Estimands: a genuine step forward in designing, running, analyzing, and communicating clinical trials

Kaspar Rufibach
Advanced Biostatistical Sciences, Merck KGaA, Darmstadt, Germany
Mayo clinic biostatistics seminar, 23th January 2025

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

Lew Sheiner
American Clinical Pharmacologist

Sheiner (1991)

Agenda

1. Case study: POLARIX
2. Impact
3. How about regulators?
4. Resources

Agenda backup

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

6. Estimands
   - ICH E9 estimands addendum
   - Intercurrent events
   - Estimation
   - How do estimands relate to "ITT"?
   - What about "per-protocol populations"?
   - Impact

7. Case study: RATIFY

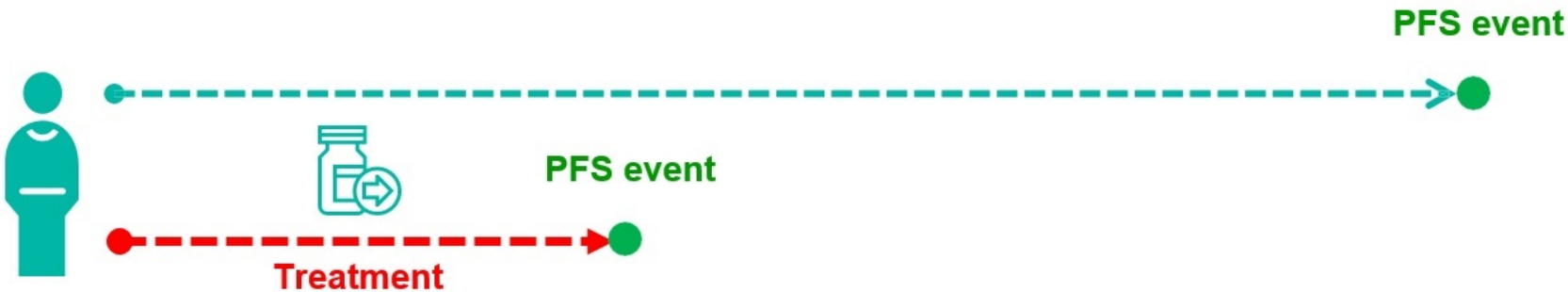
8. Backup: Hypothetical strategy to address ICEs: application to Covid-19

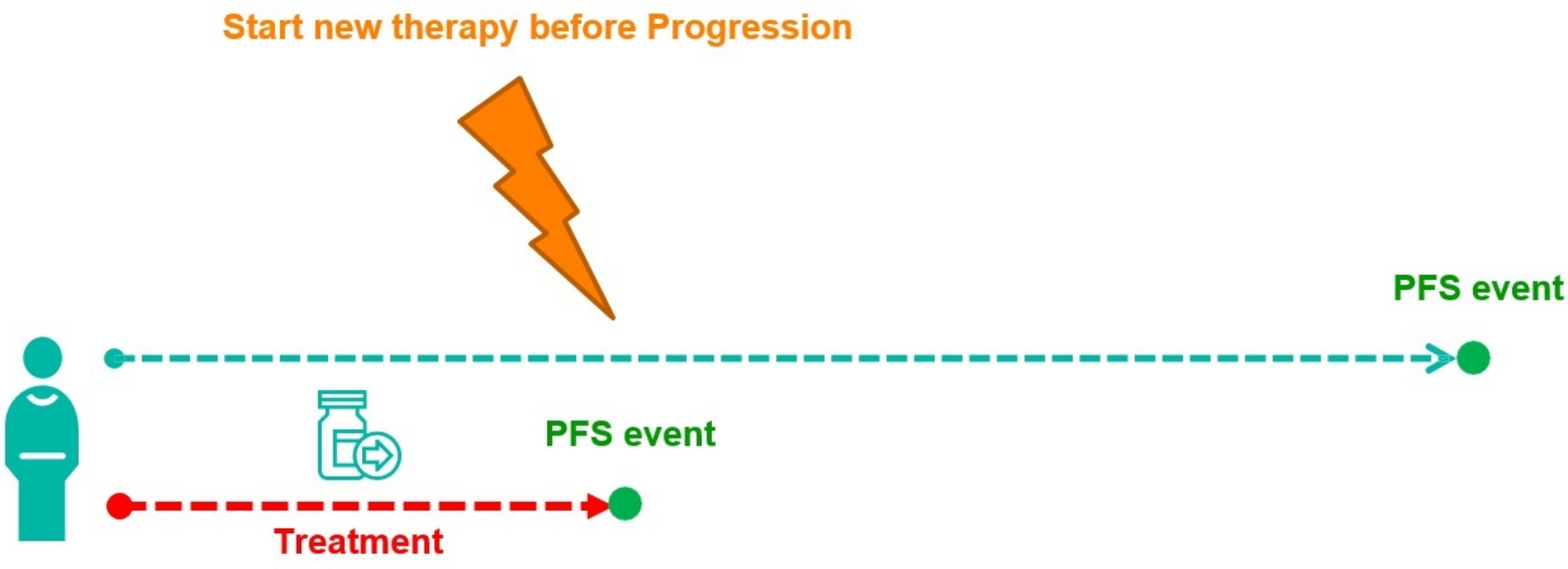
9. Backup: Case study: treatment switching

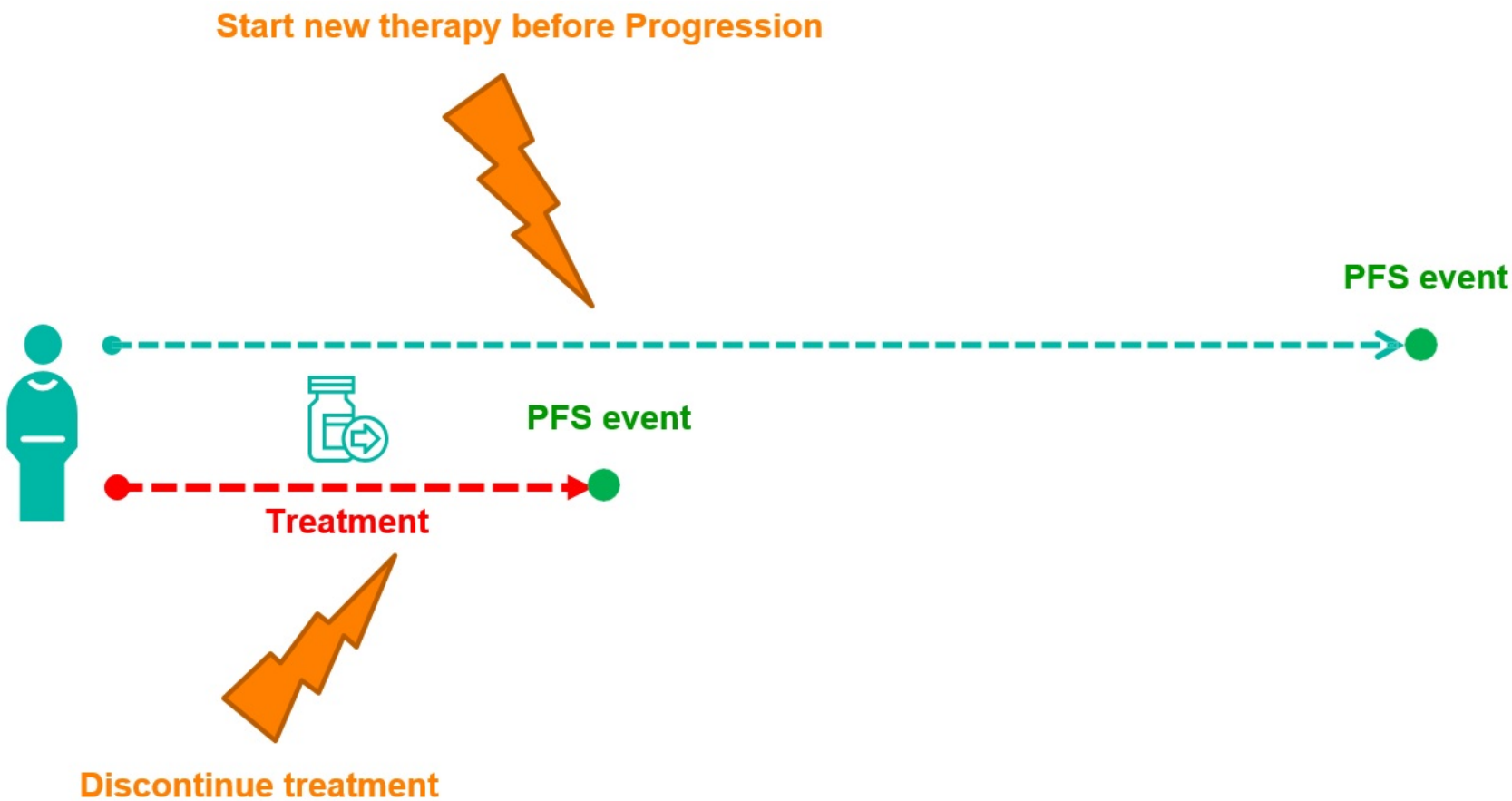
Agenda

Case study: POLARIX


PFS event
Treatment


PFS event
PFS event
Treatment


Start new therapy before Progression
PFS event
PFS event
Treatment


Start new therapy before Progression
PFS event
PFS event
Treatment
Discontinue treatment

Do these clinical events affect your interpretation of the treatment effect?

Is the treatment effect clearly defined?

What data would you collect?

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

W.E. Deming, American Statistician

Polarix Oncologic Drugs Advisory Committee (ODAC).

2-arm RCT in DLBCL.
R-CHOP vs. R-CH-Polatuzumab-P.
Primary endpoint: ”PFS”.

Is it clear what ”PFS” is?

Estimand attribute	Analysis 1 (pre-specified in SAP): PFS as per protocol	Analysis 2 (requested by FDA): PFS with censoring at NALT
Population	As per protocol	
Endpoint	PFS: time to PD or death	
Summary measure	Hazard ratio	
Treatment conditions	As per protocol	

Estimand attribute	Analysis 1 (pre-specified in SAP): PFS as per protocol	Analysis 2 (requested by FDA): PFS with censoring at NALT
Population	As per protocol	
Endpoint	PFS: time to PD or death	
Summary measure	Hazard ratio	
Treatment conditions	As per protocol	
Intercurrent events and handling strategy	NALT Treatment policy	NALT “censoring”?

