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

Kaspar Rufibach
Methods, Collaboration, and Outreach Group, F. Hoffmann-La Roche, Basel
Royal Statistical Society Session on Design for Medical and Clinical studies
16 December 2021

Acknowledgments

This material:

- was first presented at the 76th Deming Conference on Applied Statistics on 9th December 2020 link,
- by Kaspar Rufibach and **Evgeny Degtyarev (Novartis)**.
- Sections on CAR-T and switching had initially been prepared by Evgeny.

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

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)

After emailing scientific question back three times:

After emailing scientific question
back three times:

So hard exercise, it made me realise I am not sure what exactly we want.

After emailing scientific question
back three times:

So hard exercise, it made me realise I am not sure what exactly we want.

Roche quantitative scientist

Agenda

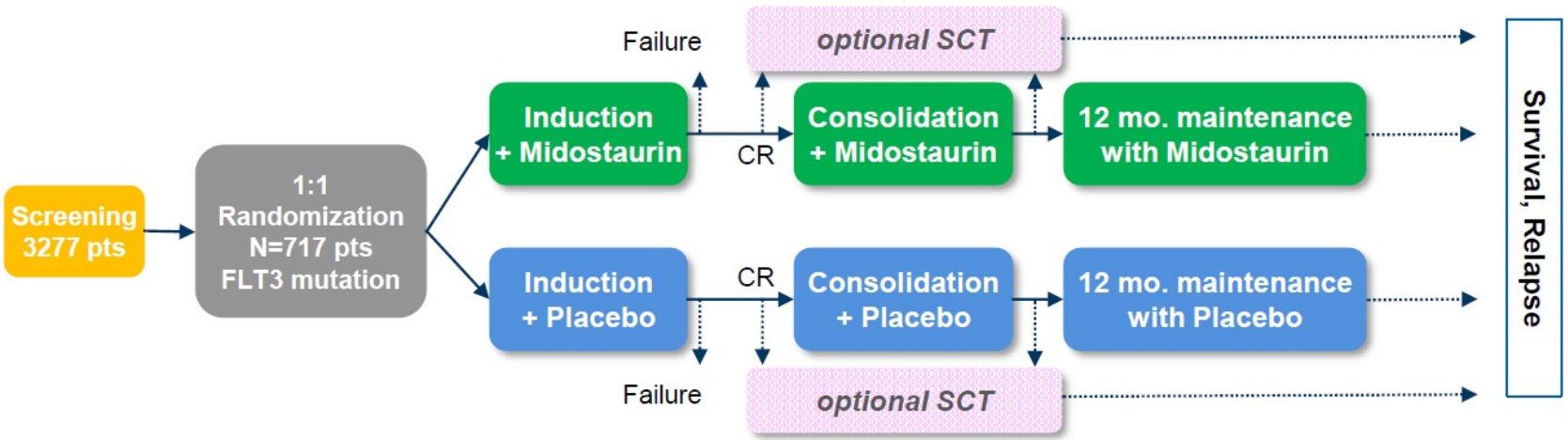
1. Case study: hematology
2. Case study: treatment switching
3. Impact and conclusions
4. Backup: ICH E9(R1) addendum: Why? And what's new?
5. Backup: Industry working group *Estimands in oncology*
6. Backup: Subgroups by post-randomization event - principal stratification
7. Backup: Estimation of average causal effect
8. Backup: Estimation of principal effects

Agenda

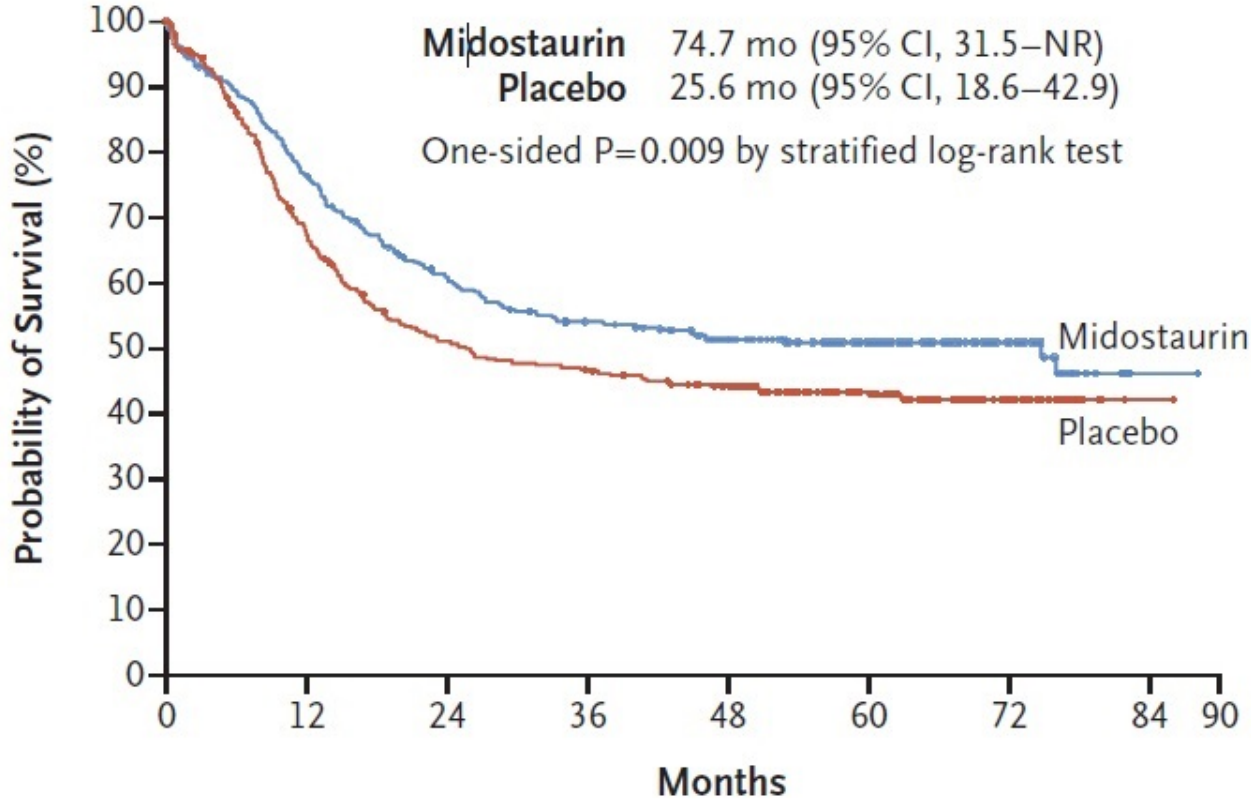
Case study: hematology

Complex treatment strategies in hematology

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



- **Randomized, phase III** 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?

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 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.

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**!

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 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?

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.

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:

AML: treatment strategy based on sequence of

AML: treatment strategy based on sequence of

- multiple decision points and

AML: treatment strategy based on sequence of

- multiple decision points and
- treatment modalities.

AML: treatment strategy based on sequence of

- multiple decision points and
- treatment modalities.

RATIFY:

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.

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:

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.

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.

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?

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.

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

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

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

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

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

Complex (multiphase) strategies:

Non-proportional hazards?

Complex (multiphase) strategies:

Non-proportional hazards?

Cure?

What do these findings have in common?

What do these findings have in common?

They can all be anticipated!

What do these findings have in common?

They can all be anticipated!

Clear formulation of
clinical trial objective is key.

DOI: 10.1002/pst.2108

MAIN PAPER

WIL

Estimands in hematologic oncology trials

Steven Sun[1] | Hans-Jochen Weber[2] | Emily Butler[3] | Kaspar Rufibach[4] | Satrajit Roychoudhury[5]

Sun *et al.* (2021):

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

Agenda

Case study: treatment switching

Good old days: Herceptin

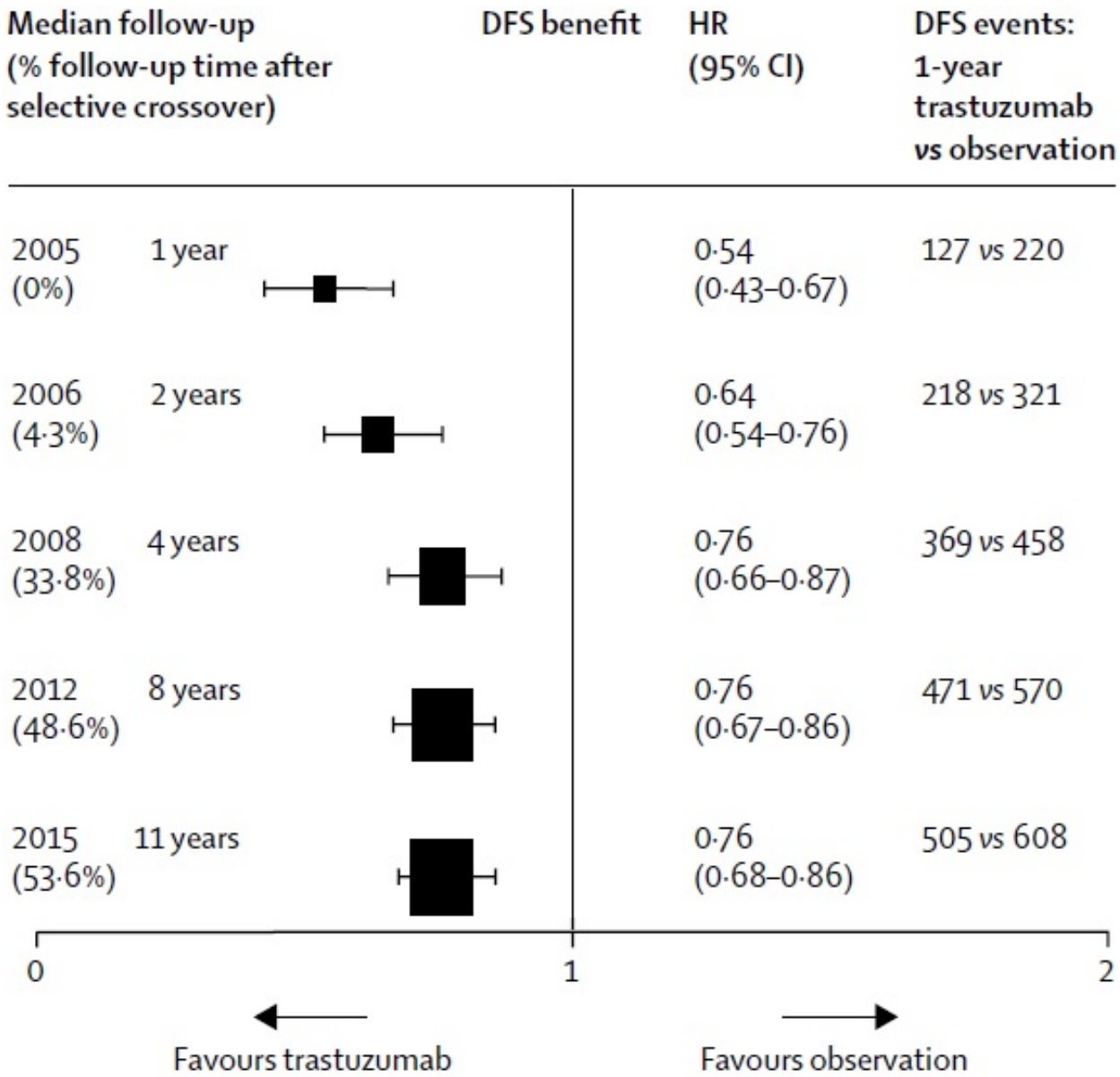
HERA

- **Population**: HER2+ early breast cancer patients.
- **Primary therapy**: surgery, chemotherapy, or radiotherapy as indicated.
- **Comparison**: after completion of primary therapy: trastuzumab vs. observation.
- **Randomized, phase III** clinical trial.
- Primary endpoint: investigator-assessed **disease-free survival**.

