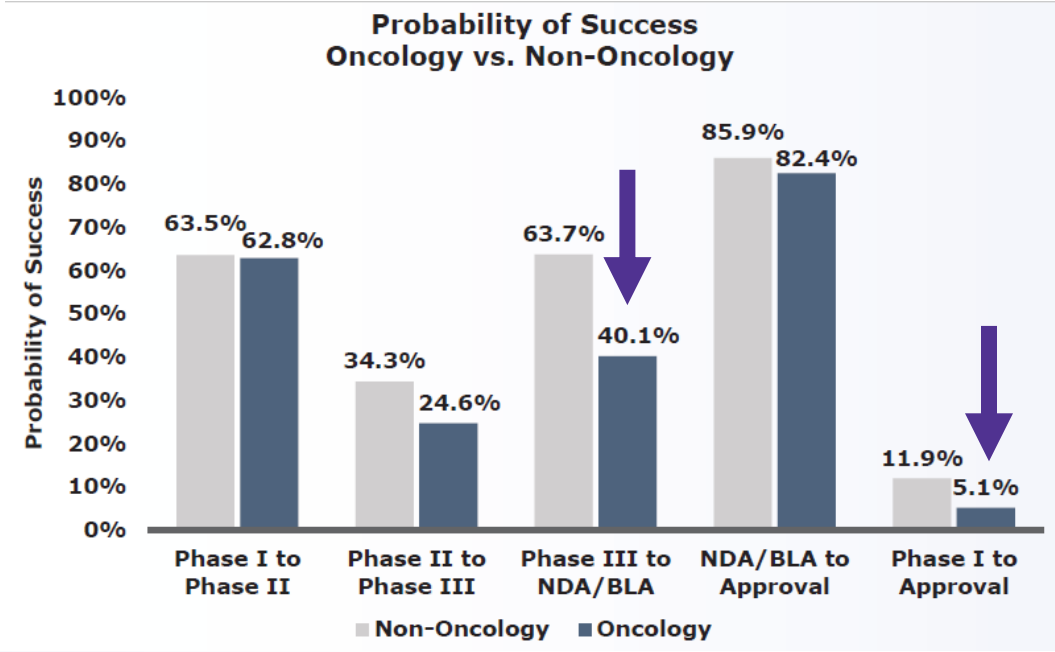
Matching endpoints and objectives in clinical trials

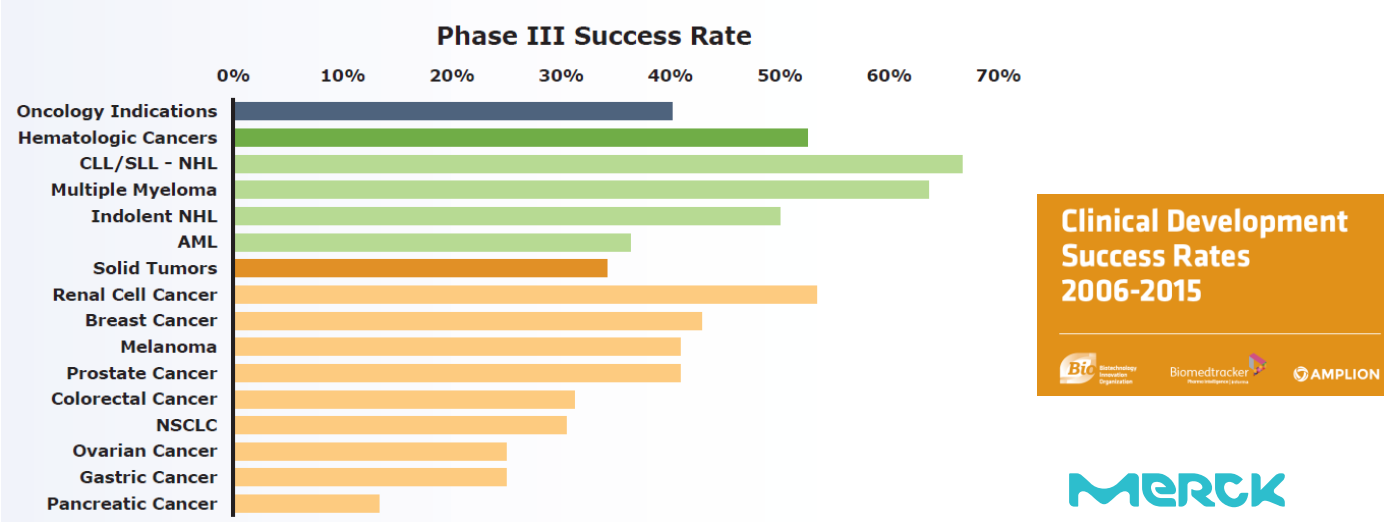
Reflections on disconnects and opportunities