Estimand attribute	Analysis 1 (pre-specified in SAP): PFS as per protocol	Analysis 2 (requested by FDA): PFS with censoring at NALT
Population	As per protocol	
Endpoint	PFS: time to PD or death	
Summary measure	Hazard ratio	
Treatment conditions	As per protocol	
Intercurrent events and handling strategy	NALT Treatment policy	NALT “censoring”?
P-value	0.0177	0.0567

Estimand attribute	Analysis 1 (pre-specified in SAP): PFS as per protocol	Analysis 2 (requested by FDA): PFS with censoring at NALT
Population	As per protocol	
Endpoint	PFS: time to PD or death	
Summary measure	Hazard ratio	
Treatment conditions	As per protocol	
Intercurrent events and handling strategy	NALT Treatment policy	NALT “censoring”?
P-value	0.0177	0.0567
Implied scientific question	What is the time to PD / death irrespective of taking NALT?	What is the time to PD / death assuming NALT would not exist?

Agenda

Impact

Past: too sloppy in translating clinical trial objectives to clear statistical quantities.

1) Stakeholders not aligned.

2) Analysis method not aligned to scientific question.

3) Data collection requirements unclear.

4) Heterogeneity between trials.

Present and future:

ICH E9(R1) estimands addendum.

Clear upfront definition of treatment effect of interest.

Broader use of causal inference methods.

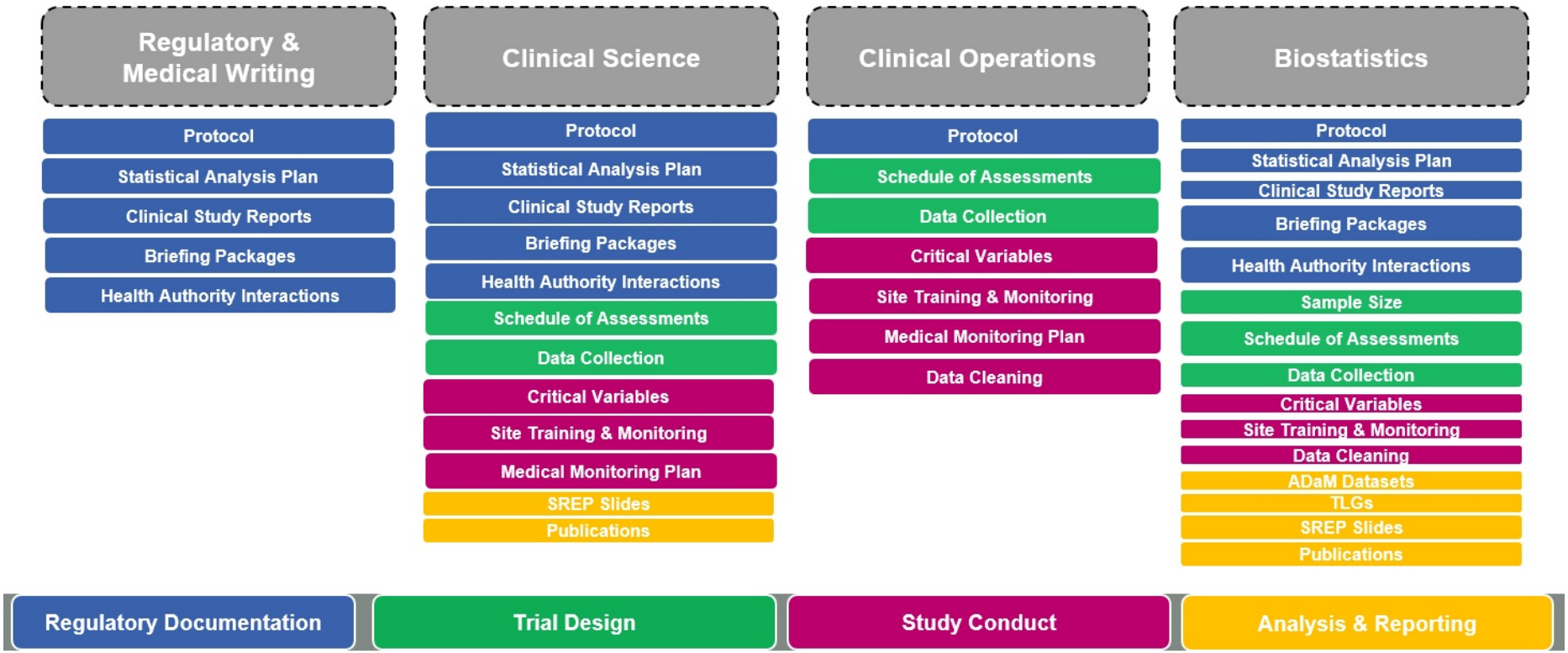
Have discussions upfront.

Get clarity early on.

Shorten filing timelines.

Do I need to care?

Yes!


Regulatory & Medical Writing
Protocol
Statistical Analysis Plan
Clinical Study Reports
Briefing Packages
Health Authority Interactions
Clinical Science
Protocol
Statistical Analysis Plan
Clinical Study Reports
Briefing Packages
Health Authority Interactions
Schedule of Assessments
Data Collection
Critical Variables
Site Training & Monitoring
Medical Monitoring Plan
SREP Slides
Publications
Clinical Operations
Protocol
Schedule of Assessments
Data Collection
Critical Variables
Site Training & Monitoring
Medical Monitoring Plan
Data Cleaning
Biostatistics
Protocol
Statistical Analysis Plan
Clinical Study Reports
Briefing Packages
Health Authority Interactions
Sample Size
Schedule of Assessments
Data Collection
Critical Variables
Site Training & Monitoring
Data Cleaning
ADaM Datasets
TLGs
SREP Slides
Publications
Regulatory Documentation
Trial Design
Study Conduct
Analysis & Reporting

Covid.

Ukraine war.

Patients!

Physicians. Investigators.

Trial developers.

Regulators.

HTA bodies.

Impact on data collection and trial planning

- Definition of estimands(s) requires **multi-disciplinary** involvement from **earliest stages** of clinical trial development.
- Estimand **dictates data that need to be collected**.
- Each trial likely to have **multiple estimands** for **multiple stakeholders** $\Rightarrow$ different estimands might require different data!
- Impacts **design of eCRF** or other data collection tools, and monitoring strategy.
- Increased effort in recording reasons underlying **treatment or trial 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, missing data, 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.

Agenda

How about regulators?

How about regulators?

Ionan *et al.* (2023) (paper written jointly by FDA and industry authors):

- "Statistical and clinical colleagues typically collaborate closely during the FDA review of regulatory submissions. **Use of the estimand framework can improve the efficiency and quality of this collaboration.** Collaborative discussions are sometimes especially challenging due to multiple complex trial design and analysis issues. **The estimand framework provides a structure to facilitate such discussions.**"
- "This framework has already proved very useful, not only in tackling new questions but also in **understanding better "old" problems**. Our subjective experience has been that **estimand thinking has been well-accepted so far and that uptake is good.**"

Shanti Gomatam (FDA statistician) in BBS seminar on 12th April 2023:

- "My appreciation of the estimand framework has increased over time."
- "The estimand framework is useful even in cases where we do not officially implement it. It helps me to get points across more precisely."

Precise formulation of clinical question of interest: not a stats thing!

Collaborative effort.

ICH M11, Transcelerate, industry protocol templates.

Estimands are here to stay.

Agenda

1. Case study: POLARIX
2. Impact
3. How about regulators?
4. Resources

Resources

Resources

Original ICH E9: ICH (1998).

ICH E9 addendum: ICH (2019).

Scientific literature.

Industry association special interest groups: `www.oncoestimand.org`, Estimands in neuroscience, Estimands implementation working group.

A problem well put is half solved.

John Dewey
American Philosopher and Educator

Design trumps analysis.

Don Rubin, American Statistician

Rubin (2008)

Thank you for your attention.

kaspar.rufibach@merckgroup.com

www.kasparrufibach.ch

References I

Akacha, M., Bretz, F., and Ruberg, S. (2017). Estimands in clinical trials - broadening the perspective. *Stat Med*, **36**(1), 5–19.

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).

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.

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.

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).

Fletcher, C., Hefting, N., Wright, M., Bell, J., Anzures-Cabrera, J., Wright, D., Lynggaard, H., and Schueler, A. (2022). Marking 2-Years of New Thinking in Clinical Trials: The Estimand Journey. *Ther Innov Regul Sci*, **56**(4), 637–650.

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.

References II

Hemmings, R. (2015). The 'estimand' - problem statement.

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.

ICH (1998). Statistical principles for clinical trials.

ICH (2019). Addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials E9(R1).

Ionan, A. C., Paterniti, M., Mehrotra, D. V., Scott, J., Ratitch, B., Collins, S., Gomatam, S., Nie, L., Rufibach, K., and Bretz, F. (2023). Clinical and statistical perspectives on the ich e9(r1) estimand framework implementation. *Statistics in Biopharmaceutical Research*, **15**(3), 554–559.

Manitz, J., Kan-Dobrosky, N., Buchner, H., Casadebaig, M.-L., Degtyarev, E., Dey, J., Haddad, V., Jie, F., Martin, E., Mo, M., Rufibach, K., Shentu, Y., Stalbovskaya, V., (Sammi) Tang, R., Yung, G., and Zhou, J. (2022). Estimands for overall survival in clinical trials with treatment switching in oncology. *Pharmaceutical Statistics*, **21**(1), 150–162.

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.

Noci, A., Wolbers, M., Abt, M., Baayen, C., Burger, H. U., Jin, M., and Robieson, W. Z. (2021). A comparison of estimand and estimation strategies for clinical trials in early parkinson's disease.

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 III

Schulz, K. F., Altman, D. G., Moher, D., Altman, D. G., Barbour, V., Boutron, I., Devereaux, P. J., Dickersin, K., Elbourne, D., Ellenberg, S., Gebski, V., Goodman, S., Gotzsche, P. C., Groves, T., Grunberg, S., Haynes, B., Hopewell, S., Juhn, P., Middleton, P., Minckler, D., Moher, D., Montori, V. M., Mulrow, C., Pocock, S., Rennie, D., Schriger, D. L., Schulz, K. F., Simera, I., and Wager, E. (2010). CONSORT 2010 statement: upyeard guidelines for reporting parallel group randomised trials. *BMJ*, **340**, c332.

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.

Wolbers, M., Noci, A., Delmar, P., Gower-Page, C., Yiu, S., and Bartlett, J. W. (2021). Standard and reference-based imputation methods based on conditional mean imputation.

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

1. Case study: POLARIX
2. Impact
3. How about regulators?
4. Resources

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

ICH E9 draft addendum

ICH E9: "Statistical principles for Clinical Trials."

1998.

Why amend E9?