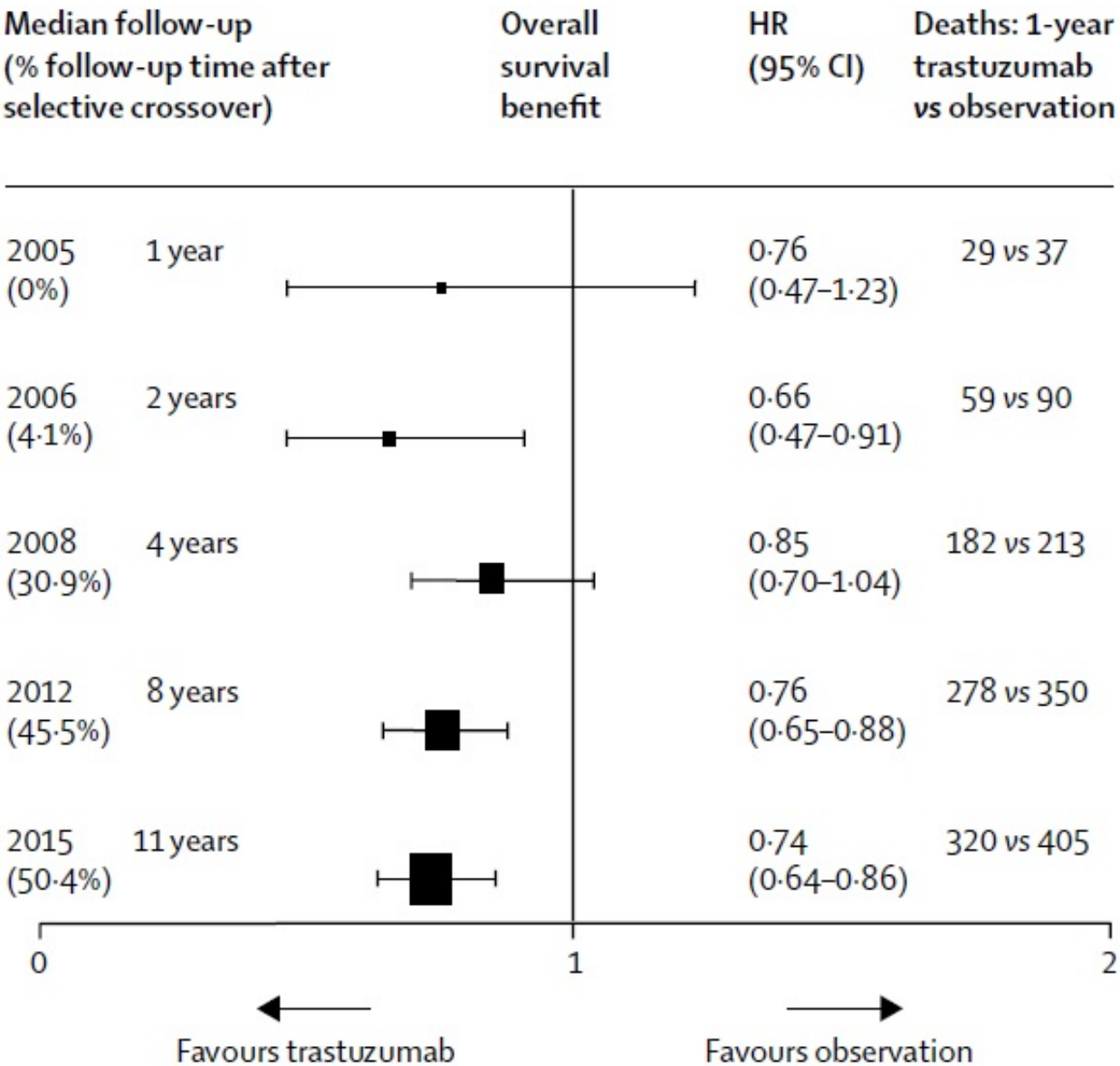
Piccart-Gebhart and Procter (2005):

- Trial stopped **early** at planned interim analysis (347 events).
- All control patients without prior disease recurrence allowed to **cross-over to trastuzumab** $\Rightarrow$ 52% did so.

Primary endpoint DFS in HERA over time



Overall survival in HERA over time



HERA: comments

OS effect establised in long-term follow-up **despite cross-over**:

HERA: comments

OS effect establised in long-term follow-up **despite cross-over**:

- Herceptin new drug class $\Rightarrow$ large treatment effect.

HERA: comments

OS effect establised in long-term follow-up **despite cross-over**:

- Herceptin new drug class $\Rightarrow$ large treatment effect.
- No alternative therapy for control arm patients $\Rightarrow$ crossover represents standard of care.

HERA: comments

OS effect establised in long-term follow-up **despite cross-over**:

- Herceptin new drug class $\Rightarrow$ large treatment effect.
- No alternative therapy for control arm patients $\Rightarrow$ crossover represents standard of care.
- **Globally**!

HERA: comments

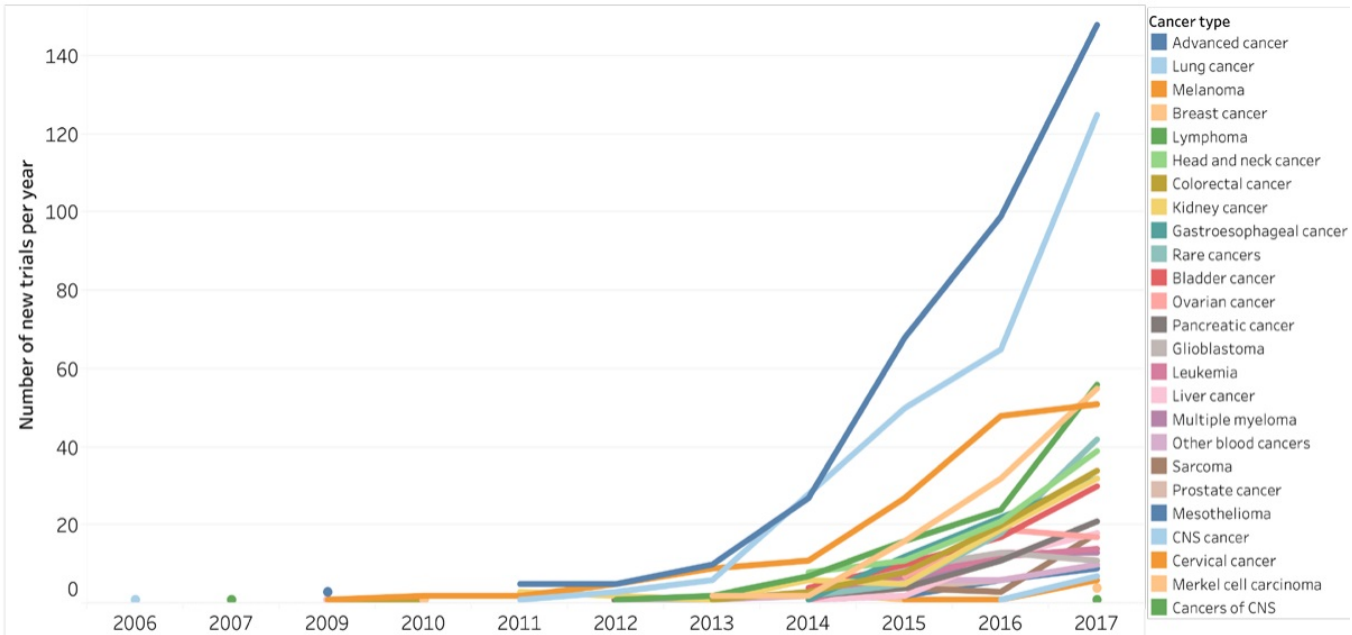
OS effect establised in long-term follow-up **despite cross-over**:

- Herceptin new drug class $\Rightarrow$ large treatment effect.
- No alternative therapy for control arm patients $\Rightarrow$ crossover represents standard of care.
- **Globally**!

Treatment policy estimand interpretable.

Oncology landscape has changed!

Clinical trials with anti-PD1/PDL1 agents

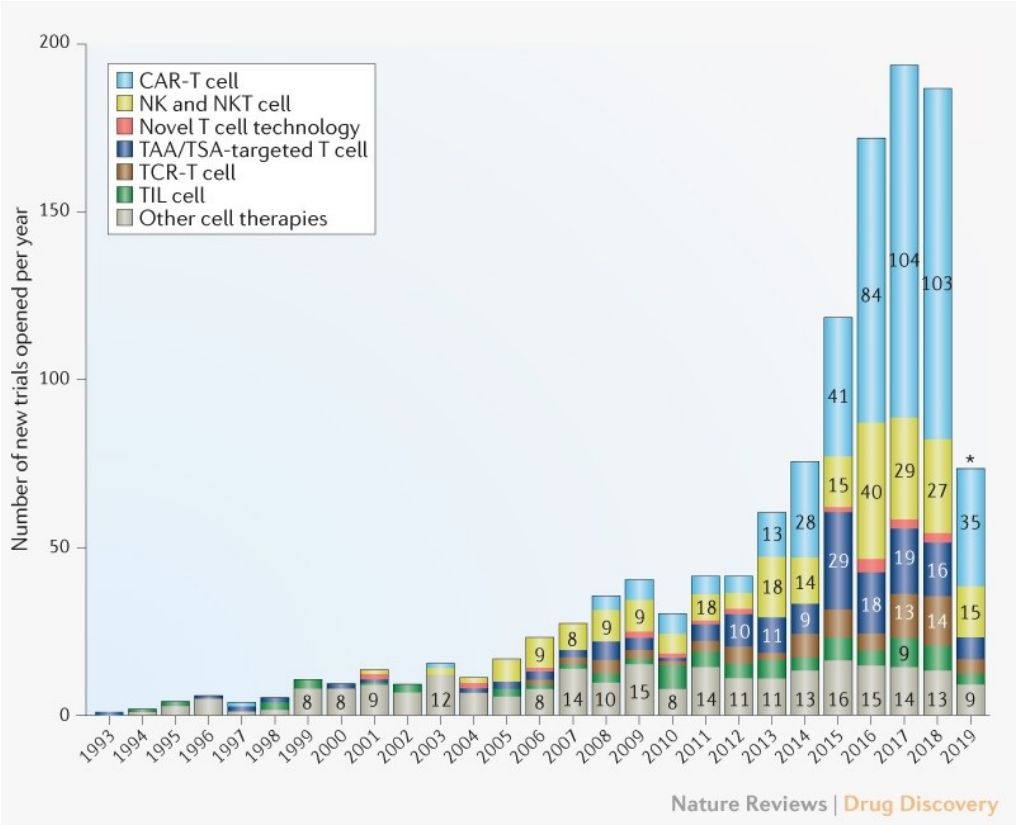


1 in 2006, **1502** in Sep 2017, **2250** in Sep 2018, **2975** in Sep 2019.

Tang *et al.* (2018)

`https://www.cancerresearch.org/scientists/immuno-oncology-landscape/pd-1-pd-l1-landscape`.

CAR-T trials



13 in 2013, $>$**100** in 2017.

Yu *et al.* (2018).

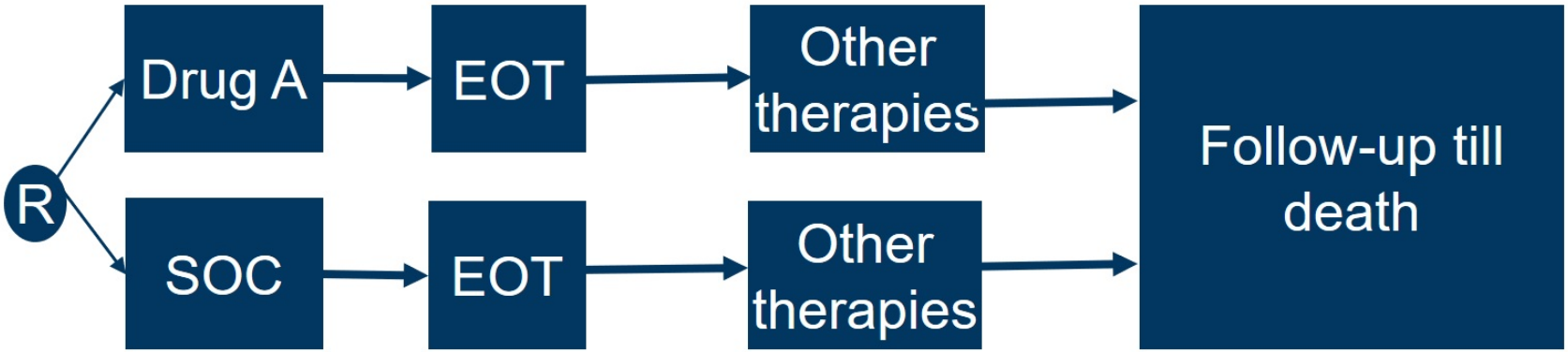
Great for patients!

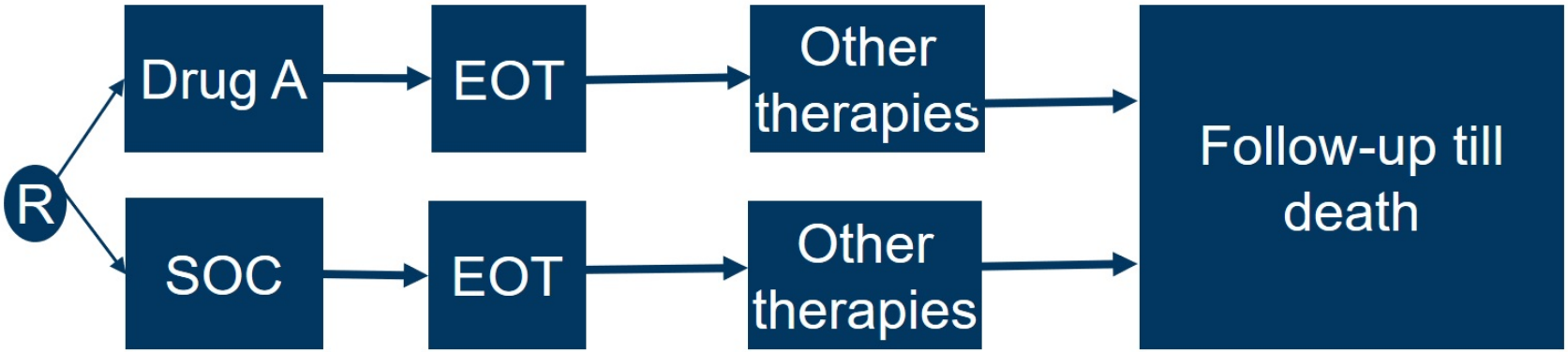
- durable responses,
- many ongoing clinical trials.

Great for patients!

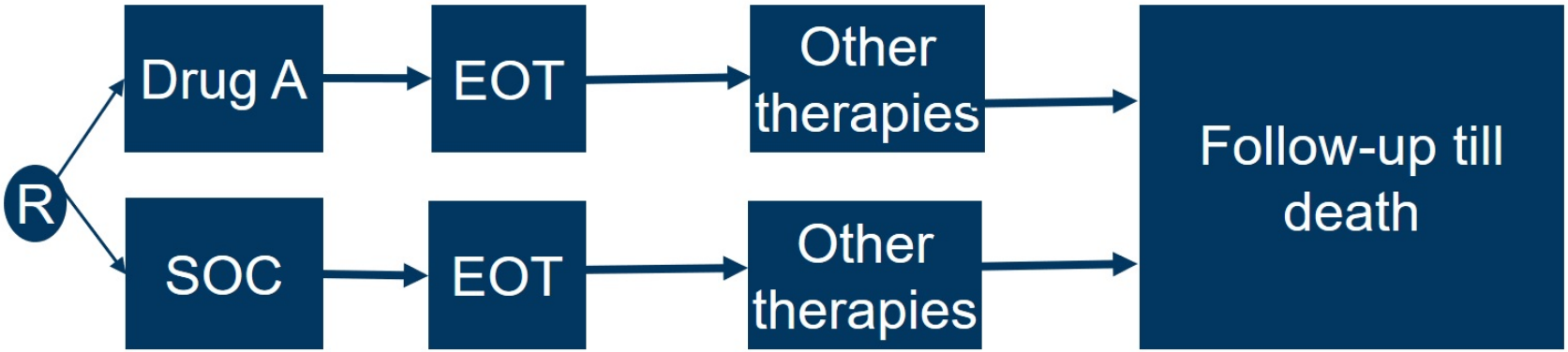
- durable responses,
- many ongoing clinical trials.