Armin Schüler,
CDDF, 2020-12-01

The Challenge
Trials often fail to show what we are expecting from them to show

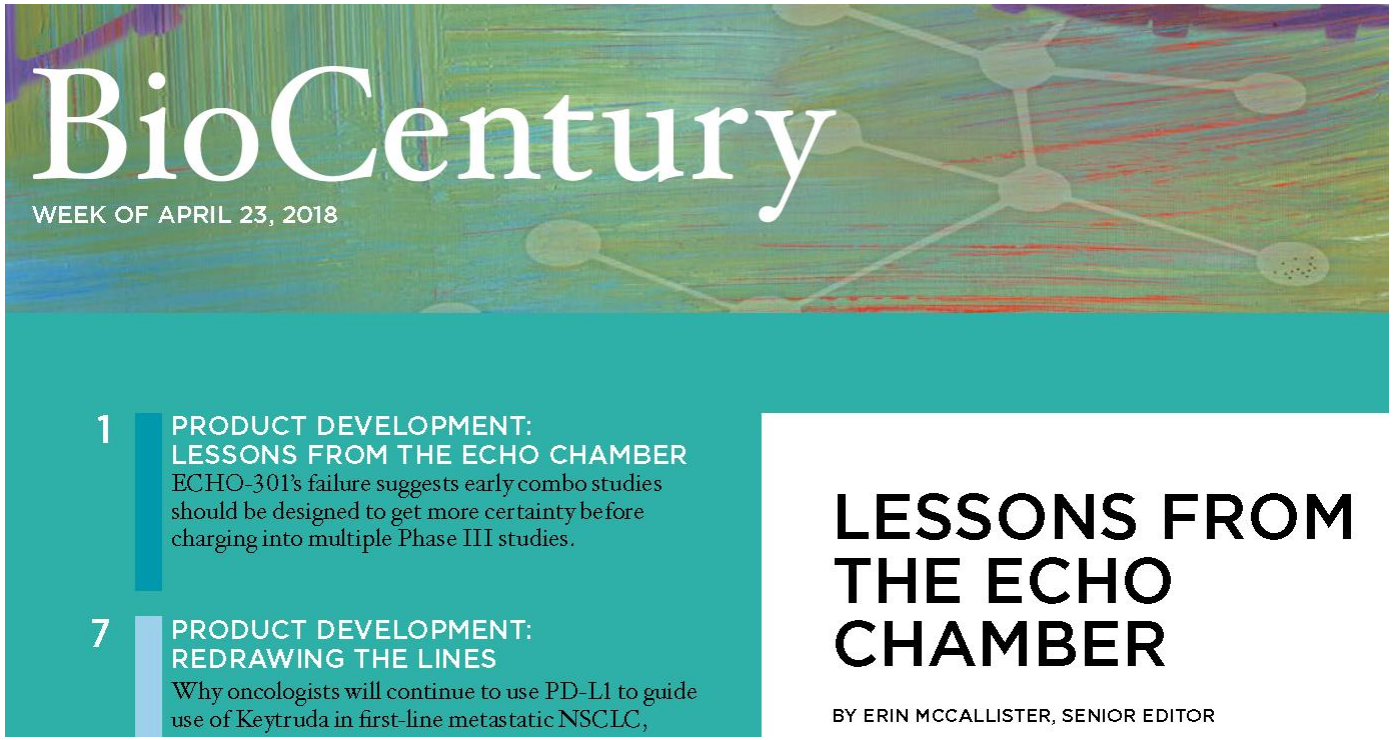
- Most drugs in early development do not have the potential for market approval which is per se not an issue
- It's the failure rates of the phase III trials which is concerning





Clinical Development Success Rates 2006-2015

The Challenge
Trials often fail to show what we are expecting from them to show



A Phase I/II trial of the combo was started in 2014, and last year produced an overall response rate (ORR) that was **75% greater** than the historical rate of Keytruda in the same setting.

Those results didn't bear out in ECHO-301.

The Challenge
Trials often fail to show what we are expecting from them to show

potential reasons

The Challenge
Trials often fail to show what we are expecting from them to show

potential reasons

1 **Impact of study population**

Different in-/exclusion criteria
- Impact of centre selection
- Different enrichment due to change in available therapies

The Challenge
Trials often fail to show what we are expecting from them to show

potential reasons

1 **Impact of study population**

Different in-/exclusion criteria
- Impact of centre selection
- Different enrichment due to change in available therapies

2 **Impact of outcome measure selected for the individual trials across phases**

- Different endpoints (Response vs. OS)
- Impact of independent read

The Challenge
Trials often fail to show what we are expecting from them to show

potential reasons

1 **Impact of study population**

Different in-/exclusion criteria
- Impact of centre selection
- Different enrichment due to change in available therapies

2 **Impact of outcome measure selected for the individual trials across phases**

- Different endpoints (Response vs. OS)
- Impact of independent read

3 **Impact of individual treatments within a combination is unclear**

The Challenge
Trials often fail to show what we are expecting from them to show

potential reasons

1 **Impact of study population**

Different in-/exclusion criteria
- Impact of centre selection
- Different enrichment due to change in available therapies