Lack of alignment between trial objectives and reported effect quantification.

Example: Dapagliflozin

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

Dapagliflozin:

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

Trial objective: Assess whether drug works compared to placebo.

Example: Dapagliflozin

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

What is going on?

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

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

What is a “treatment effect”?

Treatment effect

Not defined in original E9!

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

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

Potential outcome $\Rightarrow$ causal inference!

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

Understanding treatment effects

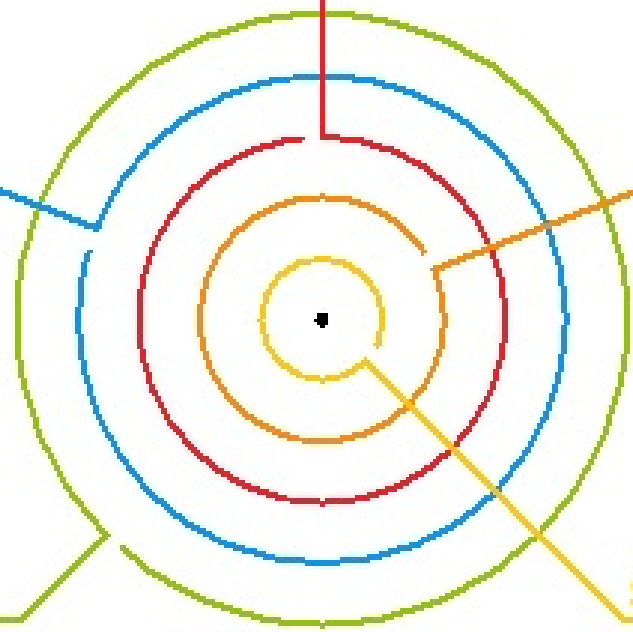
- Multiple definitions of **treatment effect**.
- Different definitions addressing **different scientific questions**.
- Not all equally acceptable for **regulatory decision making**.
- Not all alternatives can be reliably estimated! **Iterative** process of estimand - estimator definition.
- Stakeholders: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

How does the addendum fix this?

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

ESTIMAND

TARGET OF ESTIMATION



VARIABLE

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

POPULATION

The population of patients targeted by the clinical question

INTERCURRENT EVENTS

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

TREATMENT

The treatment condition of interest

SUMMARY

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

Objective pre- and post-addendum

Pre:

Treatment difference between Gazyva and Rituximab on PFS.

Post:

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

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

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

Objective pre- and post-addendum

Pre:

Treatment difference between Gazyva and Rituximab on PFS.

Post:

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

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

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

Objective pre- and post-addendum

Pre:

Treatment difference between Gazyva and Rituximab on PFS.

Post:

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

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

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

Objective pre- and post-addendum

Pre:

Treatment difference between Gazyva and Rituximab on PFS.

Post:

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

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

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

Objective pre- and post-addendum

Pre:

Treatment difference between Gazyva and Rituximab on PFS.

Post:

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

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

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

Objective pre- and post-addendum

Pre:

Treatment difference between Gazyva and Rituximab on PFS.

Post:

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

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

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

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

Agenda

1. Case study: POLARIX
2. Impact
3. How about regulators?
4. Resources

Estimands

ICH E9 addendum

ICH E9: "Statistical principles for Clinical Trials."

1998.

Why amend E9?

Trial objectives:

- Not sufficiently precise described.
- **Lack of alignment** to analysis methods and reported effect quantification.

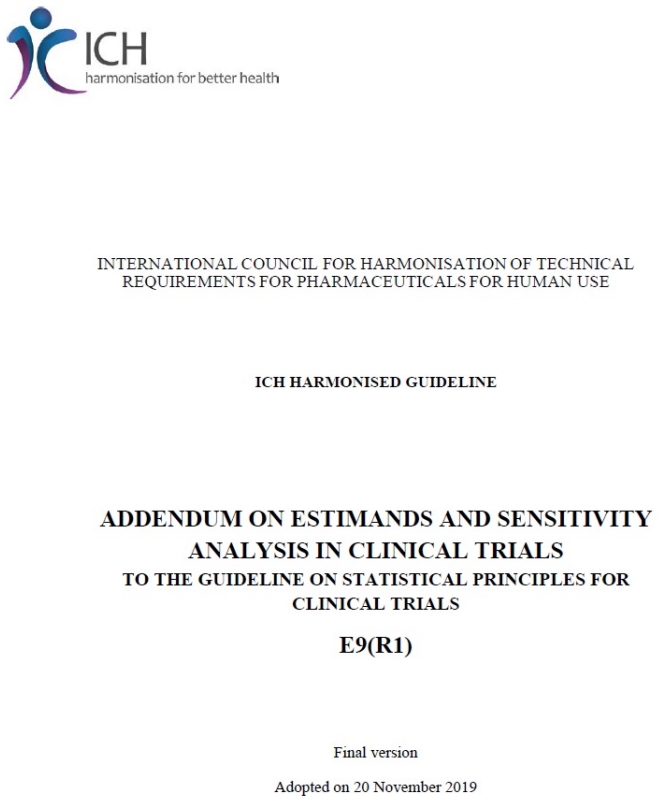
What is an estimand?

- Defines **what will be estimated** for a particular trial objective.
- Precise **description of treatment effect** reflecting clinical question posed by trial objective.
- Summarizes at **population-level** what outcomes would be in same patients under different treatment conditions being compared.
- Five components: "attributes".

Estimand framework for drug development

The principles outlined in ICH E9(R1) are relevant **whenever a treatment effect is estimated, or a hypothesis related to a treatment effect is tested, whether related to efficacy or safety**.

Applies to all trials **regardless of development phase**!



ICH harmonisation for better health

INTERNATIONAL COUNCIL FOR HARMONISATION OF TECHNICAL REQUIREMENTS FOR PHARMACEUTICALS FOR HUMAN USE

ICH HARMONISED GUIDELINE

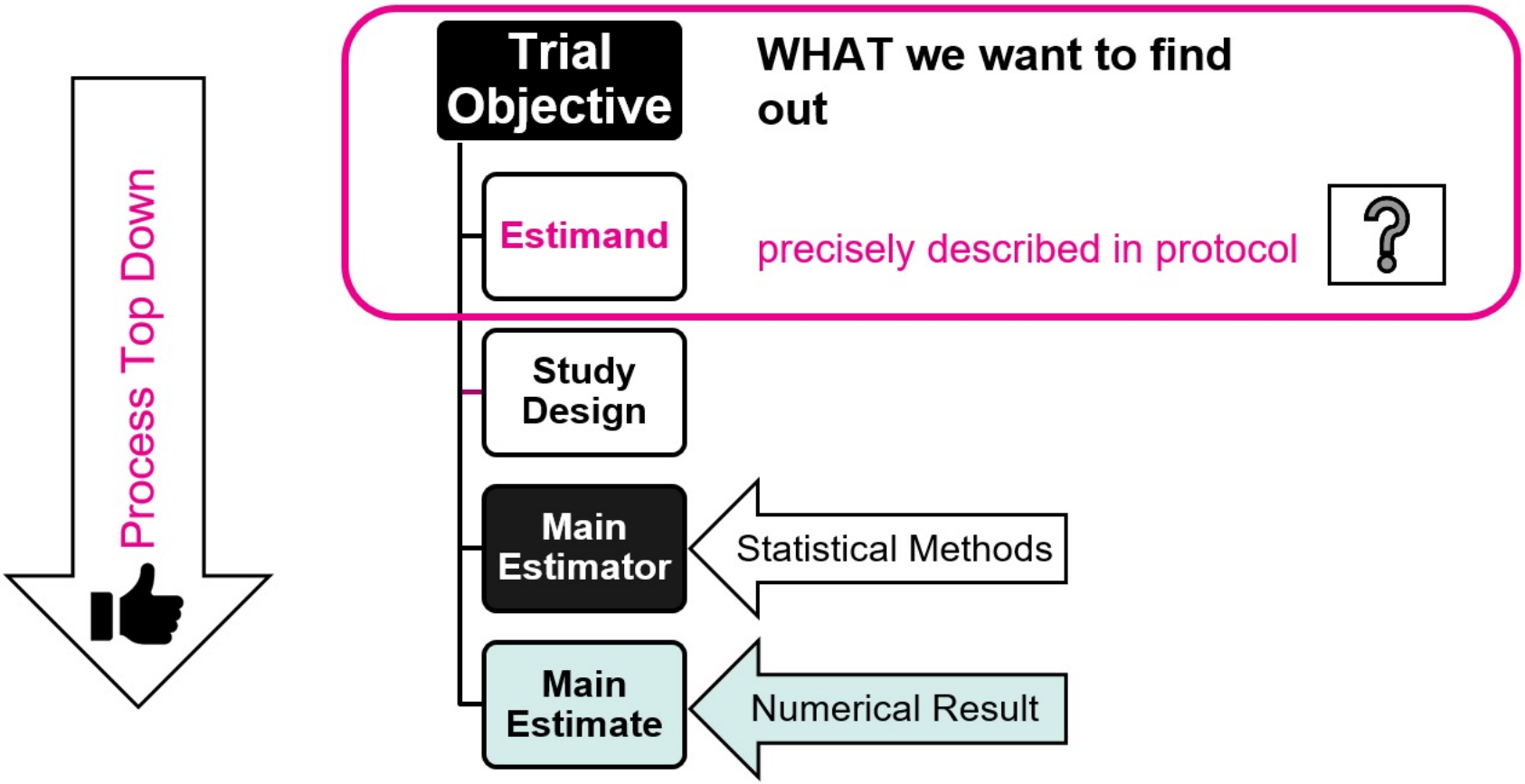
ADDENDUM ON ESTIMANDS AND SENSITIVITY ANALYSIS IN CLINICAL TRIALS
TO THE GUIDELINE ON STATISTICAL PRINCIPLES FOR CLINICAL TRIALS

E9(R1)

Final version

Adopted on 20 November 2019

Estimand framework

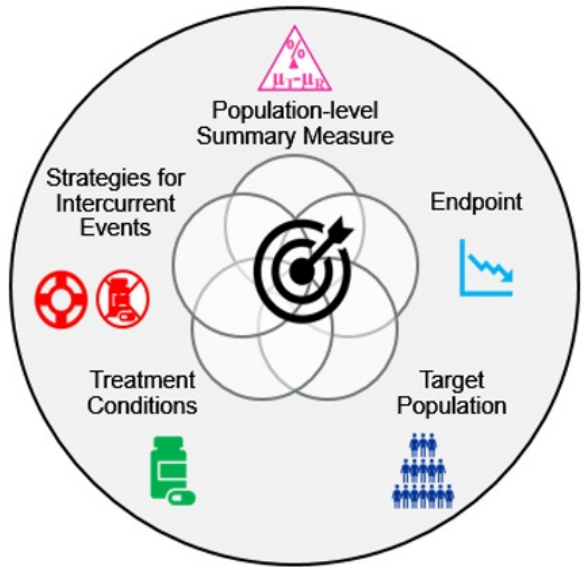


With thanks to the Estimands Implementation Working Group. Slide from the PIONEERing estimands in Clinical Development webinar

Estimand attributes

An Estimand has 5 attributes:

- Population
- Endpoint
- Population-level Summary Measure
- Treatment Conditions
- Strategies for Intercurrent Events



With thanks to the Estimands Implementation Working Group. Slide from the PIONEERing estimands in Clinical Development webinar

Intercurrent events

Intercurrent Events (ICEs)

Events occurring after treatment initiation that affect either the interpretation or the existence of the measurements associated with the clinical question of interest

Specific to scientific objective.