But what does it mean for clinical trials?


R
Drug A
EOT
Other therapies
SOC
EOT
Other therapies
Follow-up till death

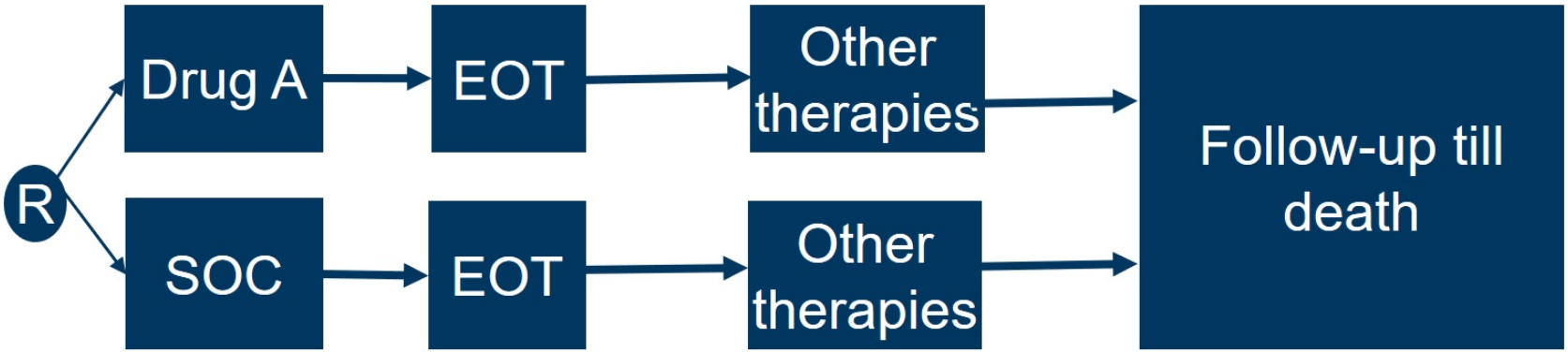


Typical OS definition:



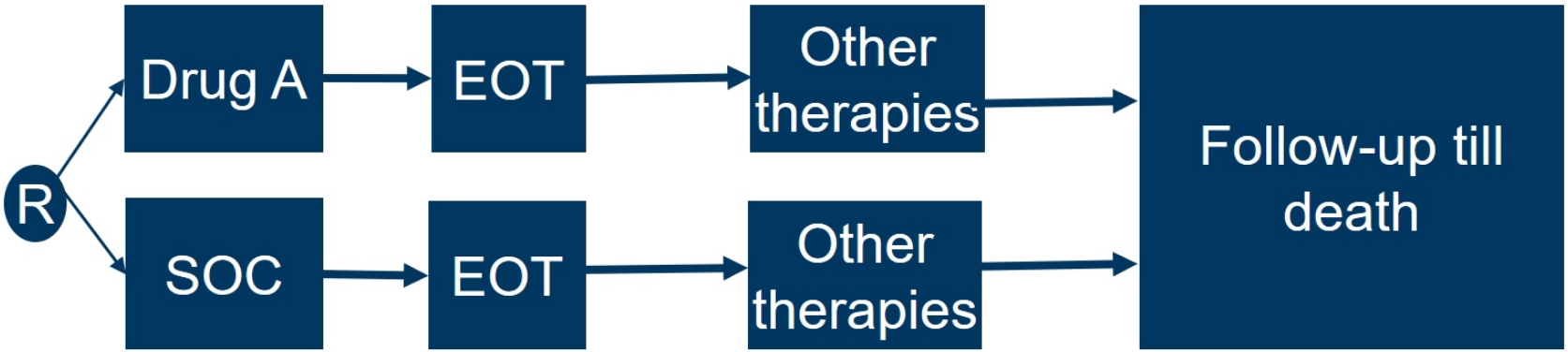
Typical OS definition:

- Time from randomization to death **regardless of patient's journey**.



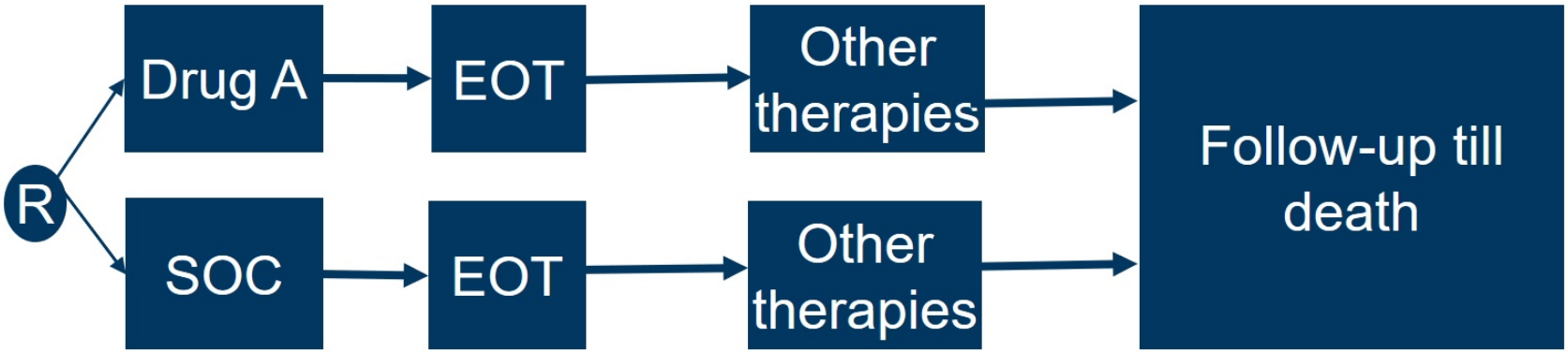
Typical OS definition:

- Time from randomization to death **regardless of patient's journey**.
- **Treatment policy** for every intercurrent event (crossover, new therapy, etc.).



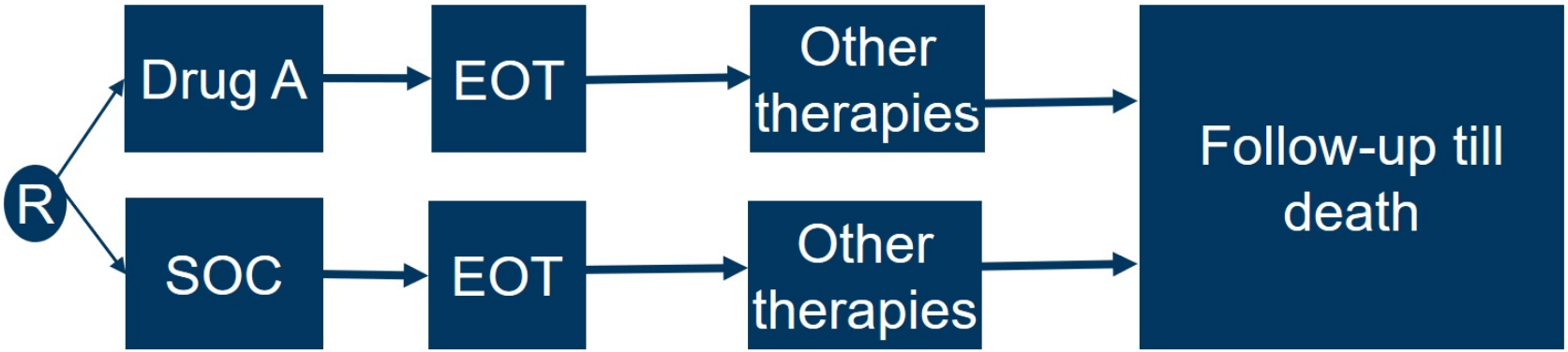
Typical OS definition:

- Time from randomization to death **regardless of patient's journey**.
- **Treatment policy** for every intercurrent event (crossover, new therapy, etc.).
- Balance in subsequent therapies generally not expected:



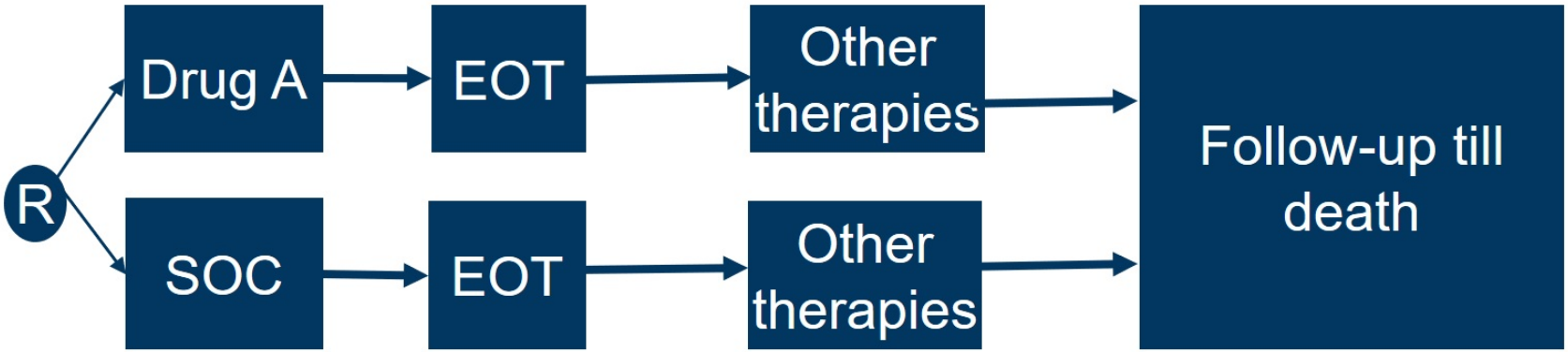
Typical OS definition:

- Time from randomization to death **regardless of patient's journey**.
- **Treatment policy** for every intercurrent event (crossover, new therapy, etc.).
- Balance in subsequent therapies generally not expected:
  - Physician choose subsequent therapy in light of previously administered therapies.



Typical OS definition:

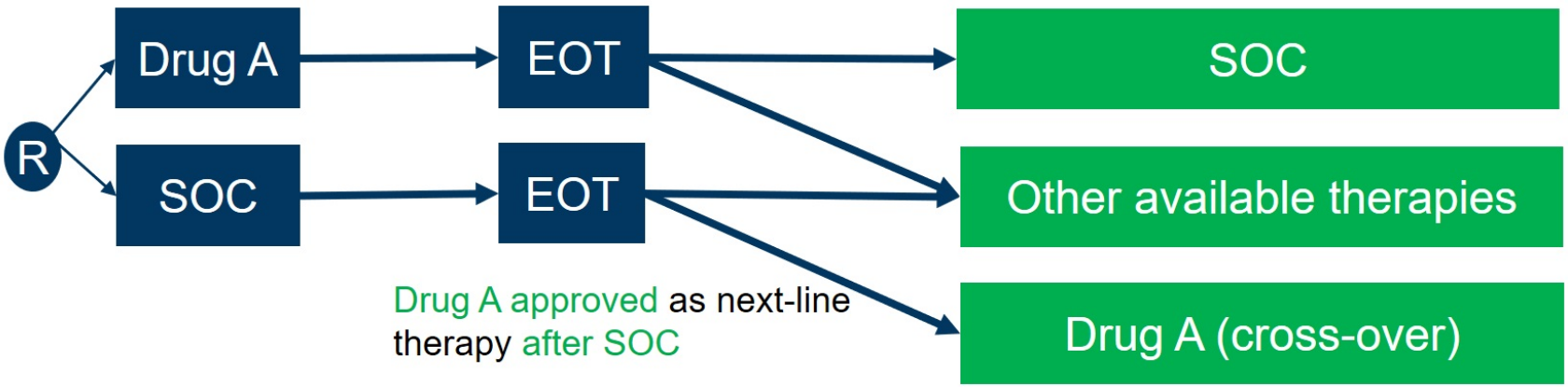
- Time from randomization to death **regardless of patient's journey**.
- **Treatment policy** for every intercurrent event (crossover, new therapy, etc.).
- Balance in subsequent therapies generally not expected:
  - Physician choose subsequent therapy in light of previously administered therapies.
  - If experimental drug works $\Rightarrow$ less switchers.