2 **Impact of outcome measure selected for the individual trials across phases**

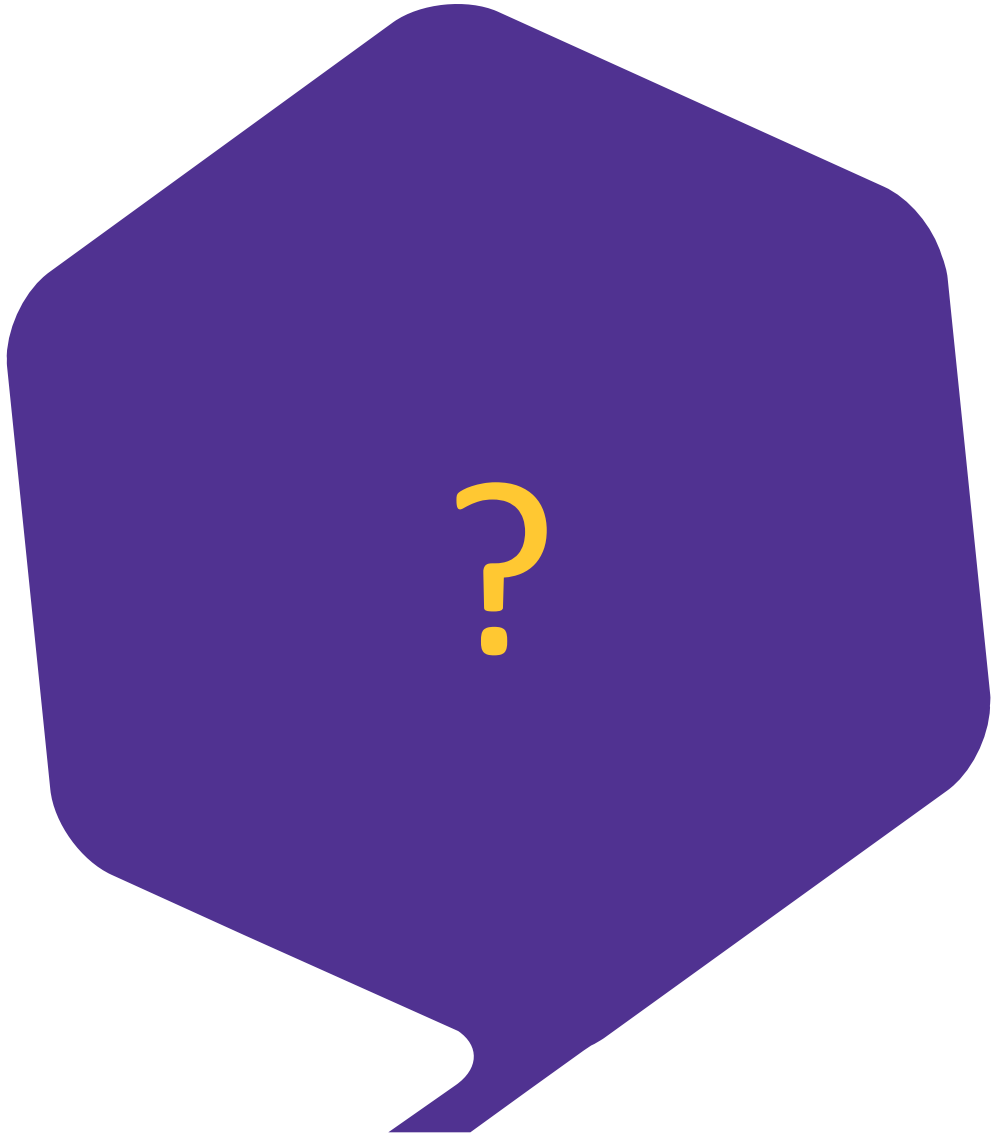
- Different endpoints (Response vs. OS)
- Impact of independent read

3 **Impact of individual treatments within a combination is unclear**

4 **By chance findings**

- Further confirmatory trials usually show lower effects
  Vice versa: confirmatory trial would not be conducted
- Power calculation on believes

5 **…**

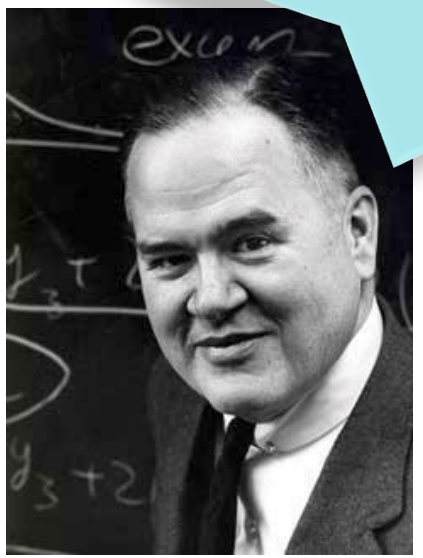


Let's step a little bit

John Tukey (1962)[1]

Far better an *approximate* answer to the *right* question, which is often vague,

than an *exact* answer to the *wrong* question, which can always be made precise.



[1]The future of data analysis. Annals of Mathematical Statistics 33 (1), (1962), p.13

Overall Survival – Same endpoint but 2 different clinical questions

1

To demonstrate improvement in OS with trt A compared to trt B

followed by subsequent therapy

with other words

What is the drug adding to the existing treatment landscape?

Are we Asking the right question?

2

To demonstrate improvement in OS with trt A compared to trt B

if there would no further line treatment existing

with other words

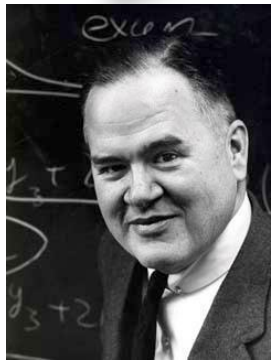
What is the benefit of A compared to B if not diluted by subsequent therapy?

MERCK

It’s about the research question, not the endpoint

John Tukey (1962)
Far better an *approximate* answer to the *right* question, which is often vague,

than an *exact* answer to the *wrong* question, which can always be made precise.



What question are we really interested to answer?

Specification of clear, specific and detailed trial objective

P – Patient, Problem or Population
I – Intervention
C – Comparison, control or comparator
O – Outcome

Estimands

What is an Estimand?

Estimand: A **precise description of the treatment effect** reflecting the clinical question posed by the trial objective. It summarizes at a population-level what the outcomes would be in the same patients under different treatment conditions being compared.

Estimator: A **method of analysis** to compute an estimate of the estimand using clinical trial data.

Estimate: A **numerical value** computed by an estimator.