Defined in protocol and are specific to the trial.

Treatment Discontinuation due to Safety

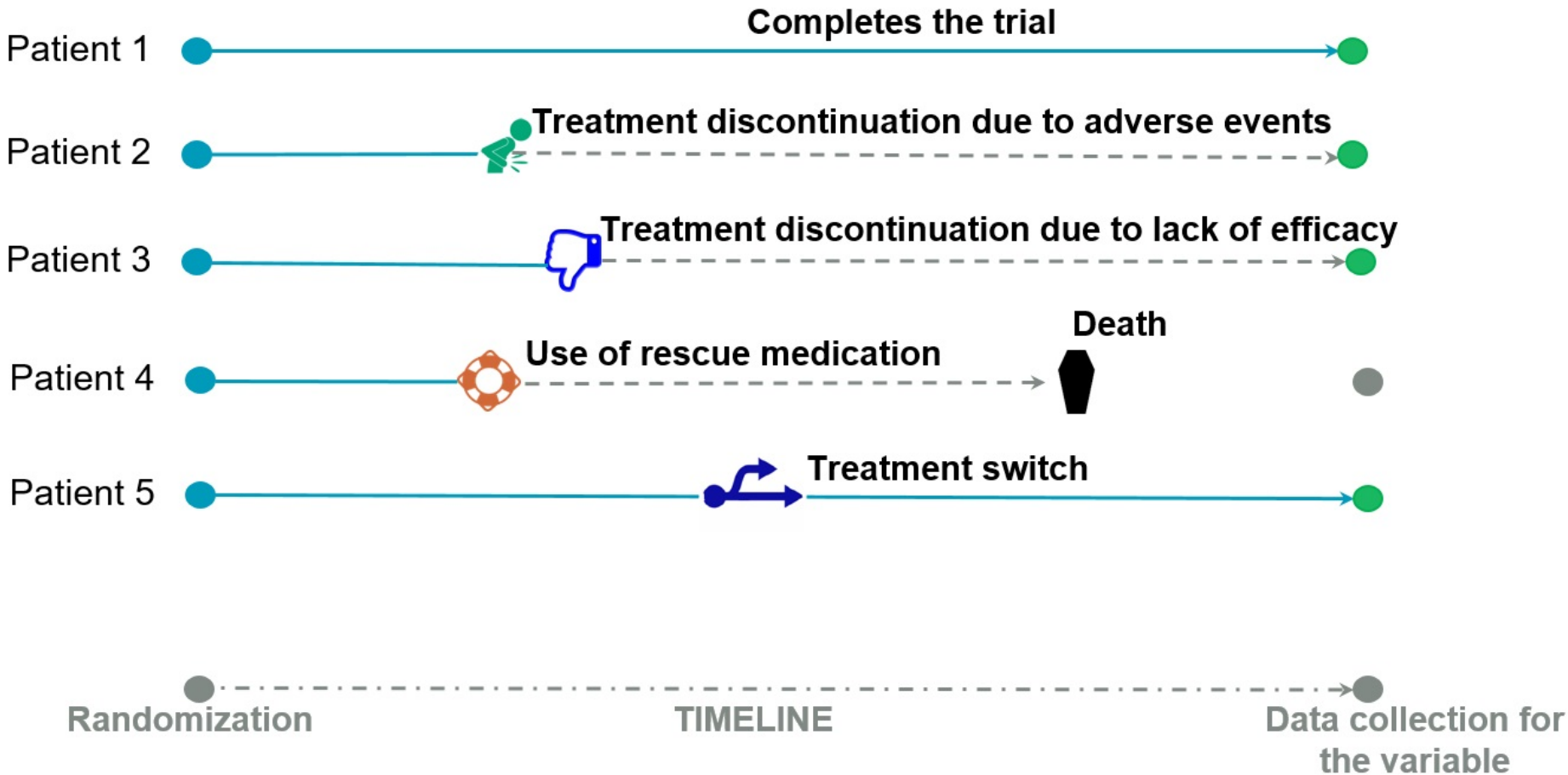
Rescue Therapy

Treatment Discontinuation due to Lack of Efficacy

Death

Treatment Switch

Intercurrent events



Patient journeys

Rescue medication **may impact treatment effect** $\Rightarrow$ intercurrent Event.

Intercurrent events: strategies

Irrespective of	Include in Outcome	Scenario in which event does not occur	Prior to occurrence	As part of target population definition
• Outcome after ICE still of interest • Data collection after ICE	• ICE informative for effect of interest • Variable includes ICE	• Scenario envisaged in which ICE would not occur	• Scientific objective: about what happened prior to ICE • Outcome after ICE considered irrelevant	• Population defined by those in whom ICE would or would not occur
Treatment Policy	**Composite**	**Hypothetical**	**While on Treatment**	**Principal Stratum**

Adapted from Oncology Estimands Working Group. Estimands in Oncology - How and Why webinar

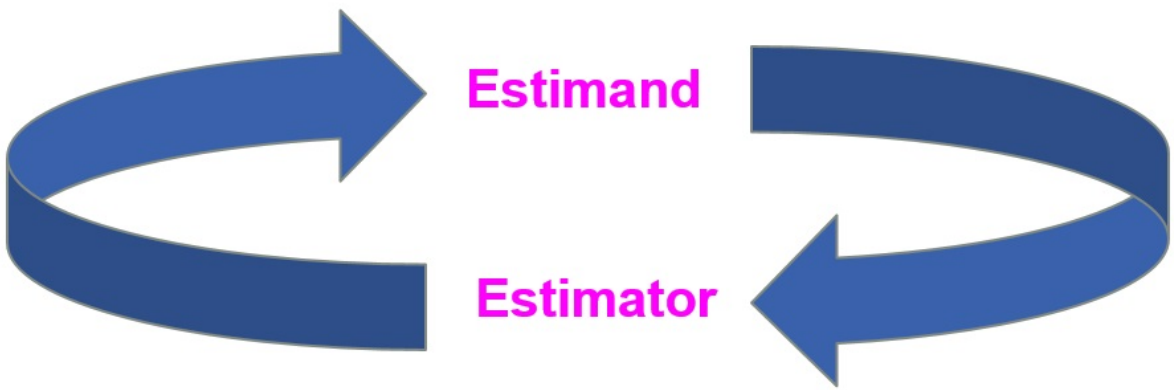
Estimation

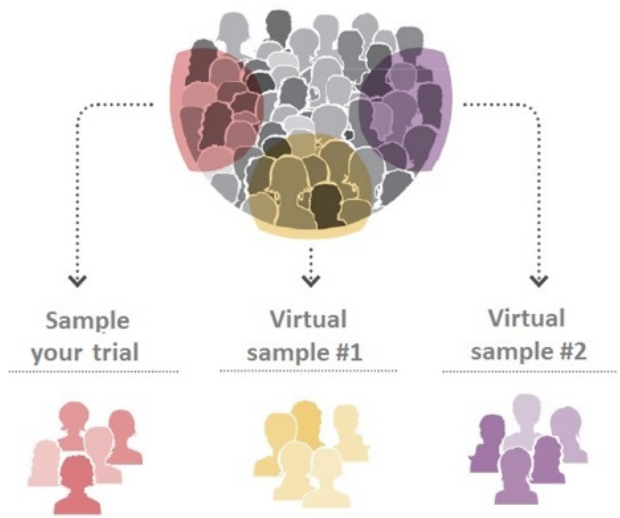
Estimand "lives" on **population level**:

- Defined based on scientific objective.
- Informs **strategy for data collection**: e.g. treatment policy vs hypothetical.

Estimator "lives" on **data level**:

- Computed based on sample.
- For some estimands no good estimator might exist.





Intention-To-Treat (ITT) principle

Effect of **treatment policy**:

- Best assessed using planned rather than actual treatment given.
- Subjects **followed, assessed and analyzed** irrespective of completion of planned course of treatment.

Estimation:

- Analysis based on all subjects.
- Subjects included in analysis as randomized.

What ITT does not tell us!

How to analyze data from subjects who **did not complete treatment as planned**?

Affects two aspects:

- What is **treatment effect of interest**?
- How to handle **missing data**?

How do estimands relate to "ITT"?

Original ICH E9:

> *The intention-to-treat principle implies that the primary analysis should include* **all randomised** *subjects. Compliance with this principle would necessitate* **complete follow-up** *of all randomised subjects for study outcomes.*

Critique:

- Estimand? Estimation? Unclear!
- What if we do not have complete follow-up?

Estimand addendum fills these gaps.

Refines and extends ITT.

Is treatment policy strategy simply translation of ITT?

No!

Need precise definition of "policy"!

Fletcher et al. (2022)

In short, "treatment policy" is not interchangeable with "ITT", and clinical researchers need to recognize that many historical "ITT analyses" do not align with analyses based on treatment policy strategies.

Fletcher et al. (2022)

Recommendations for future use of "ITT"

Recommendations for future use of ITT in light of addendum:

- Do not use "ITT" when describing "population" attribute of estimand.
- Avoid use of "ITT" when describing analysis sets.
- Be specific and **precisely define analysis sets**. Proposals for abbreviations are: "full analysis set", "all randomized patients", "efficacy-evaluable patients".

Per-protocol populations in drug development

Population of all patients in trial who

- do not violate major inclusion/exclusion criteria,
- are treated according to randomized treatment arm,
- are not lost to follow-up.

Idea: Assess **biological effect**.

Regulators generally not interested in per-protocol (superiority trials).