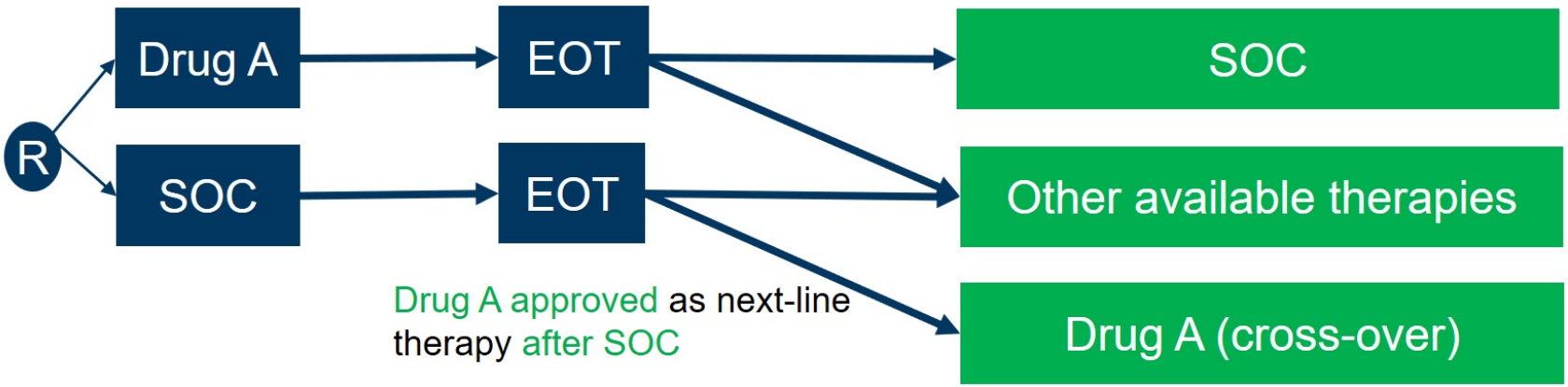


Typical OS definition:

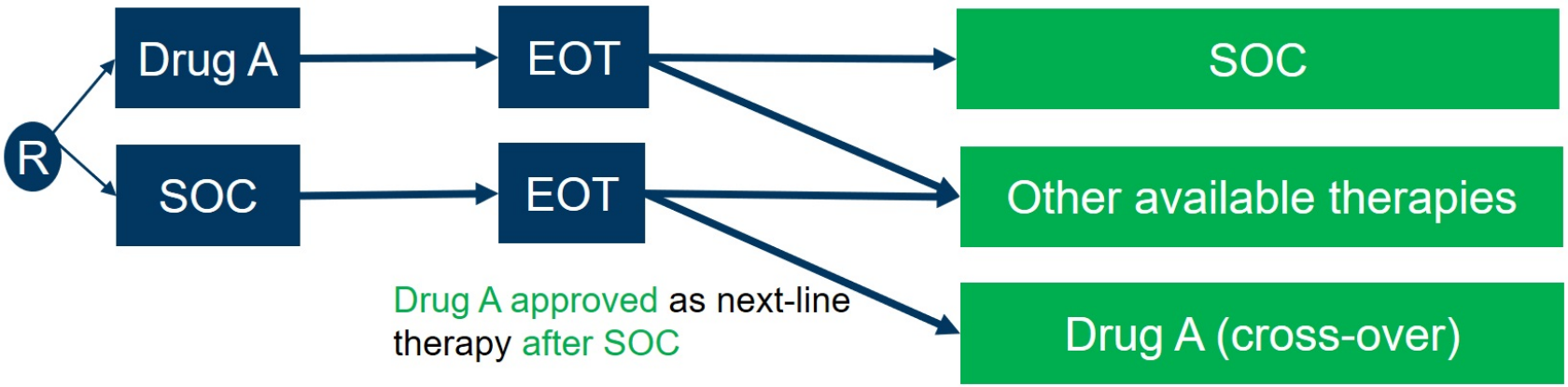
- Time from randomization to death **regardless of patient's journey**.
- **Treatment policy** for every intercurrent event (crossover, new therapy, etc.).
- Balance in subsequent therapies generally not expected:
  - Physician choose subsequent therapy in light of previously administered therapies.
  - If experimental drug works $\Rightarrow$ less switchers.

Treatment policy OS estimand **interpretable** if subsequent therapy after EOT reflects **clinical practice**.


Drug A
EOT
SOC
R
SOC
EOT
Other available therapies
Drug A approved as next-line therapy after SOC
Drug A (cross-over)

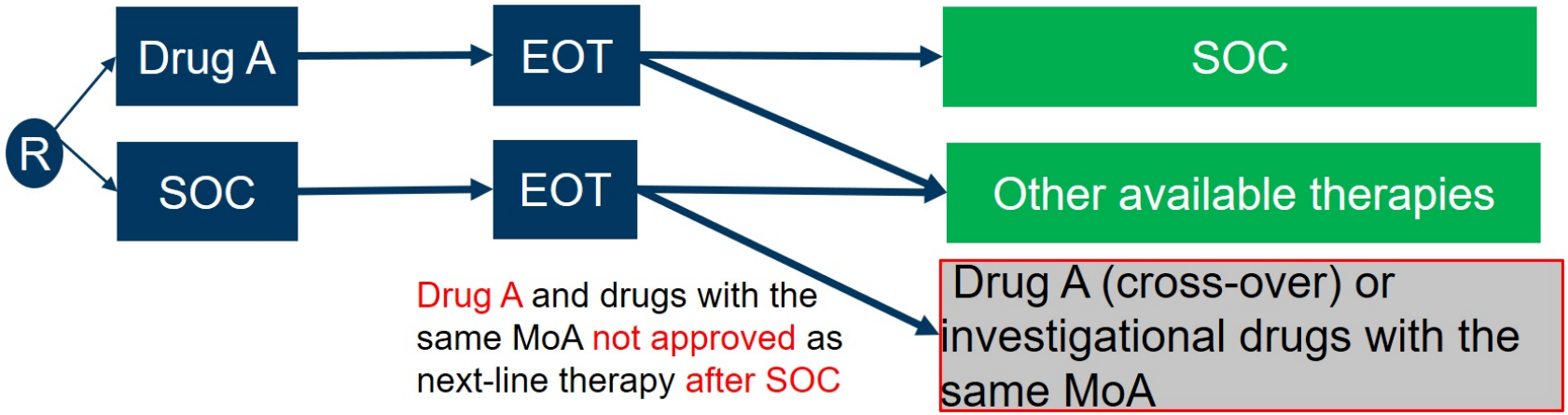


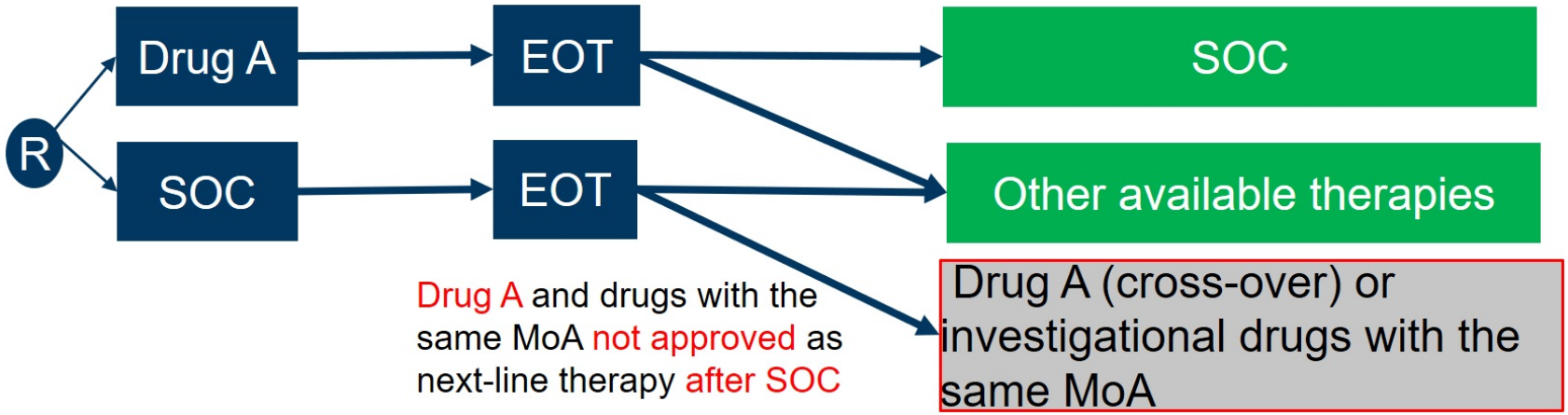
Subsequent therapy after EOT reflects clinical practice.



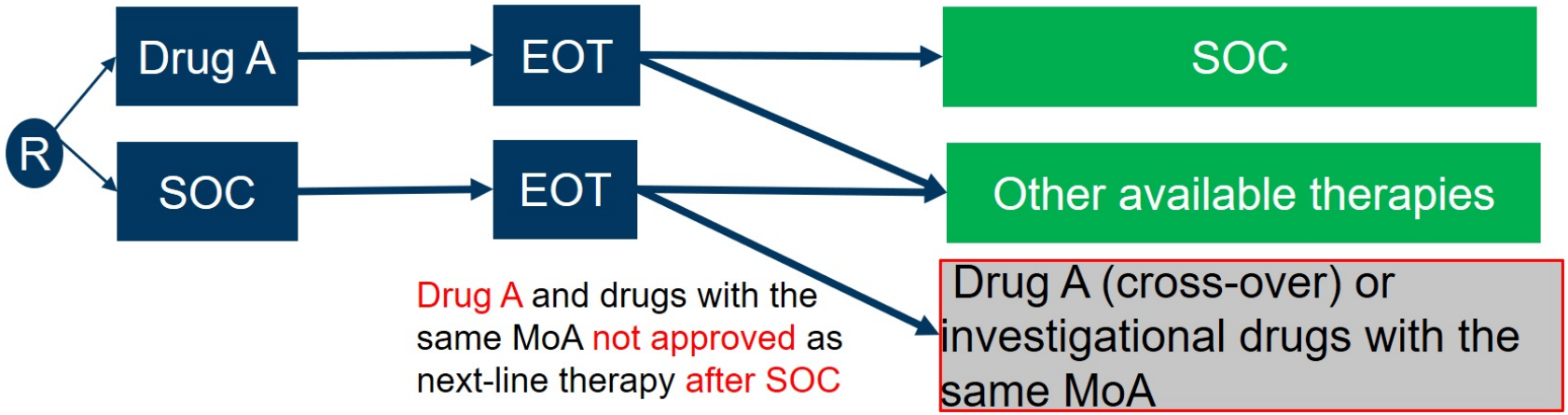
Subsequent therapy after EOT reflects clinical practice.

Treatment policy OS estimand **interpretable**.


R
Drug A
SOC
EOT
EOT
SOC
Other available therapies
Drug A (cross-over) or investigational drugs with the same MoA
Drug A and drugs with the same MoA not approved as next-line therapy after SOC

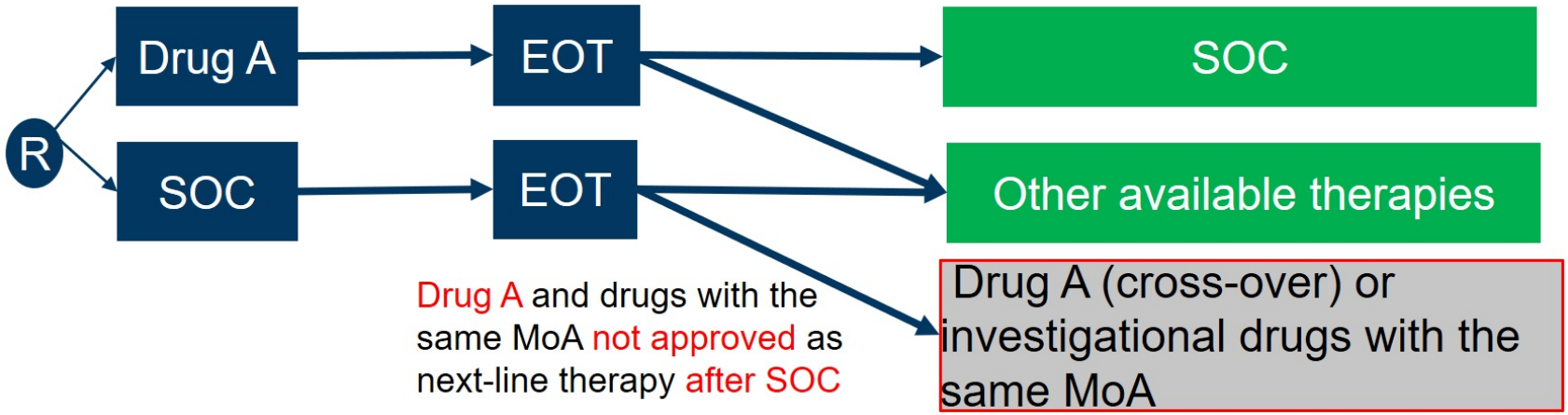


Subsequent therapy after EOT **does not** reflect clinical practice:



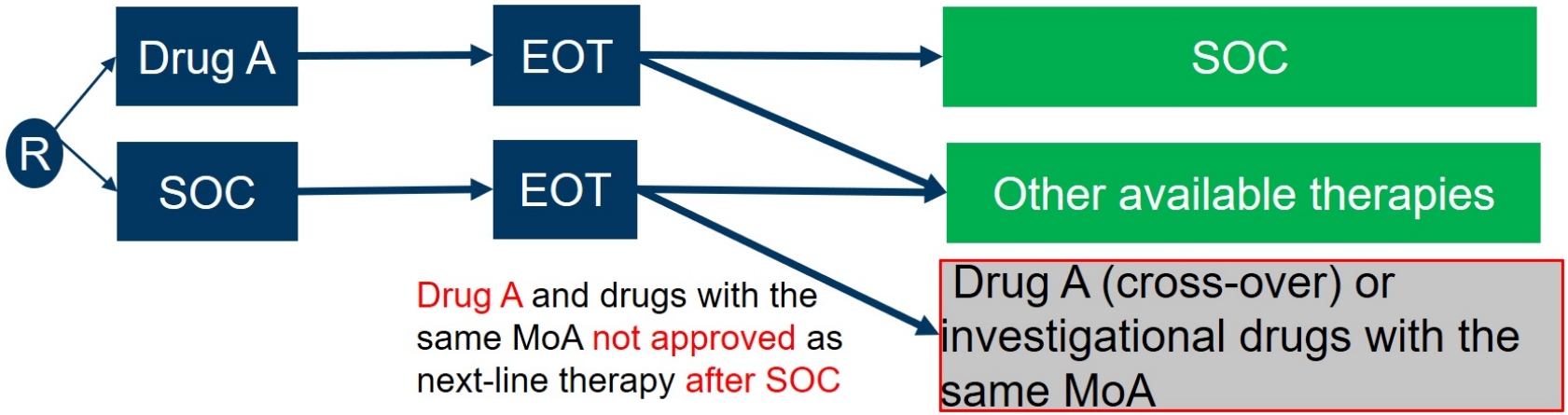
Subsequent therapy after EOT **does not** reflect clinical practice:

- Immuno-oncology.



