainty in estimated weights!
  - Plain **regression adjustment**.
  - **Matching**.
- See propensity score literature for assessment of methods, e.g. Austin (2011).

Choice of X:

- Adjust for all confounders that make $Y(1)$ and $S(0)$ (+ vice versa) independent.
- Only adjust for X that confound Y and S across worlds.
- **Do not include** covariates that "only" help predict S but have no impact on Y.
- Similar to considerations for observational studies.

Sensitivity analyses!

Assumptions **unverifiable**:

- "Across-world" $\Rightarrow$ even with **infinite number of observations** we could not test them.
- Only verifiable if we could observe both, patient receives control in one world and treatment in other.

scientific knowledge + sensitivity analyses

Conclusions principal stratification

Conclusions:

- Many relevant examples in drug development.
- Scientific question typically not primary, but important to characterize treatment effect in subgroups built by intercurrent events, such as ADA.
- Naive analyses often standard: Unclear estimand $\Rightarrow$ **causal conclusion unclear**.
- Complex question $\Rightarrow$ complex analysis needed.
- Assumptions needed: scientific input + sensitivity analyses.

Markdown:

https://oncoestimand.github.io/princ_strat_drug_dev/princ_strat_example.html

BBS seminar:

http://bbs.ceb-institute.org/?p=1587

Effective statistician podcast, together with Björn Bornkamp:

https://theeffectivestatistician.com/

a-deep-dive-into-principal-stratification-and-causal-inference

Agenda

Impact on data collection and trial planning

- Estimand **dictates data that need to be collected**.
- Each trial likely to have **multiple estimands** $\Rightarrow$ different estimands might require different data!
- Requires **multi-disciplinary** involvement from **earliest stages** of clinical trial development.
- Impacts **design of eCRF** or other data collection tools and monitoring strategy.
- Likely increased effort in recording reasons underlying **treatment or study withdrawals, or missing data**.
- Might need to reflect estimand assumptions in **sample size computation**!

Novo Nordisk:

- Focussing on retention, keeping subjects in trial even after discontinuing trial drug.
- Increased completion rates from **90% to 98%** in type 1 diabetes and from **70% to over 90%** in obesity trials.
- Source: `https://www.dsbs.dk/moder/Estimands/HLynggaard.pdf`.

Broader impact

Aligning stakeholder's expectations for target treatment effect **upfront** has potential to give:

- Increased **transparency** and **clarity** with respect to assumptions, data analysis, and inference.
- Clarity about **added value** of drugs: **meaningful** descriptions of treatment effects for licensing and prescribing decisions.
- Clinical trials with designs that are **aligned to agreed objectives**.
- Clear language to describe and discuss different estimands required by different stakeholders.
- More **predictable** regulatory assessment procedures.
- **Reduction in total number of analyses** (primary + secondary + sensitivity).
- **Shift of resources** from analysis / filing to design.
- Alternative approaches to avoid non-informative treatment policy estimand if its assumption very likely to be violated.

Design trumps analysis.

Don Rubin, American Statistician

Rubin (2008)

Agenda

1. ICH E9(R1) addendum: Why? And what's new?
2. Case study: hematology
3. Case study: CAR-T
4. Hypothetical strategy to address ICEs: application to Covid-19
5. Case study: treatment switching
6. Subgroups by post-randomization event - principal stratification
7. Impact and conclusions
8. Industry working group *Estimands in oncology*

Industry working group on estimands in oncology:

- Founded February 2018.
- Represents industry in Europe and US:
  - European special interest group "Estimands in oncology", sponsored by PSI and EFSPI.
  - ASA scientific working group of ASA biopharmaceutical section.
- **54** members (20 EU + 29 US + 5 Asia) representing **28** companies.
- Regularly interacts with **8 health authorities**.
- Presentations, webinars, papers.

www.oncoestimand.org


NOVARTIS
Roche
Pfizer
MERCK
janssen
PHARMACEUTICAL COMPANIES OF Johnson&Johnson
SANOFI
gsk
abbvie
AstraZeneca
BeiGene
BAYER
Takeda
Celgene
astellas
SERVIER
MERCK
PPD
NEKTAR
Merus battling cancer
stabu ro
Statistical Consulting
AMGEN
ICON
A Symbol of Excellence
Boehringer Ingelheim
Adelphi
ADELPHI VALUES
Daiichi-Sankyo
MERCK SERONO
zai Lab

Papers

Published or accepted:

- Lawrance *et al.* (2020): What is an estimand & how does it relate to quantifying the effect of treatment on patient-reported quality of life outcomes in clinical trials. **link**
- Degtyarev *et al.* (2020): Assessing the impact of COVID-19 on the objective and analysis of oncology clinical trials - application of the estimand framework. **link**
- Casey *et al.* (2020): Estimand framework: Are we asking the right question? A case study in the solid tumor setting. **link**

Revision submitted:

- Sun *et al.* (2020): Estimands in Hematology Trials. **link**
- Manitz *et al.* (2020): Estimands in clinical trials with treatment switching.
- Bornkamp *et al.* (2020): Principal Stratum Strategy: Potential Role in Drug Development. **link** (incl. markdown file with code).