Per-protocol in light of estimands

Addendum says (PPS = per protocol set):

> *In respect of the framework presented in this addendum, it* **may not be possible to construct a relevant estimand to which analysis of the PPS is aligned**. *... Estimands might be constructed, with aligned method of analysis, that better address the objective usually associated with the analysis of the PPS. If so, analysis of the PPS might not add additional insights.*

Spirit of addendum:

- Clearly state scientific objective and derive estimand from that.
- PPS starts at wrong end (just because we used PP in the past).
- Not clear what scientific objective it corresponds to.

How about noninferiority trials?

Traditionally, ITT and PPS **somewhat reversed** in equivalence / non-inferiority trials.

Estimand **independent of the hypothesis test** you intend to perform!

Akacha *et al.* (2017):

> *...it has become common practice to perform ITT analyses for superiority trials and per protocol analyses for non-inferiority trials. This choice seems to have* **little to do with potential differences in the underlying estimands of interest**. *The main reason for analyzing superiority and non-inferiority trials differently appears to be driven by the wish to construct conservative treatment effects.*

Conservativeness in **estimation**: do not deal with it by fiddling around with estimand!

Recommendations:

- Do not specify a PPS. **FDA feedback** on noninferiority trial: "What should we do with PPS?".
- Embed scientific question in estimand framework by clearly defining intercurrent events and describing which strategy to apply for them.

Analysis sets

Ideal world: all trial patients

- fulfill in- and exclusion criteria,
- adhere to all protocol-specified treatments and assessments,
- are followed-up as in protocol.

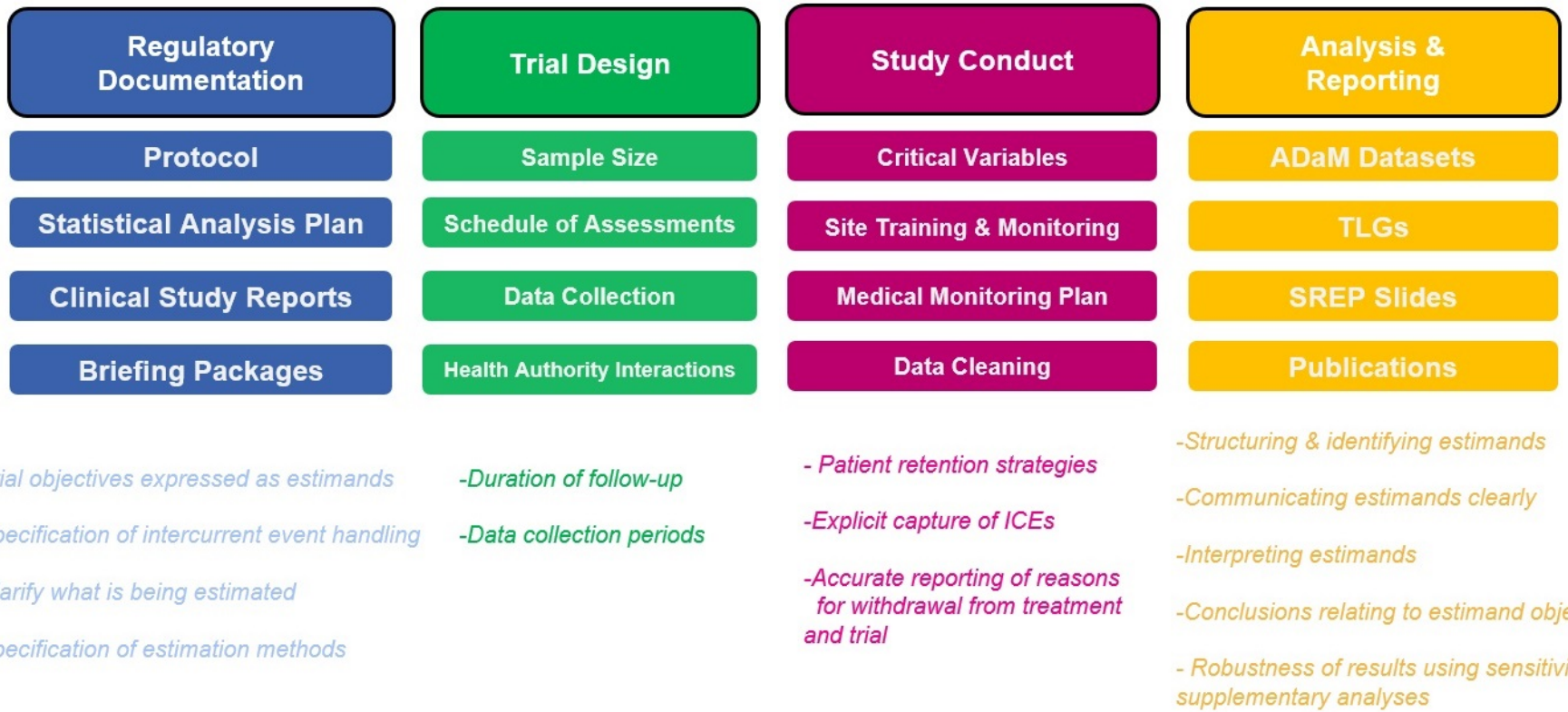
$\Rightarrow$ straightforward analysis, analysis sets are clear.

Rarely the case in **clinical reality**.

Very precisely define **analysis sets**:

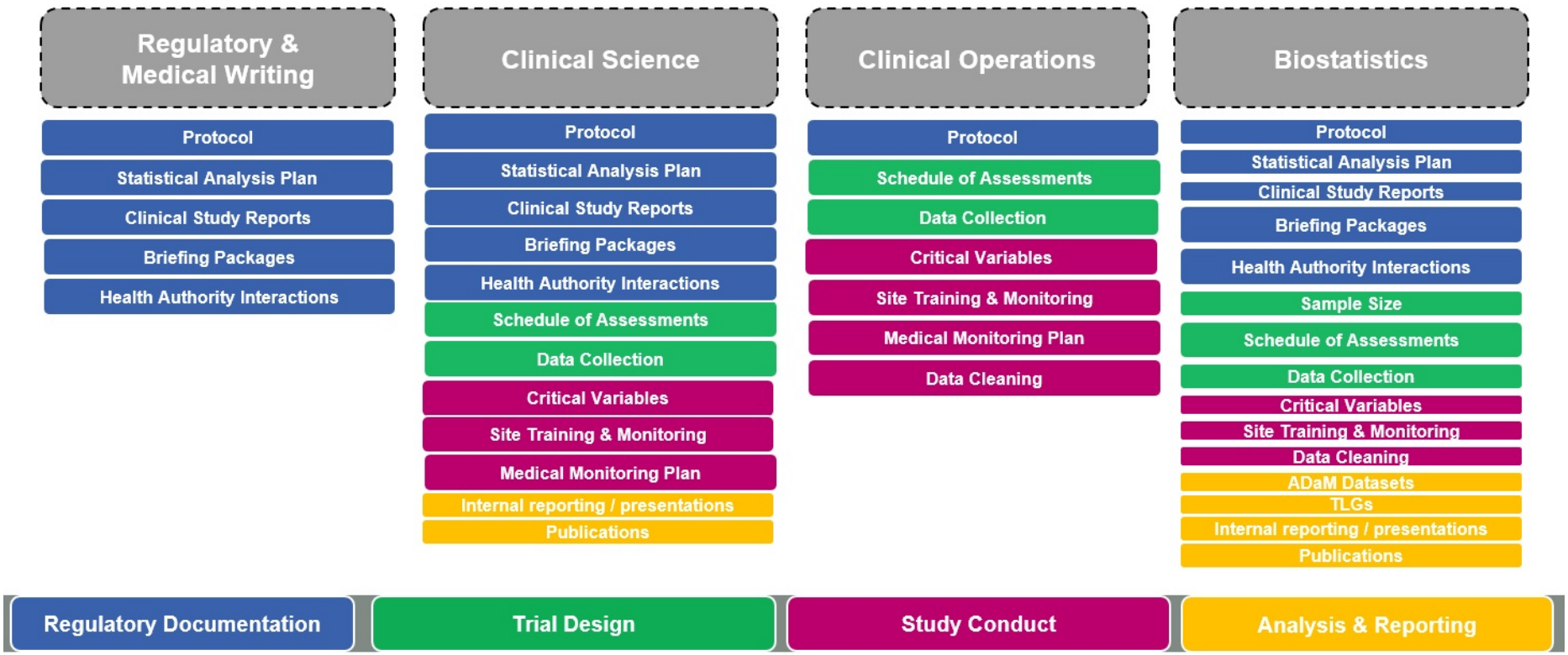
- **Which patients**?
- **Which observations**?

Impact of estimands on clinical trials



Definition of the estimand affects **virtually all aspects of clinical trial**!