Subsequent therapy after EOT **does not** reflect clinical practice:

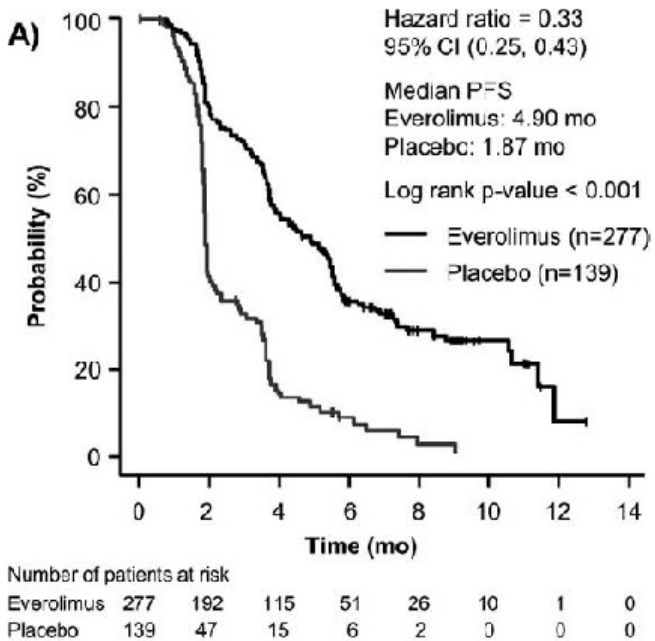
- Immuno-oncology.
- Treatment policy estimand relevant?

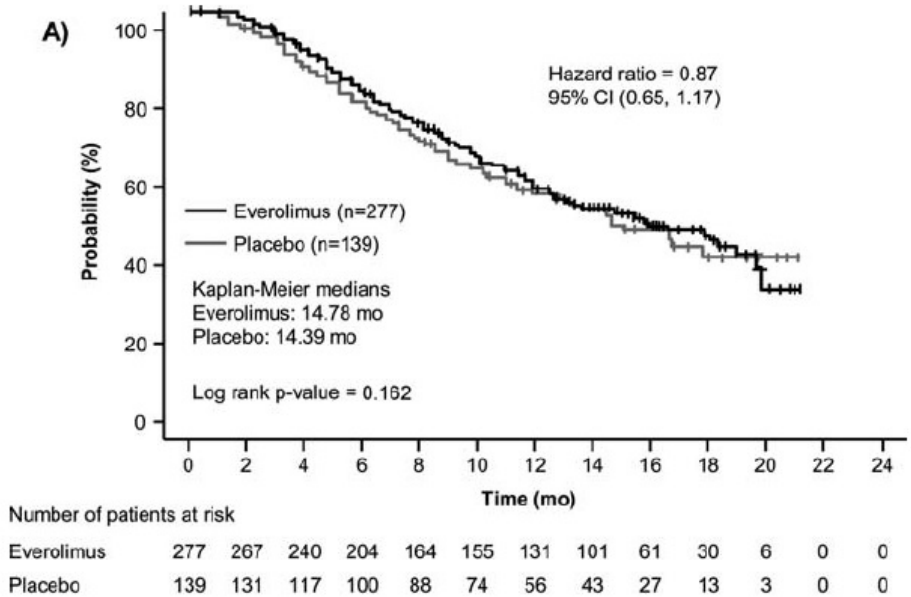


Subsequent therapy after EOT **does not** reflect clinical practice:

- Immuno-oncology.
- Treatment policy estimand relevant?
- Benefit on OS without cross-over more informative? **Hypothetical estimand**!

RECORD-1





RECORD-1: Motzer *et al.* (2010). PFS (left) and OS (right).

Further examples: GRID, Demetri *et al.* (2016); GLARIUS, Herrlinger *et al.* (2016), Javelin Lung 200, Barlesi *et al.* (2019).

Randomized but not treated

- **Blinding** often infeasible.
- Checkmate-37:
  - **20% vs 1.5%**.
  - Weber *et al.* (2015).
- Quantum-R:
  - **23% vs 1.6%**.
  - Cortes *et al.* (2019).

Randomized but not treated

- **Blinding** often infeasible.
- Checkmate-37:
  - **20% vs 1.5%**.
  - Weber *et al.* (2015).
- Quantum-R:
  - **23% vs 1.6%**.
  - Cortes *et al.* (2019).

That is quite **bothersome**, *I've been here 20 years. I haven't seen this discrepancy of randomized but not treated to this extent.*
(Rick Pazdur on Quantum-R)

Randomized but not treated

- **Blinding** often infeasible.
- Checkmate-37:
  - **20% vs 1.5%**.
  - Weber *et al.* (2015).
- Quantum-R:
  - **23% vs 1.6%**.
  - Cortes *et al.* (2019).

> *That is quite* **bothersome**, *I've been here 20 years. I haven't seen this discrepancy of randomized but not treated to this extent.*
> *(Rick Pazdur on Quantum-R)*

Overall survival in all randomized patients interpretable?

If subsequent therapies do not reflect clinical practice...

...OS description in labels is ambiguous:

If subsequent therapies do not reflect clinical practice...

...OS description in labels is ambiguous:

Regorafenib USPI:

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.

If subsequent therapies do not reflect clinical practice...

...OS description in labels is ambiguous:

Regorafenib USPI:

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 SmPC:

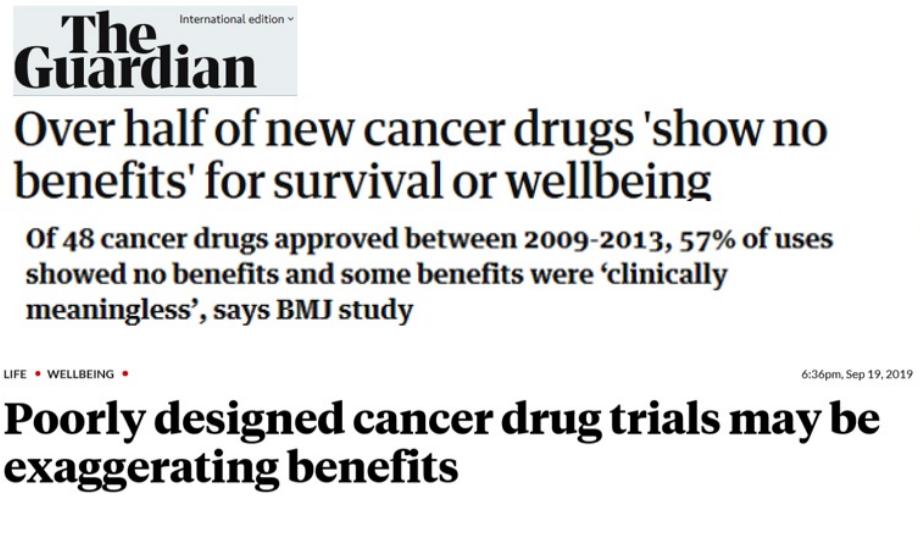
There was no statistically significant difference between nivolumab and chemotherapy in the final OS analysis. The primary OS analysis was not adjusted to account for subsequent therapies, with 54 (40.6%) patients in the chemotherapy arm subsequently receiving an anti-PD1 treatment. OS may be confounded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

If subsequent therapies do not reflect clinical practice...

...drugs are perceived as not improving survival.

If subsequent therapies do not reflect clinical practice...

...drugs are perceived as not improving survival.



The Guardian International edition

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

LIFE • WELLBEING • 6:36pm, Sep 19, 2019

Poorly designed cancer drug trials may be exaggerating benefits

HEALTH NEWS OCTOBER 13, 2017 / 8:44 PM / 7 MONTHS AGO REUTERS

Little evidence new cancer drugs improve survival

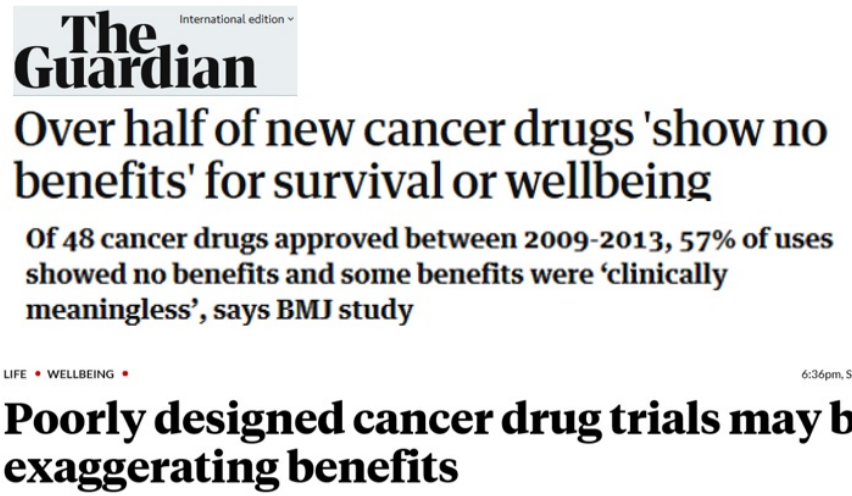
PHARMALOT STAT+

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

By ED SILVERMAN @Pharmalot / SEPTEMBER 18, 2019

If subsequent therapies do not reflect clinical practice...

...drugs are perceived as not improving survival.



Driven by

- non-significant result
- for treatment-policy OS estimand
- when subsequent therapies do not reflect clinical practice!

If subsequent therapies do not reflect clinical practice...

...regulatory standards are perceived to be low.

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 estimand represents key question of interest.

If subsequent therapies do not reflect clinical practice...

...hypothetical estimand 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.

Relevant for patients and prescribers in label: **effect of STIVARGA on OS if placebo-treated patients did not have possibility to cross-over to STIVARGA after PD?**

If subsequent therapies do not reflect clinical practice...

...hypothetical estimand 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.

Relevant for patients and prescribers in label: **effect of STIVARGA on OS if placebo-treated patients did not have possibility to cross-over to STIVARGA after PD?**

$\Rightarrow$ hypothetical strategy for intercurrent event of cross-over.

Treatment switching in immuno-oncology

Treatment switching in immuno-oncology:

Treatment switching in immuno-oncology

Treatment switching in immuno-oncology:

- Availability of **non-approved** drugs (in other clinical trials) after SOC.

Treatment switching in immuno-oncology

Treatment switching in immuno-oncology:

- Availability of **non-approved** drugs (in other clinical trials) after SOC.
- Open-label trials: Patients switch directly after randomization.

Treatment switching in immuno-oncology

Treatment switching in immuno-oncology:

- Availability of **non-approved** drugs (in other clinical trials) after SOC.
- Open-label trials: Patients switch directly after randomization.
- Additional challenge: **Varying access** to such treatment across countries.

Treatment switching in immuno-oncology

Treatment switching in immuno-oncology:

- Availability of **non-approved** drugs (in other clinical trials) after SOC.
- Open-label trials: Patients switch directly after randomization.
- Additional challenge: **Varying access** to such treatment across countries.

Treatment policy effect for OS really what we are interested in?

How DO we estimate OS effect?

How DO we estimate OS effect?

Hypothetical estimand?

Estimands for treatment switching

OBJECTIVE		*Evaluate OS benefit assuming subsequent therapies represent clinical practice*	*Evaluate OS benefit adjusted for treatment switching*	*Evaluate OS benefit adjusted for treatment cross-over at any time*	*Evaluate OS benefit adjusted for treatment cross-over upon progression*
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 (1excluding investigational drug)	Sequence of investigational drug + any subsequent therapies vs. sequence of control + any subsequent therapy (excluding investigational drug)
Strategy for addressing intercurrent events (IEs)	IE: Start of subsequent therapy at any time (other than cross-over)	Treatment policy	Hypothetical	Treatment policy	Treatment Policy
	IE: Cross-over to investigational drug without observed progression	Treatment policy	Hypothetical	Hypothetical	Treatment Policy
	IE: Cross-over to investigational drug upon progression	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

Manitz *et al.* (2021)

DOI: 10.1002/pst.2158

MAIN PAPER

Estimands for overall survival in clinical trials with treatment switching in oncology

Juliane Manitz[1] | Natalia Kan-Dobrosky[2] | Hannes Buchner[3] | Marie-Laure Casadebaig[4] | Evgeny Degtyarev[5] | Jyotirmoy Dey[6] | Vincent Haddad[7] | Fei Jie[8] | Emily Martin[1] | Mindy Mo[9] | Kaspar Rufibach[10] | Yue Shentu[11] | Viktoriya Stalbovskaya[12] | Rui (Sammi) Tang[13] | Godwin Yung[14] | Jiangxiu Zhou[15]

Manitz *et al.* (2021)

Conclusions treatment switching

All stakeholders - industry, regulators, payors - have an interest in **interpretable** OS estimates.

Conclusions treatment switching

All stakeholders - industry, regulators, payors - have an interest in **interpretable** OS estimates.

Treatment policy estimand for OS: remains main question of interest for **regulators**, patients and physicians in vast majority of situations.

Conclusions treatment switching

All stakeholders - industry, regulators, payors - have an interest in **interpretable** OS estimates.

Treatment policy estimand for OS: remains main question of interest for **regulators**, patients and physicians in vast majority of situations.

Hypothetical estimand: may be more meaningful for intercurrent events in certain situations. May help **payers** quantify **added value of new drug**.

Conclusions treatment switching

All stakeholders - industry, regulators, payors - have an interest in **interpretable** OS estimates.

Treatment policy estimand for OS: remains main question of interest for **regulators**, patients and physicians in vast majority of situations.

Hypothetical estimand: may be more meaningful for intercurrent events in certain situations. May help **payers** quantify **added value of new drug**.