More papers under preparation.

Upcoming task forces

- Clinical engagement.
- Principal stratification and treatment switching.
- Time to response and DOR.
- Estimands and PRO.
- Follow-up quantification.
- RWD.
- Conditional vs. marginal.
- Time to event endpoints with prognostic or predictive biomarker subgroups.

If you do not know how to ask the right question, you discover nothing.

W.E. Deming, American Statistician

Thank you for your attention.

kaspar.rufibach@roche.com

http://www.kasparrufibach.ch

numbersman77

numbersman77

References I

- Aalen, O. O., Cook, R. J., and Røysland, K. (2015). Does Cox analysis of a randomized survival study yield a causal treatment effect? *Lifetime Data Anal*, **21**(4), 579–593.
- Austin, P. C. (2011). An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. *Multivariate Behav Res*, **46**(3), 399–424.
- Bornkamp, B., Rufibach, K., Lin, J., Liu, Y., Mehrotra, D. V., Roychoudhury, S., Schmidli, H., Shentu, Y., and Wolbers, M. (2020). Principal stratum strategy: Potential role in drug development.
- Casey, M., Degtyarev, E., Lechuga, M., Aimone, P., Ravaud, A., Motzer, R., Liu, F., Stalbovskaya, V., Tang, R., Butler, E., Sailer, O., Halabi, S., and George, D. (2020). Estimand framework: Are we asking the right question? a case study in the solid tumor setting. *Pharm. Stat., to appear.*
- Degtyarev, E., Rufibach, K., Shentu, Y., Yung, G., Casey, M., Englert, S., Liu, F., Liu, Y., Sailer, O., Siegel, J., Sun, S., Tang, R., Zhou, J., and on behalf of the Industry Working Group on Estimands in Oncology (2020). Assessing the impact of covid-19 on the clinical trial objective and analysis of oncology clinical trials - application of the estimand framework. *Statistics in Biopharmaceutical Research*, **12**(4), 427–437.
- European Medicines Agency, Committee for Medicinal Products for Human Use (2019). Mayzent: Assessment report.
- Frangakis, C. E. and Rubin, D. B. (2002). Principal stratification in causal inference. *Biometrics*, **58**(1), 21–29.
- Hemmings, R. (2015). The 'estimand' - problem statement. Presented at PSI 'Estimands Discussion Meeting' in February 2015.
- ICH E9 working group (2019). ICH E9 (R1): addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials.
- Joffe, M. M., Small, D., Hsu, C.-Y., *et al.* (2007). Defining and estimating intervention effects for groups that will develop an auxiliary outcome. *Statistical Science*, **22**(1), 74–97.
- Lawrance, R., Degtyarev, E., Griffiths, P., Trask, P., Lau, H., D'Alessio, D., Griebsch, I., Wallenstein, G., Cocks, K., and Rufibach, K. (2020). What is an estimand & how does it relate to quantifying the effect of treatment on patient-reported quality of life outcomes in clinical trials? *J Patient Rep Outcomes*, **4**, 68.

References II

- Lipkovich, I., Ratitch, B., and Mallinckrodt, C. H. (2020). Causal inference and estimands in clinical trials. *Statistics in Biopharmaceutical Research*, **1**(12), 54–67.
- Magnusson, B. P., Schmidli, H., Rouyrre, N., and Scharfstein, D. O. (2019). Bayesian inference for a principal stratum estimand to assess the treatment effect in a subgroup characterized by postrandomization event occurrence. *Statistics in Medicine*, **38**(23), 4761–4771.
- Manitz, J., Kan-Dobrosky, N., Buchner, H., Casadebaig, M., Degtyarev, E., Dey, J., Haddad, V., Fei, J., Martin, E., Mo, M., Rufibach, K., Shentu, Y., Stalbovskaya, V., Tang, R., Yung, G., and Zhu, J. (2020). Estimands in clinical trials with treatment switching. *Under revision*.
- Rubin, D. B. (2008). For objective causal inference, design trumps analysis. *The Annals of Applied Statistics*, **2**(3), 808–840.
- Sheiner, L. B. (1991). The intellectual health of clinical drug evaluation. *Clin Pharmacol Ther*, **50**(1), 4–9.
- Stone, R. M., Mandrekar, S. J., Sanford, B. L., Laumann, K., Geyer, S., Bloomfield, C. D., Thiede, C., Prior, T. W., Döhner, K., Marcucci, G., Lo-Coco, F., Klisovic, R. B., Wei, A., Sierra, J., Sanz, M. A., Brandwein, J. M., de Witte, T., Niederwieser, D., Appelbaum, F. R., Medeiros, B. C., Tallman, M. S., Krauter, J., Schlenk, R. F., Ganser, A., Serve, H., Ehninger, G., Amadori, S., Larson, R. A., and Döhner, H. (2017). Midostaurin plus chemotherapy for acute myeloid leukemia with a flt3 mutation. *The New England journal of medicine*, **377**, 454–464.
- Sun, S., Weber, J., Butler, E., Rufibach, K., and Roychoudhury, S. (2020). Estimands in hematology trials. *Under revision*.