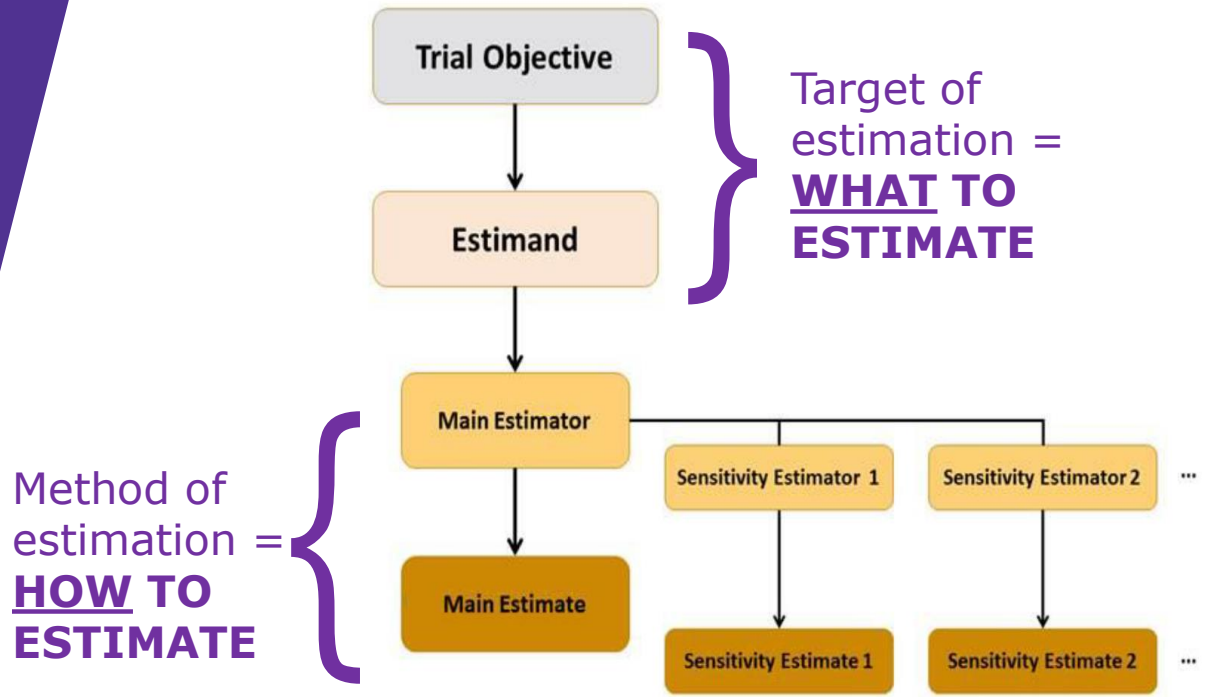
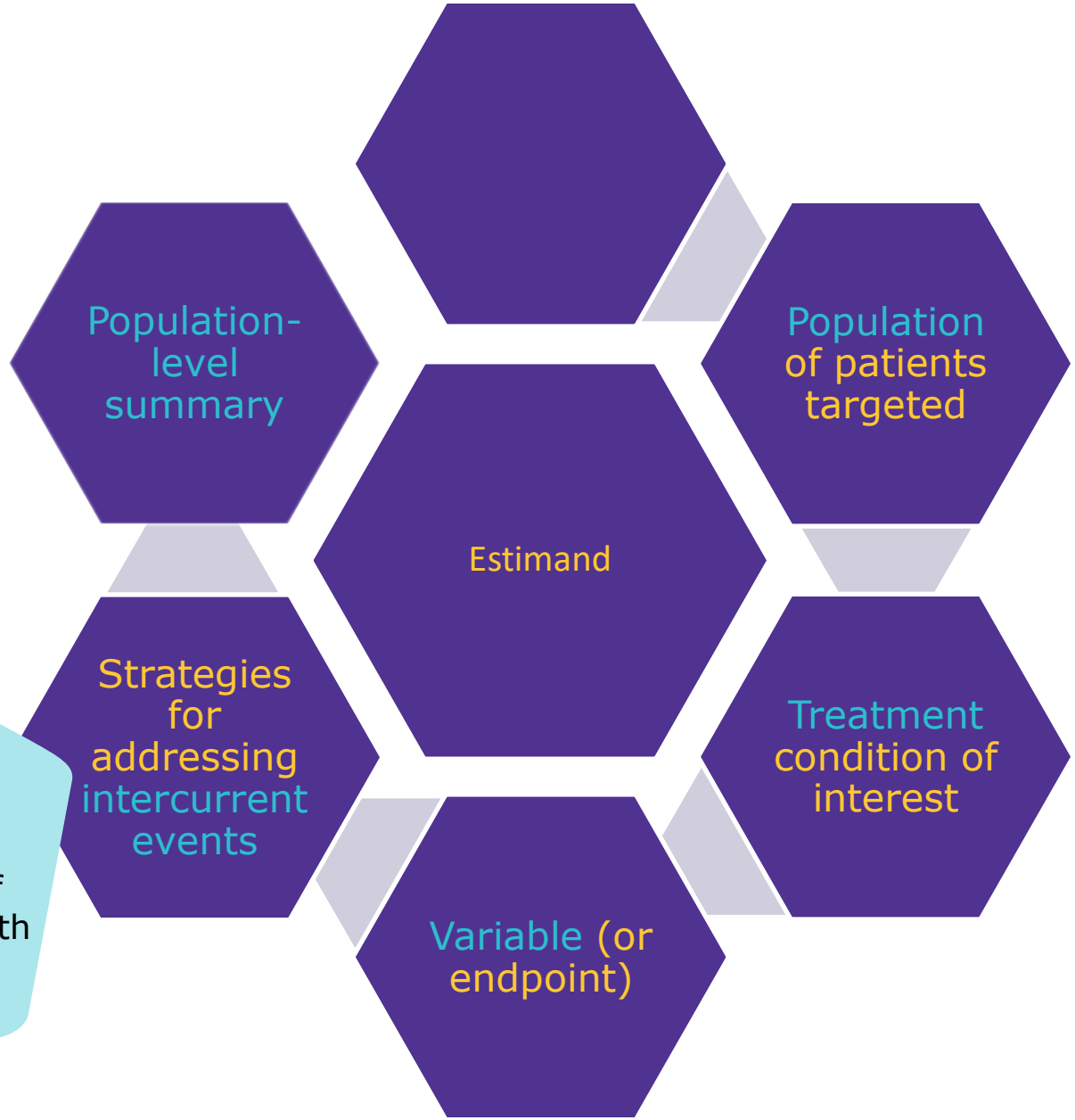


Figure 1: Aligning target of estimation, method of estimation, and sensitivity analysis, for a given trial objective

What is an Estimand?