Impact of estimands on stakeholders



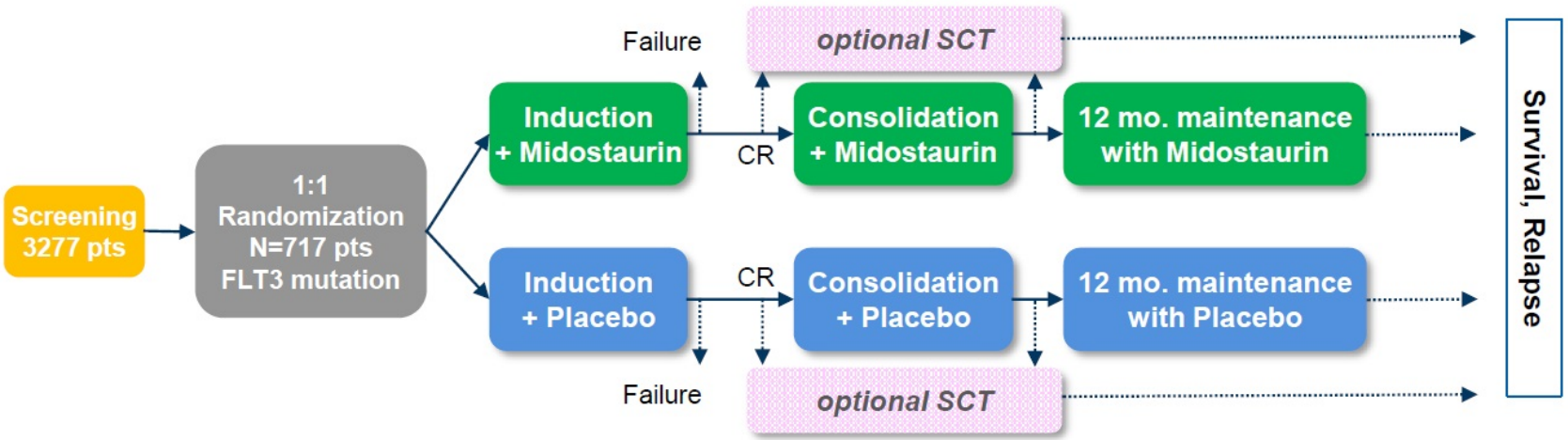
Agenda

1. Case study: POLARIX
2. Impact
3. How about regulators?
4. Resources

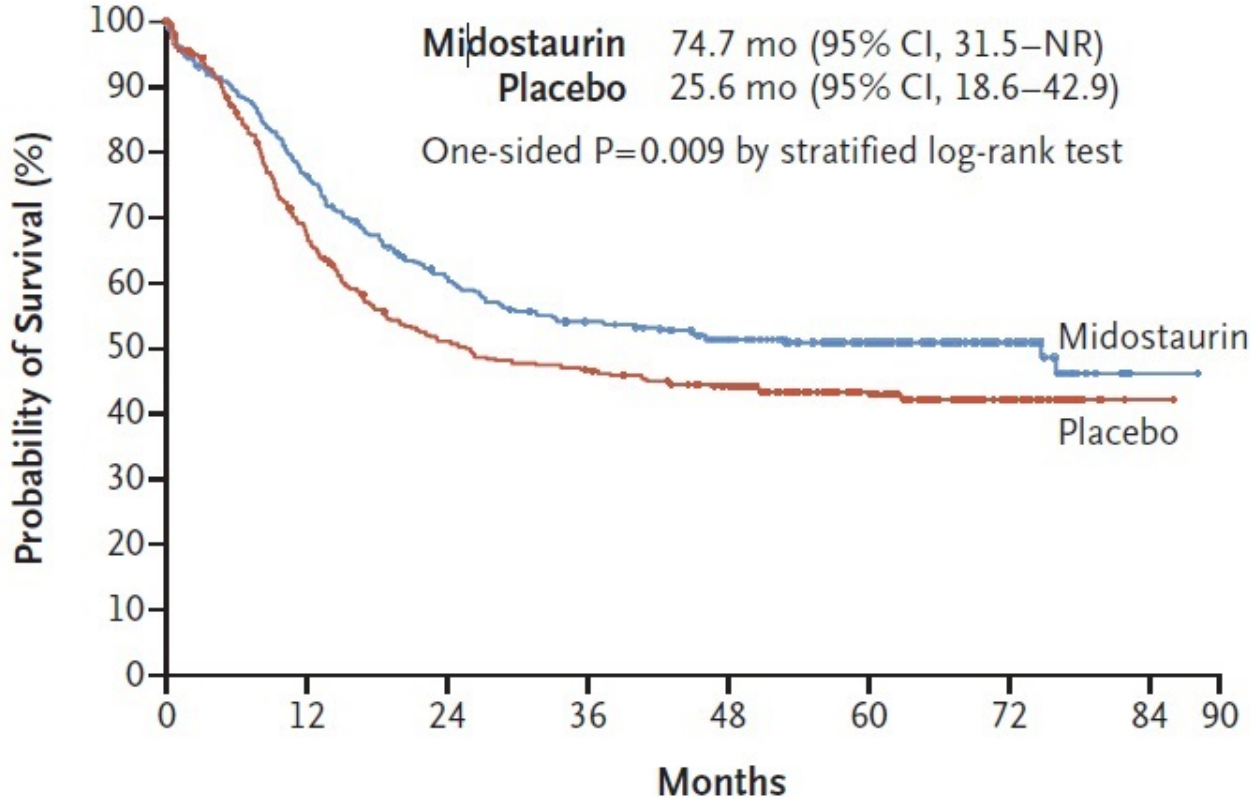
Case study: RATIFY

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?

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

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

Completely different clinical questions!

What question are we asking?

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

What ended up in the label?

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

AML: treatment strategy based on sequence of

- multiple decision points and
- treatment modalities.

RATIFY:

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

How would we define the estimand today?

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

Treatment strategy:

- Experimental: 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:

Non-proportional hazards?

Cure?

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

Backup: Hypothetical strategy to address ICEs: application to Covid-19

Hypothetical estimands

- ICH E9(R1) addendum: acknowledges that some hypothetical scenarios likely of more clinical or regulatory interest than others.
  - CAR-T example: hypothetical estimand less relevant.
- Hypothetical estimands: often **implicitly** targeted by primary analysis in pivotal trials:
  - FDA guideline: Censor PFS at initiation of new anticancer therapy.
  - Routine use of MMRM when "missing data" is present.
- More explicitly: EMA guidelines for Alzheimer and Diabetes.

COVID-19 and estimands

Primary intention of ICH E9 addendum: **alignment** between clinical trial objectives and treatment effect estimation **prior** to start of trial.

ICH E9 addendum specific for **unforeseen** clinical events during trial conduct:

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

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

Assessing impact of COVID-19 on estimand

VARIABLE

The variable (or endpoint) to be obtained for each patient.
Q: Does the current endpoint reflect the treatment effect in the original scientific objective?

POPULATION

The population of patients targeted by the clinical question.
Q: Are the enrolled patients representative of the target population?

INTERCURRENT EVENTS (ICEs)

Other ICEs not already addressed by treatment, population and variable, and how they are handled.
Q: Can the original clinical trial objective be addressed without defining new strategies for ICEs related to COVID-19? (e.g. apply pre-specified rules for discontinuations to discontinuations due to COVID-19)

TREATMENT

The treatment condition of interest.
Q: Are the treatment conditions (e.g. non-compliance, drug discontinuation, subsequent therapy) representative of what would have been administered pre-COVID-19?

SUMMARY

A population-level summary for the variable which provides a basis for treatment comparison.
Q: Is the summary measure still interpretable?

Backup: Hypothetical strategy to address ICEs: application to Covid-19

COVID-19 and hypothetical estimand

Ongoing trials: implicitly designed assuming

- **no major disruption of healthcare systems** and
- **absence of highly infectious disease** with severe complications
- for which **no effective therapy** is available.

Trial objectives should relate to world without COVID-19 pandemic.

Intercurrent events primarily caused by disruption of healthcare system or patients' desire to minimize traveling independently of disease or treatment: **hypothetical strategy reasonable**.

Implication on estimation

Change in estimand does not always requires change in analysis.

Estimates from initially planned analysis: may still be sufficiently precise to assess effect in a world without COVID-19 pandemic.

Focus on questions of interest:

- Results in more clarity in interpretation.
- regardless of whether there is a change in analysis.

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

Agenda

1. Case study: POLARIX
2. Impact
3. How about regulators?
4. Resources

Backup: Case study: treatment switching

Good old days: Herceptin

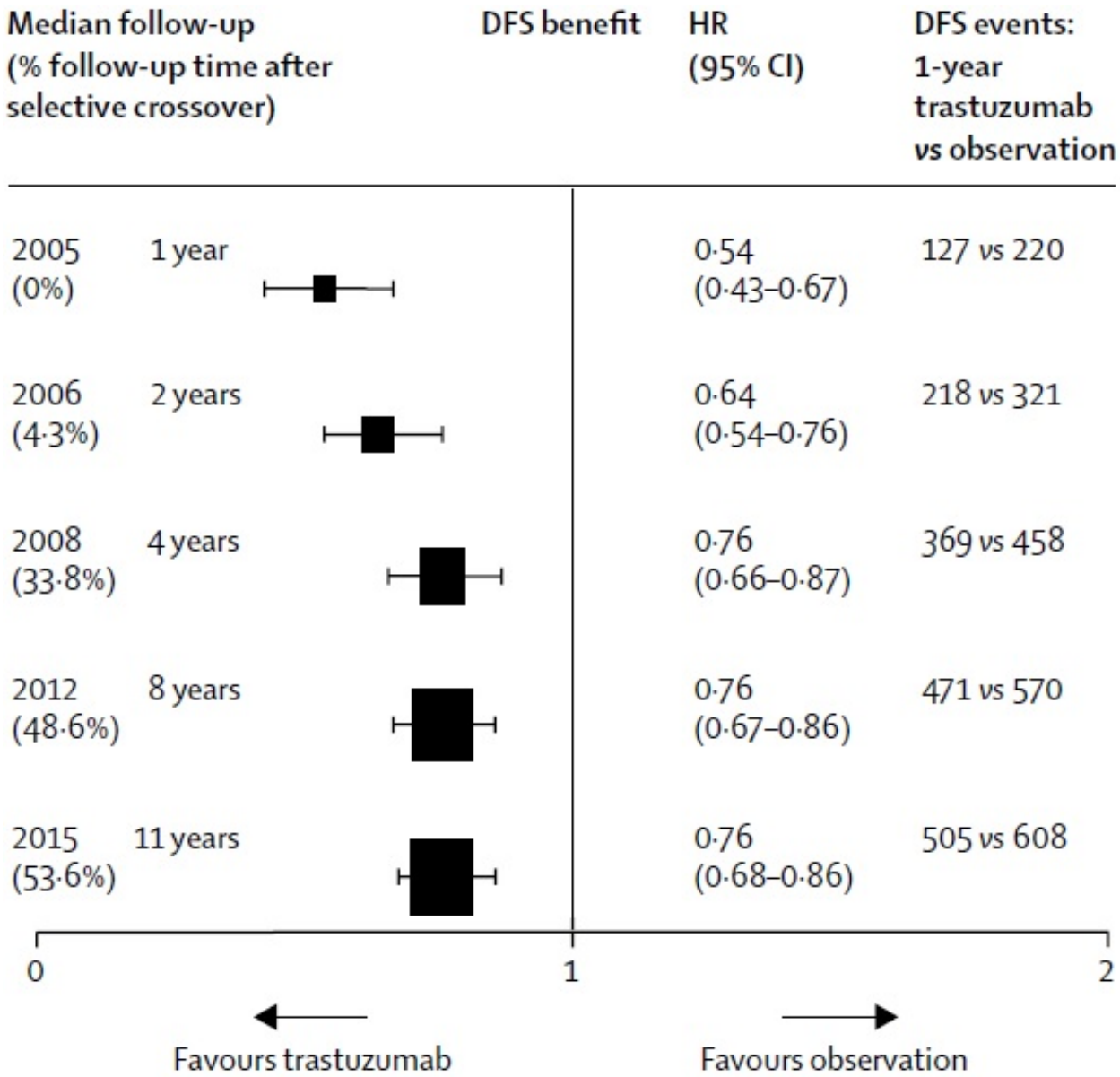
HERA

- **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**.