Methodology may not yet be perfect: all stakeholders need to

- learn together,
- understand primary and sensitivity analyses.

Conclusions treatment switching

All stakeholders - industry, regulators, payors - have an interest in **interpretable** OS estimates.

Treatment policy estimand for OS: remains main question of interest for **regulators**, patients and physicians in vast majority of situations.

Hypothetical estimand: may be more meaningful for intercurrent events in certain situations. May help **payers** quantify **added value of new drug**.

Methodology may not yet be perfect: all stakeholders need to

- learn together,
- understand primary and sensitivity analyses.

Enables to communicate added value of drugs better.

Agenda

Impact and conclusions

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**!

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)

Thank you for your attention.

kaspar.rufibach@roche.com
http://go.roche.com/dss-mco

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.

Angrist, J. D., Imbens, G. W., and Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, **91**(434), 444–455.

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.

Barlesi, F., Özgüroglu, M., Vansteenkiste, J., Spigel, D., Yang, J. C.-H., Bajars, M., Ruisi, M., Manitz, J., and Park, K. (2019). Assessing the impact of subsequent checkpoint inhibitor (cpi) treatment on overall survival: Post hoc analyses from the phase iii javelin lung 200 study of avelumab vs docetaxel in platinum-treated locally advanced/metastatic non-small cell lung cancer (nsclc). *Annals of Oncology*, **30**(Issue Supplement 5).

Bornkamp, B., Rufibach, K., Lin, J., Liu, Y., Mehrotra, D. V., Roychoudhury, S., Schmidli, H., Shentu, Y., and Wolbers, M. (2021). Principal stratum strategy: Potential role in drug development. *Pharm Stat, to appear*.

Cameron, D., Piccart-Gebhart, M. J., Gelber, R. D., Procter, M., Goldhirsch, A., de Azambuja, E., Castro, G., Untch, M., Smith, I., Gianni, L., Baselga, J., Al-Sakaff, N., Lauer, S., McFadden, E., Leyland-Jones, B., Bell, R., Dowsett, M., and Jackisch, C. (2017). 11 years' follow-up of trastuzumab after adjuvant chemotherapy in HER2-positive early breast cancer: final analysis of the HERceptin Adjuvant (HERA) trial. *Lancet*, **389**(10075), 1195–1205.

Casey, M., Degtyarev, E., Lechuga, M. J., Aimone, P., Ravaud, A., Motzer, R. J., Liu, F., Stalbovskaya, V., Tang, R., Butler, E., Sailer, O., Halabi, S., and George, D. (2021). Estimand framework: Are we asking the right questions? a case study in the solid tumor setting. *Pharmaceutical Statistics*, **20**(2), 324–334.

Cortes, J. E., Khaled, S., Martinelli, G., Perl, A. E., Ganguly, S., Russell, N., Krämer, A., Dombret, H., Hogge, D., Jonas, B. A., Leung, A. Y.-H., Mehta, P., Montesinos, P., Radsak, M., Sica, S., Arunachalam, M., Holmes, M., Kobayashi, K., Namuyinga, R., Ge, N., Yver, A., Zhang, Y., and Levis, M. J. (2019). Quizartinib versus salvage chemotherapy in relapsed or refractory flt3-itd acute myeloid leukaemia (quantum-r): a multicentre, randomised, controlled, open-label, phase 3 trial. *The Lancet. Oncology*, **20**, 984–997.

References II

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.

Demetri, G. D., Reichardt, P., Kang, Y.-K., Blay, J.-Y., Joensuu, H., Schaefer, K., Wagner, A., Casali, P. G., and Kappeler, C. (2016). Final overall survival (os) analysis with modeling of crossover impact in the phase iii grid trial of regorafenib vs placebo in advanced gastrointestinal stromal tumors (gist).

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.

Gianni, L., Dafni, U., Gelber, R. D., Azambuja, E., Muehlbauer, S., Goldhirsch, A., Untch, M., Smith, I., Baselga, J., Jackisch, C., Cameron, D., Mano, M., Pedrini, J. L., Veronesi, A., Mendiola, C., Pluzanska, A., Semiglazov, V., Vrdoljak, E., Eckart, M. J., Shen, Z., Skiadopoulos, G., Procter, M., Pritchard, K. I., Piccart-Gebhart, M. J., and Bell, R. (2011). Treatment with trastuzumab for 1 year after adjuvant chemotherapy in patients with HER2-positive early breast cancer: a 4-year follow-up of a randomised controlled trial. *Lancet Oncol.*, **12**(3), 236–244.

Hampson, L. V., Bornkamp, B., Holzhauer, B., Kahn, J., Lange, M. R., Luo, W.-L., Cioppa, G. D., Stott, K., and Ballerstedt, S. (2021). Improving the assessment of the probability of success in late stage drug development.

Hemmings, R. (2015). The 'estimand' - problem statement. Presented at PSI 'Estimands Discussion Meeting' in February 2015.

Hernán, M. A. and Robins, J. M. (2020). *Causal Inference: What If*. Chapman & Hall/CRC, Boca Raton.

Herrlinger, U., Schäfer, N., Steinbach, J. P., Weyerbrock, A., Hau, P., Goldbrunner, R., Friedrich, F., Rohde, V., Ringel, F., Schlegel, U., Sabel, M., Ronellenfitsch, M. W., Uhl, M., Maciaczyk, J., Grau, S., Schnell, O., Hänel, M., Krex, D., Vajkoczy, P., Gerlach, R., Kortmann, R.-D., Mehdorn, M., Tüttenberg, J., Mayer-Steinacker, R., Fietkau, R., Brehmer, S., Mack, F., Stuplich, M., Kebir, S., Kohnen, R., Dunkl, E., Leutgeb, B., Proescholdt, M., Pietsch, T., Urbach, H., Belka, C., Stummer, W., and Glas, M. (2016). Bevacizumab plus irinotecan versus temozolomide in newly diagnosed o6-methylguanine-dna methyltransferase nonmethylated glioblastoma: The randomized glarius trial. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, **34**, 1611–1619.

References III

ICH (2019). Addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials E9(R1). Accessible via `https://database.ich.org/sites/default/files/E9-R1_Step4_Guideline_2019_1203.pdf`.

Imbens, G. W. and Rubin, D. B. (2015). *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press, Cambridge.

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.

Kong, S., Heinzmann, D., Lauer, S., and Lu, T. (2021). Weighted approach for estimating effects in principal strata with missing data for a categorical post-baseline variable in randomized controlled trials. *Statistics in Biopharmaceutical Research*, **to appear**.

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.

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. (2021). Estimands in clinical trials with treatment switching. *Pharm Stat, to apper*.

Motzer, R. J., Escudier, B., Oudard, S., Hutson, T. E., Porta, C., Bracarda, S., Grünwald, V., Thompson, J. A., Figlin, R. A., Hollaender, N., Kay, A., Ravaud, A., and Group, R.-. S. (2010). Phase 3 trial of everolimus for metastatic renal cell carcinoma : final results and analysis of prognostic factors. *Cancer*, **116**, 4256–4265.

Pearl, J. (2009). *Reflections, Elaborations, and Discussions with Readers*, pages 331–400. Cambridge University Press.

Piccart-Gebhart, M. J. and Procter, M. e. a. (2005). Trastuzumab after adjuvant chemotherapy in her2-positive breast cancer. *N. Engl. J. Med.*, **353**(16), 1659–1672.

Rubin, D. B. (2008). For objective causal inference, design trumps analysis. *The Annals of Applied Statistics*, **2**(3), 808–840.

References IV

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, H.-J., Butler, E., Rufibach, K., and Roychoudhury, S. (2021). Estimands in hematologic oncology trials. *Pharmaceutical Statistics*, **20**(4), 793–805.

Tang, J., Yu, J. X., Hubbard-Lucey, V. M., Neftelinov, S. T., Hodge, J. P., and Lin, Y. (2018). Trial watch: The clinical trial landscape for pd1/pdl1 immune checkpoint inhibitors. *Nature reviews. Drug discovery*, **17**, 854–855.

Weber, J. S., D'Angelo, S. P., Minor, D., Hodi, F. S., Gutzmer, R., Neyns, B., Hoeller, C., Khushalani, N. I., Miller, W. H., Lao, C. D., Linette, G. P., Thomas, L., Lorigan, P., Grossmann, K. F., Hassel, J. C., Maio, M., Sznol, M., Ascierto, P. A., Mohr, P., Chmielowski, B., Bryce, A., Svane, I. M., Grob, J.-J., Krackhardt, A. M., Horak, C., Lambert, A., Yang, A. S., and Larkin, J. (2015). Nivolumab versus chemotherapy in patients with advanced melanoma who progressed after anti-ctla-4 treatment (checkmate 037): a randomised, controlled, open-label, phase 3 trial. *The Lancet. Oncology*, **16**, 375–384.

Yu, J. X., Hubbard-Lucey, V. M., and Tang, J. (2018). The global pipeline of cell therapies for cancer. *Nat. Rev. Drug Discov*, **17**, 465–466.

Backup

Agenda

Backup: ICH E9(R1) addendum: Why? And what's new?

ICH E9 draft addendum

ICH E9: “Statistical principles for Clinical Trials.”

ICH E9 draft addendum

ICH E9: "Statistical principles for Clinical Trials."

1998.

ICH E9 draft addendum

ICH E9: "Statistical principles for Clinical Trials."

1998.

Why amend E9?

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".

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!

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!

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

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,

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.

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!

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

Understanding treatment effects

- Multiple definitions of **treatment effect**.

Understanding treatment effects

- Multiple definitions of **treatment effect**.
- Different definitions addressing **different scientific questions**.

Understanding treatment effects

- Multiple definitions of **treatment effect**.
- Different definitions addressing **different scientific questions**.
- Not all equally acceptable for **regulatory decision making**.

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.

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**.

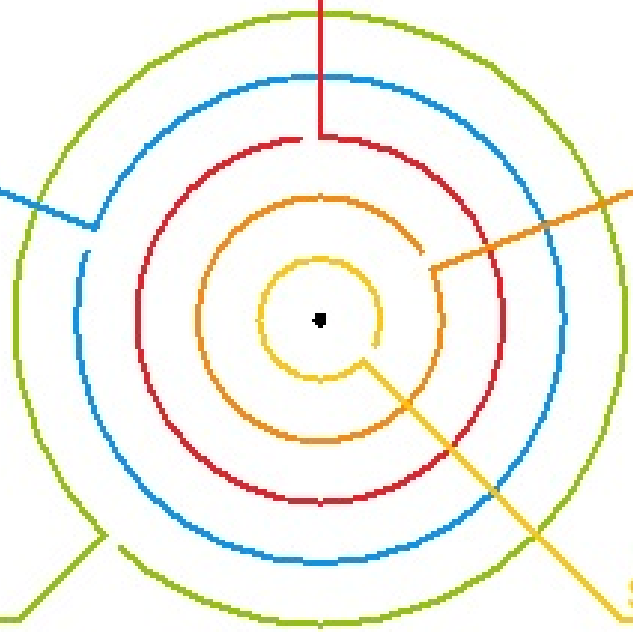
How does the addendum fix this?

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.

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.

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

Backup: 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.
- **77** members (30 EU + 38 US + 9 Asia) representing **37** companies / institutions.
- Regularly interacts with **8 health authorities**.
- Presentations, webinars, papers.

www.oncoestimand.org


NOVARTIS
Roche
Pfizer
MERCK
janssen
PHARMACEUTICAL COMPANIES OF Johnson&Johnson
statisticon
SANOFI
gsk
abbvie
AstraZeneca
BeiGene
IQVIA
BAYER
Takeda
Incyte
astellas
SERVIER
MODUS OUTCOMES
MERCK
PPD
NEKTAR
Merus battling cancer
CLINICAL OUTCOMES SOLUTIONS
Innovative science to capture the patient voice
stabu ro
Statistical Consulting
AMGEN
ICON
A Symbol of Excellence
Boehringer Ingelheim
phastar
Adelphi
ADELPHI VALUES
Daiichi-Sankyo
MERCK SERONO
zai Lab
Evidera
Bristol Myers Squibb
Genmab
CANCER RESEARCH UK
BIRMINGHAM CANCER RESEARCH UK CLINICAL TRIALS UNIT
UNIVERSITY OF BIRMINGHAM

Papers

Published:

- 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.* (2021): Estimand framework: Are we asking the right question? A case study in the solid tumor setting. **link**
- Sun *et al.* (2021): Estimands in Hematology Trials. **link**
- Manitz *et al.* (2021): Estimands in clinical trials with treatment switching. **link**
- Bornkamp *et al.* (2021): Principal Stratum Strategy: Potential Role in Drug Development. **link** (incl. markdown file with code).
- Hampson *et al.* (2021): Comment on FDA paper on *Biostatistical Considerations When Using RWD and RWE in Clinical Studies for Regulatory Purposes.* **link**

More papers under preparation.

Task forces

- Estimands engagement.
- Principal stratification in clinical trials.
- Patient-reported outcomes.
- Duration of responses.
- Quantification of follow-up.
- Real-world data and estimands.
- Conditional vs. marginal effects.
- Time to event endpoints with prognostic or predictive biomarker subgroups.

Agenda

Backup: Subgroups by post-randomization event - principal stratification

"... 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 (2019)

Principal stratification:

Principal stratification:

- Originates in causal inference: Frangakis and Rubin (2002).

Principal stratification:

- Originates in causal inference: Frangakis and Rubin (2002).
- Framework for comparing treatments adjusting for **posttreatment** variables.

Principal stratification:

- Originates in causal inference: Frangakis and Rubin (2002).
- Framework for comparing treatments adjusting for **posttreatment** variables.
- Formulated within **potential outcomes** framework.

Principal stratification:

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

Principal stratification:

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

Introductory books causal inference: Imbens and Rubin (2015), Hernán and Robins (2020).

First, let us summarize what does not work.