Estimand is the target of estimation to address the scientific question of interest posed by the study objective.

An estimand is described by five attributes, defining together the treatment effect of interest.



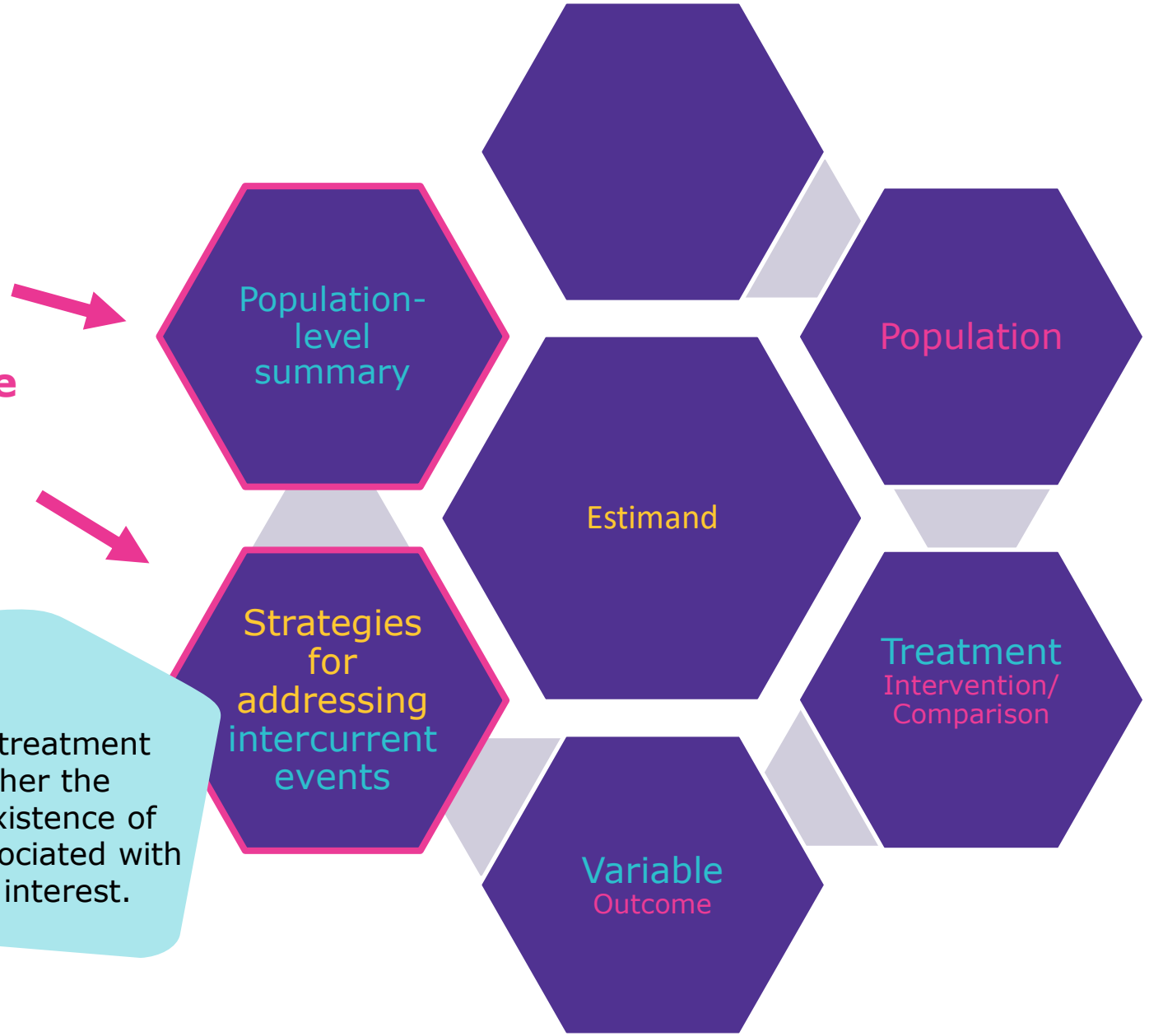
Def.: Intercurrent Events

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

What is an Estimand?

It's building on existing frameworks.

Compared to PICO there are 2 new elements introduced



Def.: Intercurrent Events

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

Intercurrent event in cancer trials

Relevant intercurrent events are

- Treatment discontinuation (e.g. due to an adverse event)
- Subsequent therapy (potentially unknown)

Example

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

- **65%** of patients in the BEV-IRI received TMZ, and
- **66%** of patients in the TMZ arm received BEV (+IRI)

Herrlinger U, et al. Bevacizumab Plus Irinotecan Versus Temozolomide in Newly Diagnosed O 6-Methylguanine-DNA Methyltransferase Nonmethylated Glioblastoma: The Randomized GLARIUS Trial. Journal of Clinical Oncology. 2016; 34 (14): 1611-1619. doi: 10.1200/JCO.2015.63.4691 29.06.2020 14

Intercurrent event in cancer trials

Relevant intercurrent events are

- Treatment discontinuation (e.g. due to an adverse event)
- Subsequent therapy (potentially unknown)