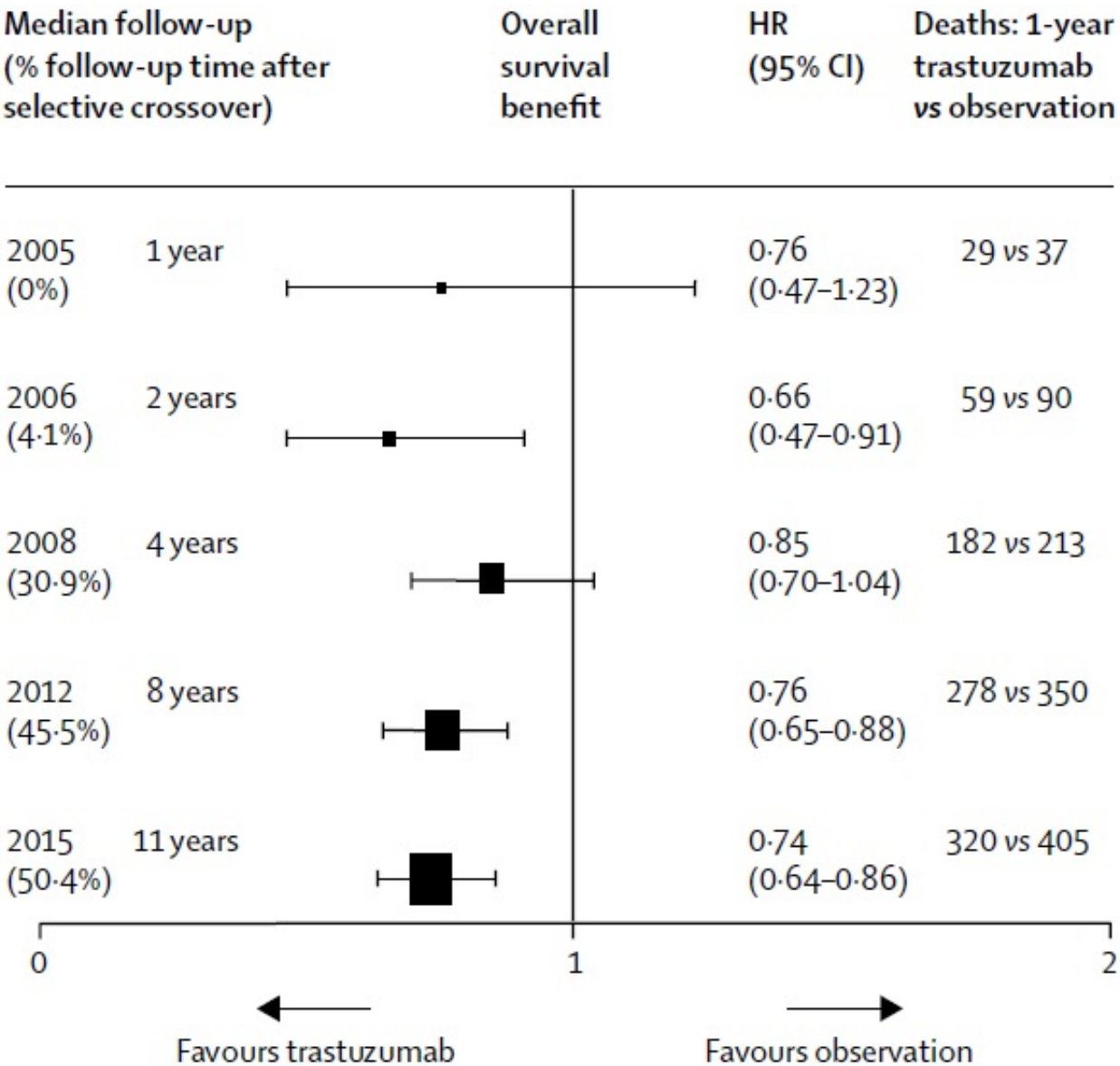
Piccart-Gebhart and Procter (2005):

- 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

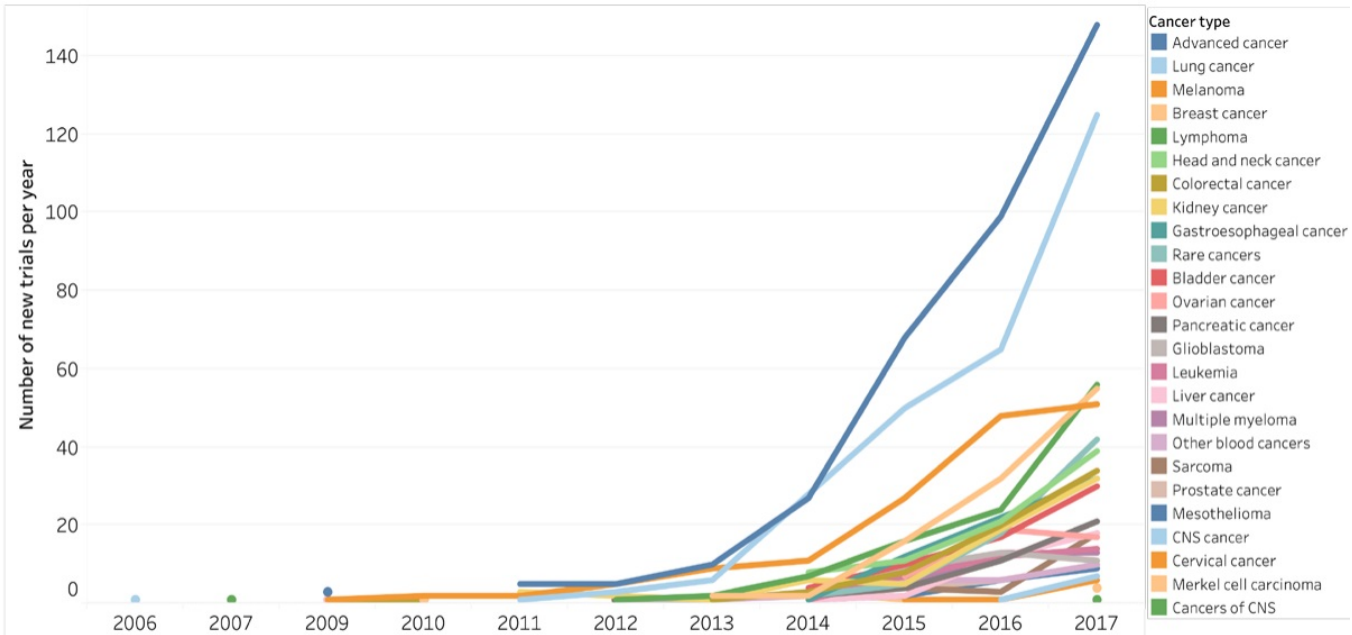
OS effect establised in long-term follow-up **despite cross-over**:

- 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

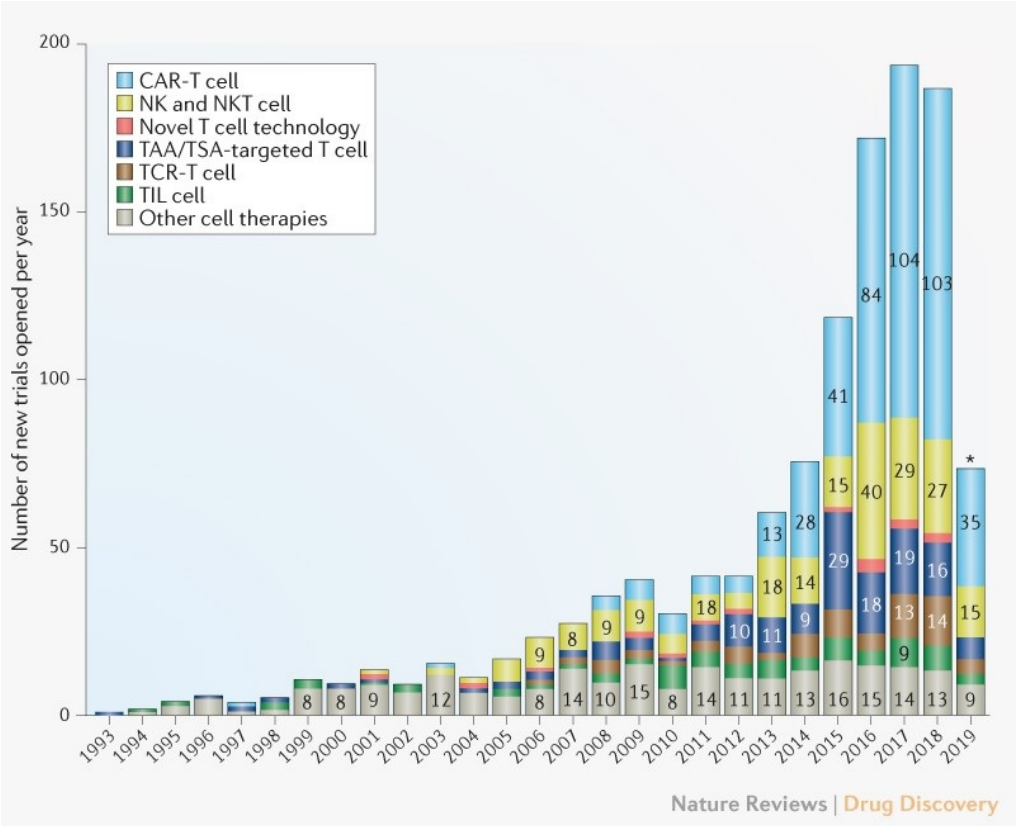


1 in 2006, **1502** in Sep 2017, **2250** in Sep 2018, **2975** in Sep 2019.

Tang *et al.* (2018)

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

CAR-T trials



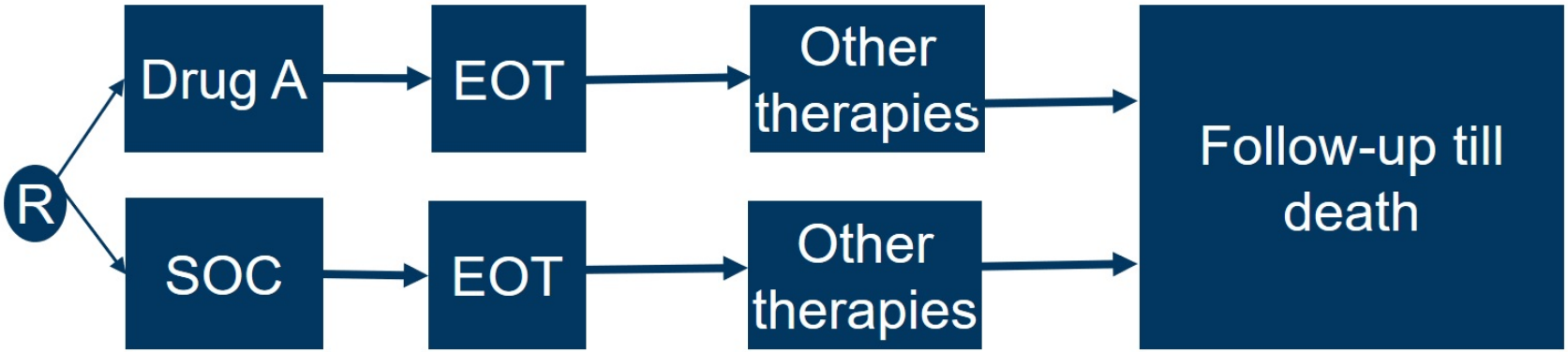
13 in 2013, >**100** in 2017.