2-arm RCT test (T) vs. control (C)

2-arm RCT test (T) vs. control (C)

Do responders have higher treatment effect?

2-arm RCT test (T) vs. control (C)

Do responders have higher treatment effect?

"Subgroup" built by post-randomization event!

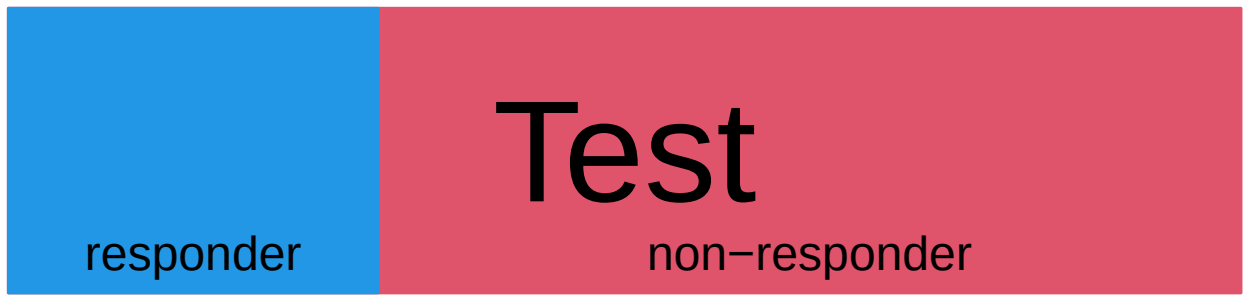
How can we make valid causal statements?

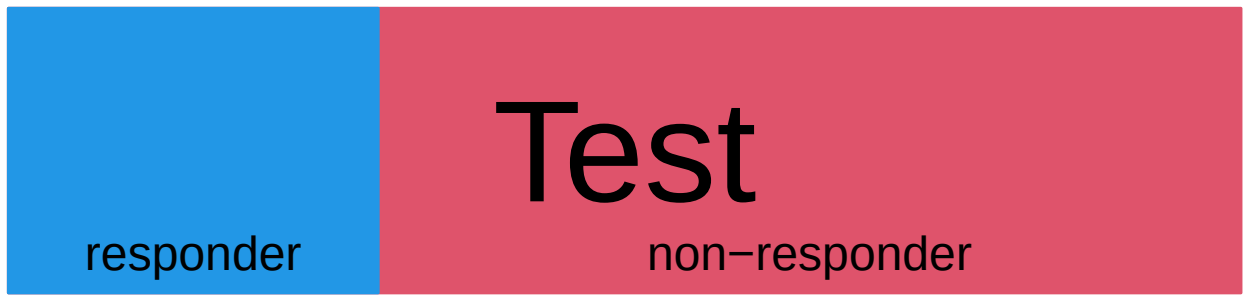
How can we make valid causal statements?

Need "matched control patients"!


Test


Control


Test
responder
non−responder



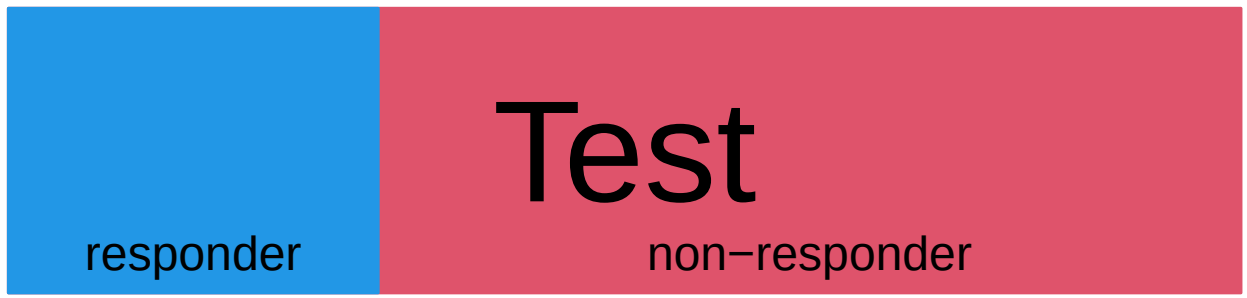
Patients who respond
if randomized to Test
had they received control

Test

Control

Test

Control


Test
responder
non−responder

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

Naive analyses are misleading and do not answer causal question

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

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 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 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 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

Bornkamp *et al.* (2021).

CAR-T example - see later!

OS / PFS by response.

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).

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

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: $Y(Z = 0) = Y(0)$,

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, $Y(Z = 1) = Y(1)$.

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!

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 effect** $Y(1) - Y(0)$ not observed.

What are causal effects?

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

What are causal effects?

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

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

What are causal effects?

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

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

Causal treatment effect:

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. $\{Y(0)_i, i \in \mathcal{S}\}$.

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". **Hypothetical scenario**.

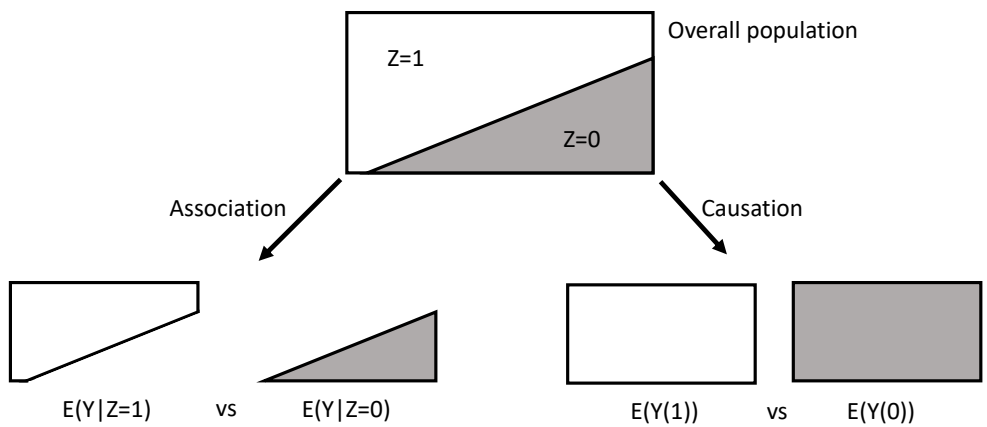
What are causal effects?

$Y(1)_i$: 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". **Hypothetical scenario**.



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

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 event, e.g. responder.

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 event, e.g. responder.

- Compare patients with $S = 1$ on both test and control arm.

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 event, e.g. responder.

- Compare patients with $S = 1$ on both test and control arm.
- RCT: $S(Z)$ **post-randomization** $\Rightarrow$ S **depends on** Z!

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 event, e.g. responder.

- 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**!

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 event, e.g. responder.

- 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**.

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 event, e.g. responder.

- 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

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 event, e.g. responder.

- 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

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 event, e.g. responder.

- 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.

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 event, e.g. responder.

- 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 $S(1) = S(0) \Rightarrow$ response 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.

Principal stratification

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

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

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.

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.

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

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 $t^* > \tilde{t}$:

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

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

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

Backup: Subgroups by post-randomization event - principal stratification #98 /

Potential outcomes, estimands, and PS

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

Potential outcomes, estimands, and PS

All estimand strategies can be 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.

Sensitivity analyses!

Assumptions for estimation (see backup) **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.

Sensitivity analyses!

Assumptions for estimation (see backup) **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:

Conclusions principal stratification

Conclusions:

- Many relevant examples in drug development.

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 or CAR-T. **Both explicitly requested by HAs**!

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 or CAR-T. **Both explicitly requested by HAs**!
- Naive analyses often standard: Unclear estimand $\Rightarrow$ **causal conclusion unclear**.

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 or CAR-T. **Both explicitly requested by HAs**!
- Naive analyses often standard: Unclear estimand $\Rightarrow$ **causal conclusion unclear**.
- Complex question $\Rightarrow$ complex analysis needed.

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 or CAR-T. **Both explicitly requested by HAs**!
- Naive analyses often standard: Unclear estimand $\Rightarrow$ **causal conclusion unclear**.
- Complex question $\Rightarrow$ complex analysis needed.
- Assumptions needed: scientific input + sensitivity analyses.

DOI: 10.1002/pst.2104

MAIN PAPER

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)

Markdown:

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

DOI: 10.1002/pst.2104

MAIN PAPER

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)

Markdown:

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

Effective statistician podcast, Björn Bornkamp and Kaspar Rufibach:

https://theeffectivestatistician.com/

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

arXiv.org > stat > arXiv:2101.04263

Search...

Help | Ad

Statistics > Methodology

[Submitted on 12 Jan 2021]

Weighted Approach for Estimating Effects in Principal Strata with Missing Data for a Categorical Post-Baseline Variable in Randomized Controlled Trials

Shengchun Kong, Dominik Heinzmann, Sabine Lauer, Tian Lu

This research was motivated by studying anti-drug antibody (ADA) formation and its potential impact on long-term benefit of a biologic treatment in a randomized controlled trial, in which ADA status was not only unobserved in the control arm but also in a subset of patients from the experimental treatment arm. Recent literature considers the principal stratum estimand strategy to estimate treatment effect in groups of patients defined by an intercurrent status, i.e. in groups defined by a post-randomization variable only observed in one arm and potentially associated with the outcome. However, status information might be missing even for a non-negligible number of patients in the experimental arm. For this setting, a novel weighted principal stratum approach is presented: Data from patients with missing intercurrent event status were re-weighted based on baseline covariates and additional longitudinal information. A theoretical justification of the proposed approach is provided for different types of outcomes, and assumptions allowing for causal conclusions on treatment effect are specified and investigated. Simulations demonstrated that the proposed method yielded valid inference and was robust against certain violations of assumptions. The method was shown to perform well in a clinical study with ADA status as an intercurrent event.

Kong et al. (2021)

Github repository: https://github.com/openpharma/BBS-causality-training

arXiv.org > stat > arXiv:2101.04263

Search...

Help | Ad

Statistics > Methodology

[Submitted on 12 Jan 2021]

Weighted Approach for Estimating Effects in Principal Strata with Missing Data for a Categorical Post-Baseline Variable in Randomized Controlled Trials

Shengchun Kong, Dominik Heinzmann, Sabine Lauer, Tian Lu

This research was motivated by studying anti-drug antibody (ADA) formation and its potential impact on long-term benefit of a biologic treatment in a randomized controlled trial, in which ADA status was not only unobserved in the control arm but also in a subset of patients from the experimental treatment arm. Recent literature considers the principal stratum estimand strategy to estimate treatment effect in groups of patients defined by an intercurrent status, i.e. in groups defined by a post-randomization variable only observed in one arm and potentially associated with the outcome. However, status information might be missing even for a non-negligible number of patients in the experimental arm. For this setting, a novel weighted principal stratum approach is presented: Data from patients with missing intercurrent event status were re-weighted based on baseline covariates and additional longitudinal information. A theoretical justification of the proposed approach is provided for different types of outcomes, and assumptions allowing for causal conclusions on treatment effect are specified and investigated. Simulations demonstrated that the proposed method yielded valid inference and was robust against certain violations of assumptions. The method was shown to perform well in a clinical study with ADA status as an intercurrent event.

Kong et al. (2021)

Github repository: https://github.com/openpharma/BBS-causality-training

Talk Dominik in BBS seminar: http://bbs.ceb-institute.org/?p=1668

Agenda

Backup: Estimation of average causal effect

Estimation of average causal effect

Key assumptions:

Estimation of average causal effect

Key assumptions:

- **Exchangeability**: Counterfactual outcomes independent of treatment assignment

Estimation of average causal effect

Key assumptions:

- **Exchangeability**: Counterfactual outcomes independent of treatment assignment $\Leftrightarrow$ $Y(1)$ and $Y(0)$ independent of Z.

Estimation of average causal effect

Key assumptions:

- **Exchangeability**: Counterfactual outcomes independent of treatment assignment $\Leftrightarrow$ $Y(1)$ and $Y(0)$ independent of Z. Trivially fulfilled in **RCT**. Via propensity scores otherwise.
- **Consistency**: No multiple versions of treatment

Estimation of average causal effect

Key assumptions:

- **Exchangeability**: Counterfactual outcomes independent of treatment assignment $\Leftrightarrow$ $Y(1)$ and $Y(0)$ independent of Z. Trivially fulfilled in **RCT**. Via propensity scores otherwise.
- **Consistency**: No multiple versions of treatment $\Leftrightarrow$ individual's PO under observed exposure IS her observed outcome

Estimation of average causal effect

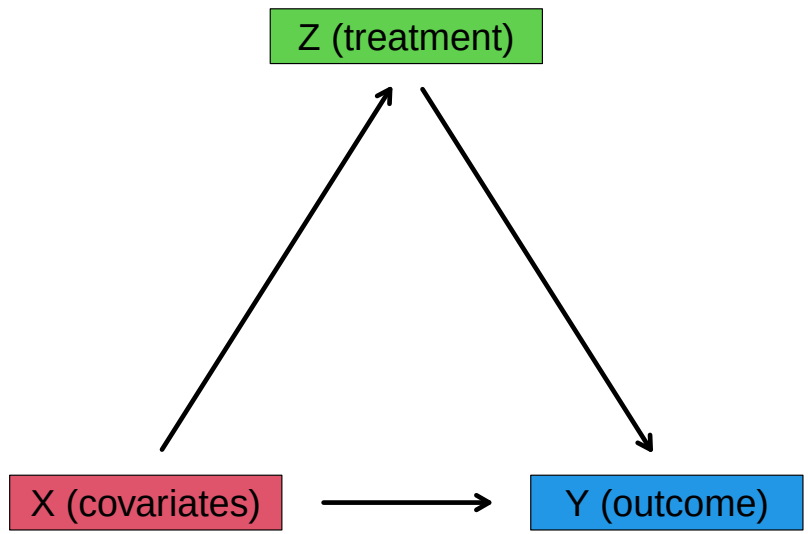
Key assumptions:

- **Exchangeability**: Counterfactual outcomes independent of treatment assignment $\Leftrightarrow$ $Y(1)$ and $Y(0)$ independent of Z. Trivially fulfilled in **RCT**. Via propensity scores otherwise.
- **Consistency**: No multiple versions of treatment $\Leftrightarrow$ individual's PO under observed exposure IS her observed outcome $\Leftrightarrow$ $E(Y(x)|Z = x) = E(Y|Z = x), x = 0, 1.$

Estimation of average causal effect

Key assumptions:

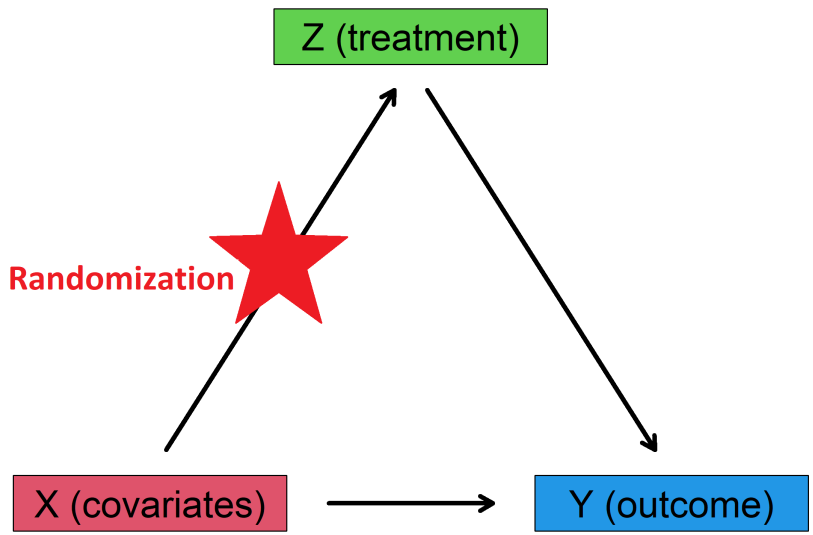
- **Exchangeability**: Counterfactual outcomes independent of treatment assignment $\Leftrightarrow$ $Y(1)$ and $Y(0)$ independent of Z. Trivially fulfilled in **RCT**. Via propensity scores otherwise.
- **Consistency**: No multiple versions of treatment $\Leftrightarrow$ individual's PO under observed exposure IS her observed outcome $\Leftrightarrow$ $E(Y(x)|Z = x) = E(Y|Z = x), x = 0, 1.$



Estimation of average causal effect

Key assumptions:

- **Exchangeability**: Counterfactual outcomes independent of treatment assignment $\Leftrightarrow$ $Y(1)$ and $Y(0)$ independent of Z. Trivially fulfilled in **RCT**. Via propensity scores otherwise.
- **Consistency**: No multiple versions of treatment $\Leftrightarrow$ individual's PO under observed exposure IS her observed outcome $\Leftrightarrow$ $E(Y(x)|Z = x) = E(Y|Z = x), x = 0, 1$.



Estimation of average causal effect

Key assumptions:

- **Exchangeability**: Counterfactual outcomes independent of treatment assignment $\Leftrightarrow$ $Y(1)$ and $Y(0)$ independent of Z. Trivially fulfilled in **RCT**. Via propensity scores otherwise.
- **Consistency**: No multiple versions of treatment $\Leftrightarrow$ individual's PO under observed exposure IS her observed outcome $\Leftrightarrow$ $E(Y(x)|Z = x) = E(Y|Z = x), x = 0, 1.$

Estimation of average causal effect

Key assumptions:

- **Exchangeability**: Counterfactual outcomes independent of treatment assignment $\Leftrightarrow$ $Y(1)$ and $Y(0)$ independent of Z. Trivially fulfilled in **RCT**. Via propensity scores otherwise.
- **Consistency**: No multiple versions of treatment $\Leftrightarrow$ individual's PO under observed exposure IS her observed outcome $\Leftrightarrow$ $E(Y(x)|Z = x) = E(Y|Z = x), x = 0, 1$.

$$
\begin{aligned}
E(Y(1) - Y(0)) &\overset{\textit{linearity of } E}{=} E(Y(1)) - E(Y(0)) \\
&\overset{\textit{exchangeability}}{=} E(Y(1)|Z = 1) - E(Y(0)|Z = 0) \\
&\overset{\textit{consistency}}{=} E(Y|Z = 1) - E(Y|Z = 0).
\end{aligned}
$$

Estimation of average causal effect

Key assumptions:

- **Exchangeability**: Counterfactual outcomes independent of treatment assignment $\Leftrightarrow$ $Y(1)$ and $Y(0)$ independent of Z. Trivially fulfilled in **RCT**. Via propensity scores otherwise.
- **Consistency**: No multiple versions of treatment $\Leftrightarrow$ individual's PO under observed exposure IS her observed outcome $\Leftrightarrow$ $E(Y(x)|Z = x) = E(Y|Z = x), x = 0, 1$.

$$
\begin{aligned}
E(Y(1) - Y(0)) &\overset{\textit{linearity of } E}{=} E(Y(1)) - E(Y(0)) \\
&\overset{\textit{exchangeability}}{=} E(Y(1)|Z = 1) - E(Y(0)|Z = 0) \\
&\overset{\textit{consistency}}{=} E(Y|Z = 1) - E(Y|Z = 0).
\end{aligned}
$$

So - why do we randomize?

Estimation of average causal effect

Key assumptions:

- **Exchangeability**: Counterfactual outcomes independent of treatment assignment $\Leftrightarrow$ $Y(1)$ and $Y(0)$ independent of Z. Trivially fulfilled in **RCT**. Via propensity scores otherwise.
- **Consistency**: No multiple versions of treatment $\Leftrightarrow$ individual's PO under observed exposure IS her observed outcome $\Leftrightarrow$ $E(Y(x)|Z = x) = E(Y|Z = x), x = 0, 1$.

$$
\begin{aligned}
E(Y(1) - Y(0)) &\overset{\textit{linearity of } E}{=} E(Y(1)) - E(Y(0)) \\
&\overset{\textit{exchangeability}}{=} E(Y(1)|Z = 1) - E(Y(0)|Z = 0) \\
&\overset{\textit{consistency}}{=} E(Y|Z = 1) - E(Y|Z = 0).
\end{aligned}
$$

So - why do we randomize?

- To balance covariates?

Estimation of average causal effect

Key assumptions:

- **Exchangeability**: Counterfactual outcomes independent of treatment assignment $\Leftrightarrow$ $Y(1)$ and $Y(0)$ independent of Z. Trivially fulfilled in **RCT**. Via propensity scores otherwise.
- **Consistency**: No multiple versions of treatment $\Leftrightarrow$ individual's PO under observed exposure IS her observed outcome $\Leftrightarrow$ $E(Y(x)|Z = x) = E(Y|Z = x), x = 0, 1$.

$$
\begin{aligned}
E(Y(1) - Y(0)) &\overset{\textit{linearity of } E}{=} E(Y(1)) - E(Y(0)) \\
&\overset{\textit{exchangeability}}{=} E(Y(1)|Z = 1) - E(Y(0)|Z = 0) \\
&\overset{\textit{consistency}}{=} E(Y|Z = 1) - E(Y|Z = 0).
\end{aligned}
$$

So - why do we randomize?

- To balance covariates? **NO!**

Estimation of average causal effect

Key assumptions:

- **Exchangeability**: Counterfactual outcomes independent of treatment assignment $\Leftrightarrow$ $Y(1)$ and $Y(0)$ independent of Z. Trivially fulfilled in **RCT**. Via propensity scores otherwise.
- **Consistency**: No multiple versions of treatment $\Leftrightarrow$ individual's PO under observed exposure IS her observed outcome $\Leftrightarrow$ $E(Y(x)|Z = x) = E(Y|Z = x), x = 0, 1$.

$$
\begin{aligned}
E(Y(1) - Y(0)) &\overset{\textit{linearity of } E}{=} E(Y(1)) - E(Y(0)) \\
&\overset{\textit{exchangeability}}{=} E(Y(1)|Z = 1) - E(Y(0)|Z = 0) \\
&\overset{\textit{consistency}}{=} E(Y|Z = 1) - E(Y|Z = 0).
\end{aligned}
$$

So - why do we randomize?

- To balance covariates? **NO!**
- Covariates do not appear at all in above computation!

Estimation of average causal effect

Key assumptions:

- **Exchangeability**: Counterfactual outcomes independent of treatment assignment $\Leftrightarrow$ $Y(1)$ and $Y(0)$ independent of Z. Trivially fulfilled in **RCT**. Via propensity scores otherwise.
- **Consistency**: No multiple versions of treatment $\Leftrightarrow$ individual's PO under observed exposure IS her observed outcome $\Leftrightarrow$ $E(Y(x)|Z = x) = E(Y|Z = x), x = 0, 1$.

$$
\begin{aligned}
E(Y(1) - Y(0)) &\overset{\textit{linearity of } E}{=} E(Y(1)) - E(Y(0)) \\
&\overset{\textit{exchangeability}}{=} E(Y(1)|Z = 1) - E(Y(0)|Z = 0) \\
&\overset{\textit{consistency}}{=} E(Y|Z = 1) - E(Y|Z = 0).
\end{aligned}
$$

So - why do we randomize?

- To balance covariates? **NO!**
- Covariates do not appear at all in above computation!
- Randomization generates equal distributions (in both groups) of **potential outcomes**!

For example, one would be extremely hard pressed to find a statistics textbook, even at the graduate level, containing a mathematical proof that randomization indeed produces unbiased estimates of the quantities we wish estimated – i.e., efficacy of treatments or policies.

Judea Pearl, American computer scientist and philosopher

Pearl (2009)

Estimation of average causal effect

Observational study:

Estimation of average causal effect

Observational study:

- Decision between $Z = 0$ and $Z = 1$ might depend on X (measured or unmeasured).

Estimation of average causal effect

Observational study:

- Decision between $Z = 0$ and $Z = 1$ might depend on X (measured or unmeasured).
- $Y(1)$ and $Y(0)$ **not** independent of $Z \Rightarrow$ exchangeability violated $\Rightarrow E(Y(1)) \neq E(Y(1)|Z = 1)$ and $E(Y(0)) \neq E(Y(0)|Z = 0)$.

Estimation of average causal effect

Observational study:

- Decision between $Z = 0$ and $Z = 1$ might depend on X (measured or unmeasured).
- $Y(1)$ and $Y(0)$ **not** independent of $Z \Rightarrow$ exchangeability violated $\Rightarrow E(Y(1)) \neq E(Y(1)|Z = 1)$ and $E(Y(0)) \neq E(Y(0)|Z = 0)$.
- 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 $Y(0)$).

Estimation of average causal effect

Observational study:

- Decision between $Z = 0$ and $Z = 1$ might depend on X (measured or unmeasured).
- $Y(1)$ and $Y(0)$ **not** independent of $Z \Rightarrow$ exchangeability violated $\Rightarrow E(Y(1)) \neq E(Y(1)|Z = 1)$ and $E(Y(0)) \neq E(Y(0)|Z = 0)$.
- 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 $Y(0)$).
- Patients receiving $Z = 0$ **not representative** of overall population.

Estimation of average causal effect

Observational study:

- Decision between $Z = 0$ and $Z = 1$ might depend on X (measured or unmeasured).
- $Y(1)$ and $Y(0)$ **not** independent of $Z \Rightarrow$ exchangeability violated $\Rightarrow E(Y(1)) \neq E(Y(1)|Z = 1)$ and $E(Y(0)) \neq E(Y(0)|Z = 0)$.
- 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 $Y(0)$).
- Patients receiving $Z = 0$ **not representative** of overall population.

$$
\begin{aligned}
E(Y(1) - Y(0)) &\overset{\textit{linearity of } E}{=} E(Y(1)) - E(Y(0)) \\
&\overset{\cancel{\textit{exchangeability}}}{\neq} E(Y(1)|Z = 1) - E(Y(0)|Z = 0) \\
&\overset{\textit{consistency}}{=} E(Y|Z = 1) - E(Y|Z = 0).
\end{aligned}
$$

Agenda

Backup: Estimation of principal effects

Assumptions

Randomization not enough to estimate 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.
  - Infectious diseases: treatment may change depending on who else is vaccinated $\Rightarrow$ violation.

Estimation

SUTVA:

- Underpins virtually all estimation methods.
- POs for any patient do not change with treatment assigned to other patients.
  - Infectious diseases: treatment may change depending on who else is vaccinated $\Rightarrow$ violation.

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.

Estimation

Exclusion-restriction:

- Assume $Y(0) = Y(1)$ (no treatment effect) for patients $\{S(0) = 0\} \cap \{S(1) = 0\}$ and $\{S(0) = 1\} \cap \{S(1) = 1\}$.

	$S(0) = 1$	$S(0) = 0$
$S(1) = 1$	no causal effect of Z on Y	$\{S(1) = 1\} \cap \{S(0) = 0\}$
$S(1) = 0$	$\{S(1) = 0\} \cap \{S(0) = 1\}$	no causal effect of Z on Y

- Randomization Z exclusively affects outcome through intercurrent event S.
- Angrist *et al.* (1996), 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** for 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**.

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

Estimation under principal ignorability for ADA example

- Estimate $P(S(1) = 1|X)$ on treatment arm using logistic regression.

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**.

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.

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:

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!

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**.

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**.

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).

Estimation under principal ignorability for ADA example

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: predictors of S and Y. Similar to observational studies: $X =$ predictors of treatment and outcome.
- **Do not include** covariates that "only" help predict S but have no impact on Y.
- Similar to considerations for observational studies.

Doing now what patients need next

R version and packages used to generate these slides:

R version: R version 4.1.1 (2021-08-10)

Base packages: stats / graphics / grDevices / utils / datasets / methods / base

Other packages: prodlim

This document was generated on 2021-12-14 at 16:35:43.