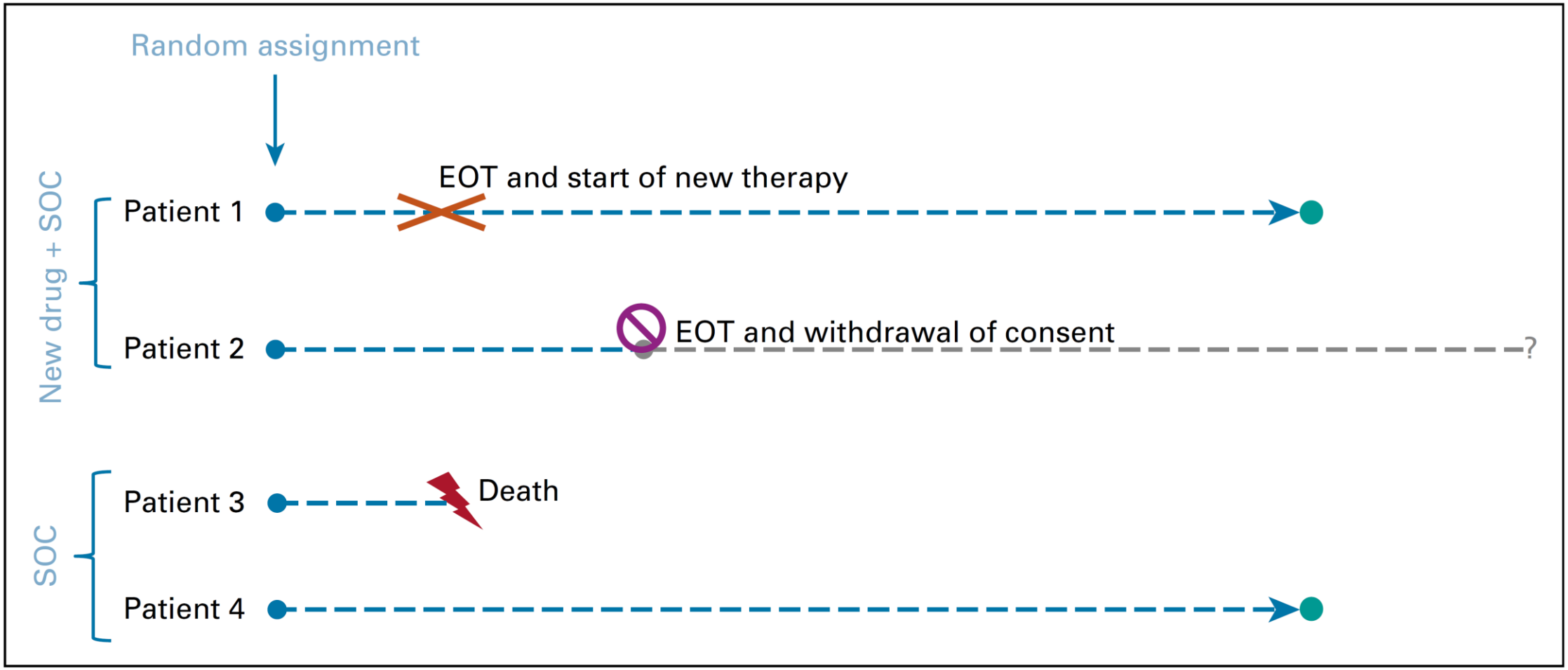
Example

The **JAVELIN Lung 200** trial, a randomised, phase 3 trial in patients with stage IIIB or IV or recurrent NSCLC and disease progression after treatment with a platinum-containing doublet with the primary objective to show improvement in overall survival (OS)

- **26%** of patients in the docetaxel arm, and
- **6%** of patients in the aveluamb arm
  received as subsequent therapy immune checkpoint inhibitors (like nivolumab, pembrolizumab, etc.)

Barlesi F. et al. 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, Volume 30, Issue Supplement_5, October 2019, doi: 10.1093/annonc/mdz260.01429.06.2020 12,

Intercurrent event in cancer trials – an illustration



Evgeny Degtyarev et al. Estimands and the Patient Journey: Addressing the Right Question in Oncology Clinical Trials, JCO Precision Oncology, 2019, doi: 10.1200/PO.18.00381

The devil is in the detail
Missing Data vs Intercurrent Events

	Missing Data	Intercurrent Event
Definition	Data that would be *meaningful* for the analysis of a given estimand *but* were *not collected*.	Events occurring after treatment initiation that *affect either the interpretation or the existence* of the measurements associated with the clinical question of interest.
Example	Study withdrawal	Discontinuation of randomized treatment
Handling	In the *statistical analysis* as a missing data problem; justification and sensitivity analysis is expected	Occurrence needs to be *considered explicitly* when defining the clinical question of interest and respective estimand
Impact on Clinical Trial Quality	It is *encouraged to avoid missing data* and it is seen as a *drawback* of clinical trial quality.	Measures to reduce or avoid intercurrent events that would *normally occur in clinical practice* risk reducing the external validity of the trial.

Population
It’s the target population which counts

Implicit assumptions during the analysis

SAP:
PRO analysis set are those patients with baseline and at least one postbaseline assessment for a specific PRO like EORTC QLQ-30. ..... Analysis will be on the PRO analysis set.

This approach (completer analysis) assumes that those patients with assessments are representative for those w/o post-baseline assessments. It’s rather not the case, is it?