Yu *et al.* (2018).

Great for patients!

- durable responses,
- many ongoing clinical trials.

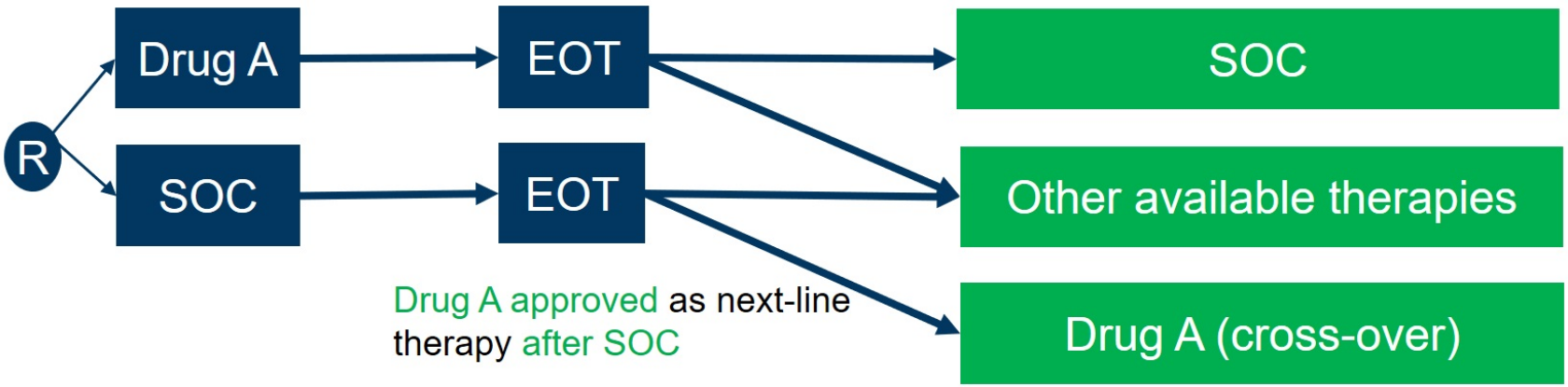
But what does it mean for clinical trials?



Typical OS definition:

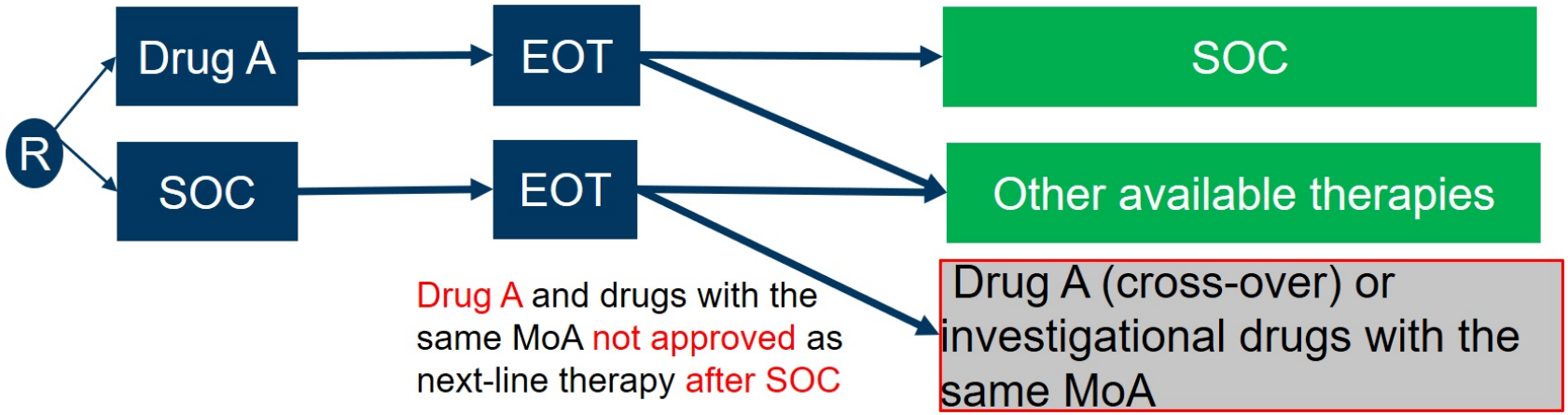
- Time from randomization to death **regardless of patient's journey**.
- **Treatment policy** for every intercurrent event (crossover, new therapy, etc.).
- Balance in subsequent therapies generally not expected:
  - Physician choose subsequent therapy in light of previously administered therapies.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

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



Subsequent therapy after EOT reflects clinical practice.

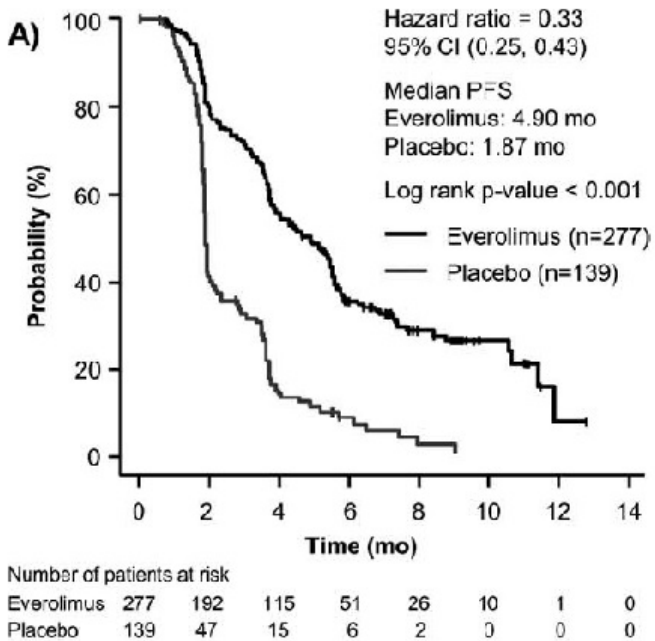
Treatment policy OS estimand **interpretable**.

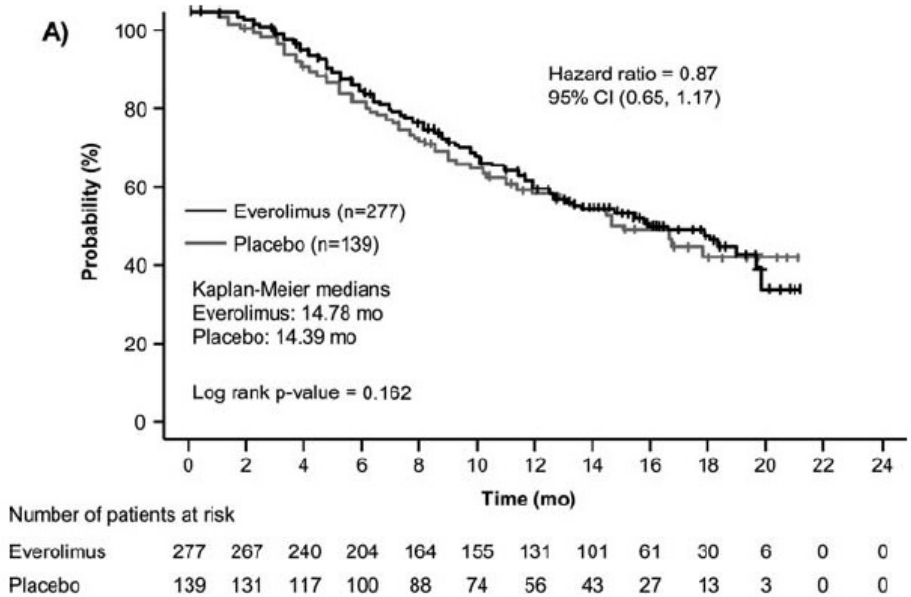


Subsequent therapy after EOT **does not** reflect clinical practice:

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

RECORD-1





RECORD-1: Motzer *et al.* (2010).

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).

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:

Regorafenib USPI:

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

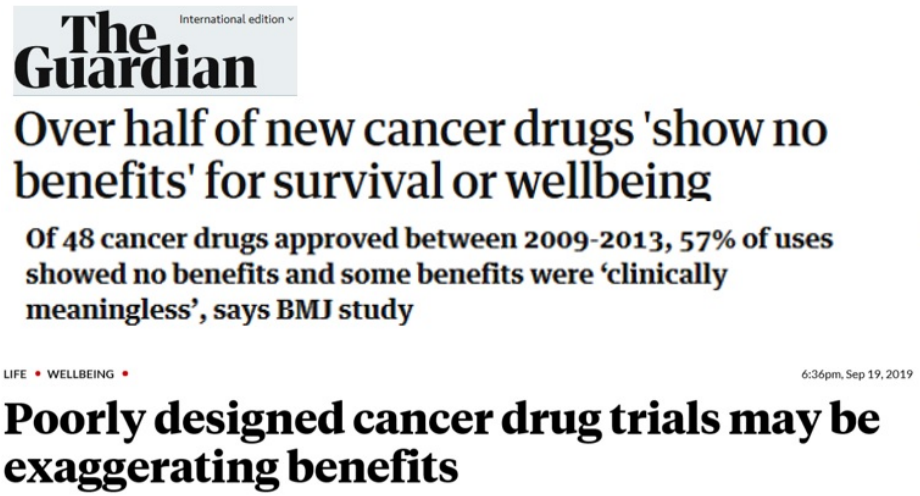
There was no statistically significant difference in overall survival at the final OS analysis, conducted at 162 OS events (Table 8). Cross-over to open label STIVARGA occurred in 58 (88%) placebo-treated patients after disease progression.

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.



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.

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.

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:

- 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?

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.* (2022)

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.

R version and packages used to generate these slides:

R version: R version 4.3.2 (2023-10-31 ucrt)

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

Other packages:

This document was generated on 2025-01-23 at 21:48:59.