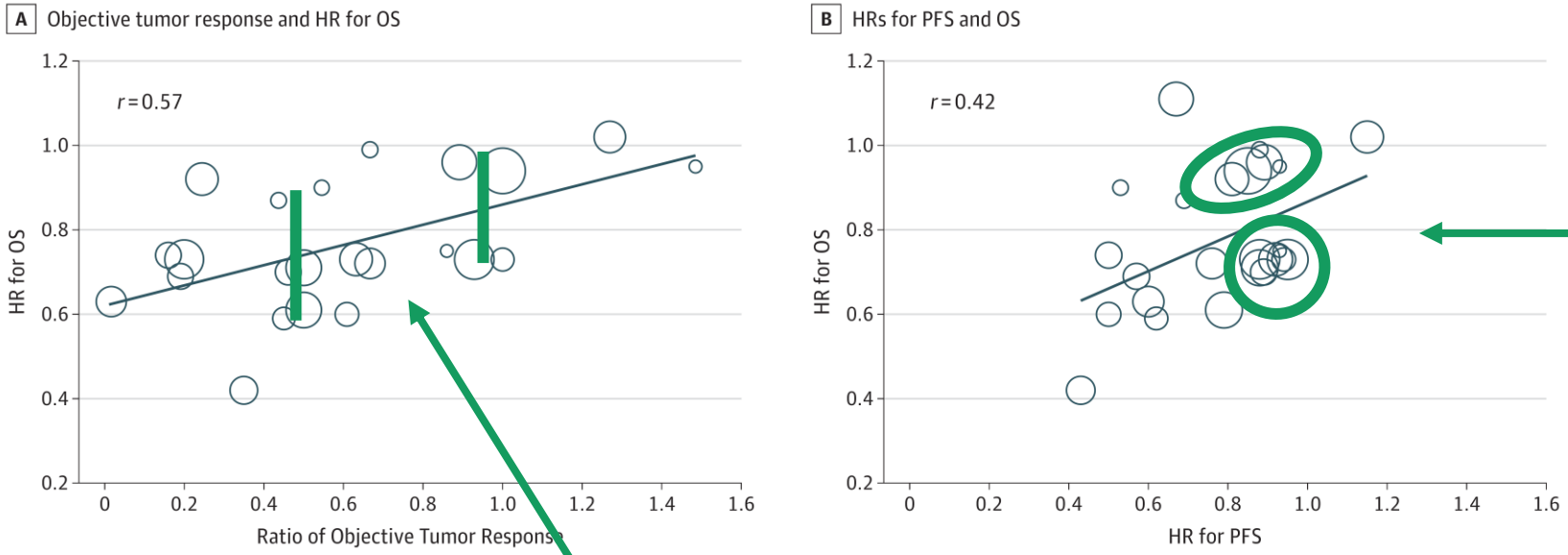
Centre selection

Phase II trial with focus on academic centres. Might lead to include patients who are not representative
- Able to travel to the academic centre
- Enrichment of severe cases as they might not be treated by the local hospitals

Endpoints

Association between endpoints for Immune Checkpoint Inhibitors

Figure 2. Correlations in Relative Treatment Effect



A, Ratio of objective tumor response and hazard ratio (HR) for overall survival (OS). B, Hazard ratio for progression-free survival (PFS) and HR for OS. Each circle represents a trial or treatment comparison. The circle size is proportional to the number of patients.

HR for PFS of 0.9 could lead to a HR for OS of ~0.95 or ~0.7

⇒Not a good predictor of OS treatment effect

- Doubling response rate could lead to a HR for OS of ~0.9 to ~0.6
- No effect on response rate could lead to a HR for OS of ~0.95 to ~0.7

⇒Not a good predictor of OS treatment effect

Richie G. et al., Defining the Most Appropriate Primary End Point in Phase 2 Trials of Immune Checkpoint Inhibitors for Advanced Solid Cancers, JAMA Oncol. 2018;4(4):522-528. doi:10.1001/jamaoncol.2017.5236

Are we asking the right question?
Example: renal cell carcinoma

References & citations driven by

Casey M, Degtyarev E, Lechuga MJ, et al. Estimand framework: Are we asking the right questions? A case study in the solid tumor setting. Pharmaceutical Statistics. 2020;1–11. https://doi.org/10.1002/pst.2079

and the ICH E9 (R1) guidance

Scientific question: Does the new treatment prolong patients' DFS time?

Received: 25 February 2020 | Revised: 11 September 2020 | Accepted: 22 October 2020
DOI: 10.1002/pst.2079

MAIN PAPER

WILEY

Estimand framework: Are we asking the right questions? A case study in the solid tumor setting

Michelle Casey[1] | Evgeny Degtyarev[2] | María José Lechuga[3] | Paola Aimone[2] | Alain Ravaud[4] | Robert J. Motzer[5] | Feng Liu[6] | Viktoriya Stalbovskaya[7] | Rui Tang[8] | Emily Butler[9] | Oliver Sailer[10] | Susan Halabi[11] | Daniel George[12]

[1]Pfizer, Inc., Collegeville, Pennsylvania
[2]Novartis Pharma AG, Basel, Switzerland
[3]Pfizer Srl, Milan, Italy
[4]Bordeaux University Hospital, Bordeaux, France
[5]Memorial Sloan Kettering Cancer Center, New York, New York
[6]AstraZeneca, Gaithersburg, Maryland
[7]Merus NV, Utrecht, The Netherlands
[8]Servier Pharmaceuticals, Boston, Massachusetts
[9]GlaxoSmithKline, Collegeville, Pennsylvania
[10]Boehringer Ingelheim Pharma GmbH & Co, Biberach, Germany
[11]Duke University, Durham, North Carolina
[12]Duke Cancer Center, Durham, North Carolina

Correspondence
Michelle Casey, Pfizer Inc., 500 Arcola Road, Collegeville, PA 19462, USA.
Email: michelle.casey@pfizer.com

Summary
The estimand framework requires a precise definition of the clinical question of interest (the estimand) as different ways of accounting for "intercurrent" events post randomization may result in different scientific questions. The initiation of subsequent therapy is common in oncology clinical trials and is considered an intercurrent event if the start of such therapy occurs prior to a recurrence or progression event. Three possible ways to account for this intercurrent event in the analysis are to censor at initiation, consider recurrence or progression events (including death) that occur before and after the initiation of subsequent therapy, or consider the start of subsequent therapy as an event in and of itself. The new estimand framework clarifies that these analyses address different questions ("does the drug delay recurrence if no patient had received subsequent therapy?" vs "does the drug delay recurrence with or without subsequent therapy?" vs "does the drug delay recurrence or start of subsequent therapy?"). The framework facilitates discussions during clinical trial planning and design to ensure alignment between the key question of interest, the analysis, and interpretation. This article is a result of a cross-industry collaboration to connect the International Council for Harmonisation E9 addendum concepts to applications. Data from previously reported randomized phase 3 studies in the renal cell carcinoma setting are used to consider common intercurrent events in solid tumor studies, and to illustrate different scientific questions and the consequences of the estimand choice for study design, data collection, analysis, and interpretation.

KEYWORDS
EFSPI SIG estimands in oncology, estimand, ICH E9, intercurrent events, time-to-event data

1 | INTRODUCTION AND BACKGROUND

An addendum to the International Council for Harmonisation (ICH) E9 guideline[1] on Statistical Principles for Clinical Trials, adopted in November 2019, introduced an estimand framework. The framework recognizes that there are multiple ways to quantify the treatment effect, and that changing the derivation of the endpoint or handling intercurrent events (ie, events that occur after baseline but before observing the endpoint of interest) differently targets different scientific questions. Different disease indications may require different scientific questions, and thus the same intercurrent

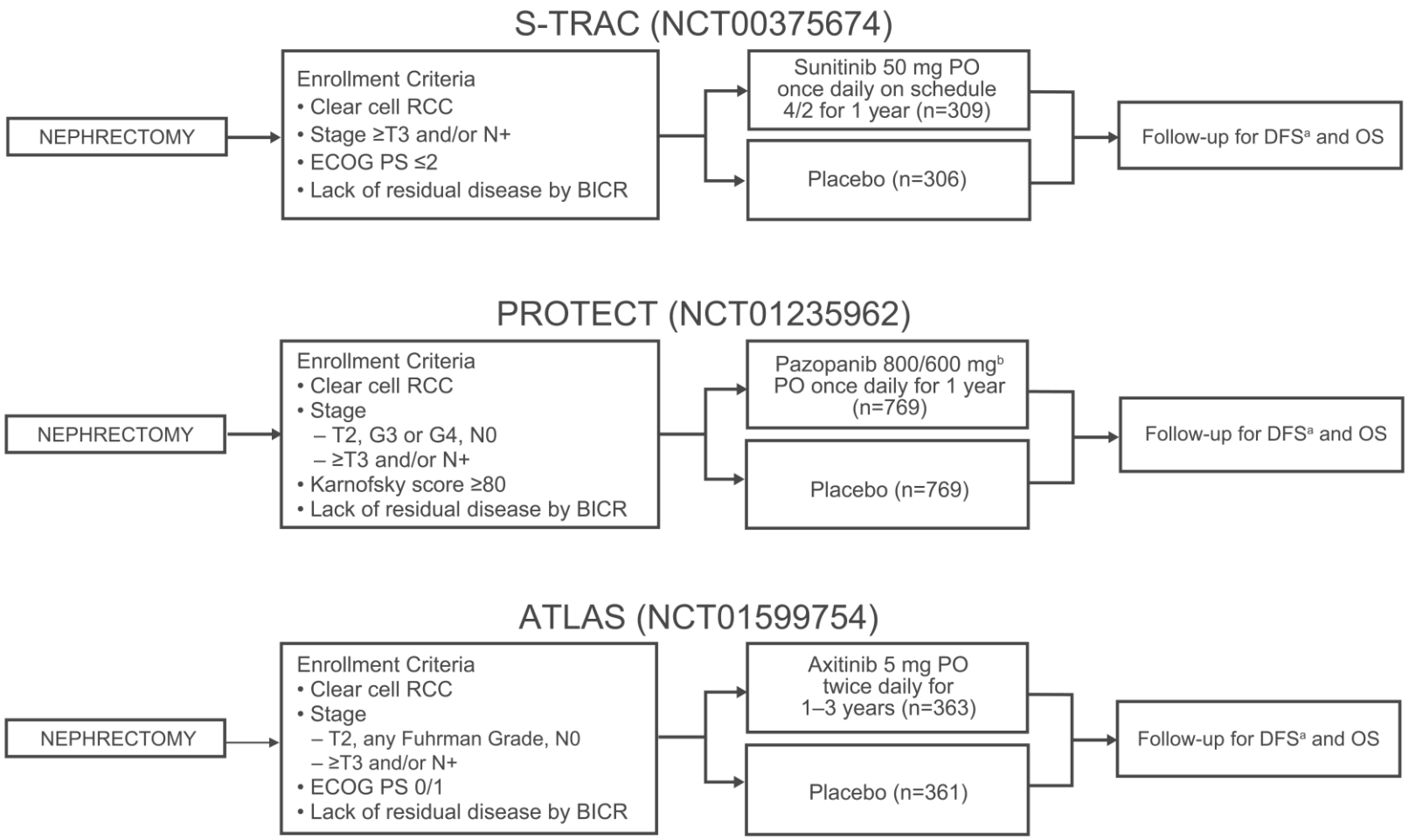
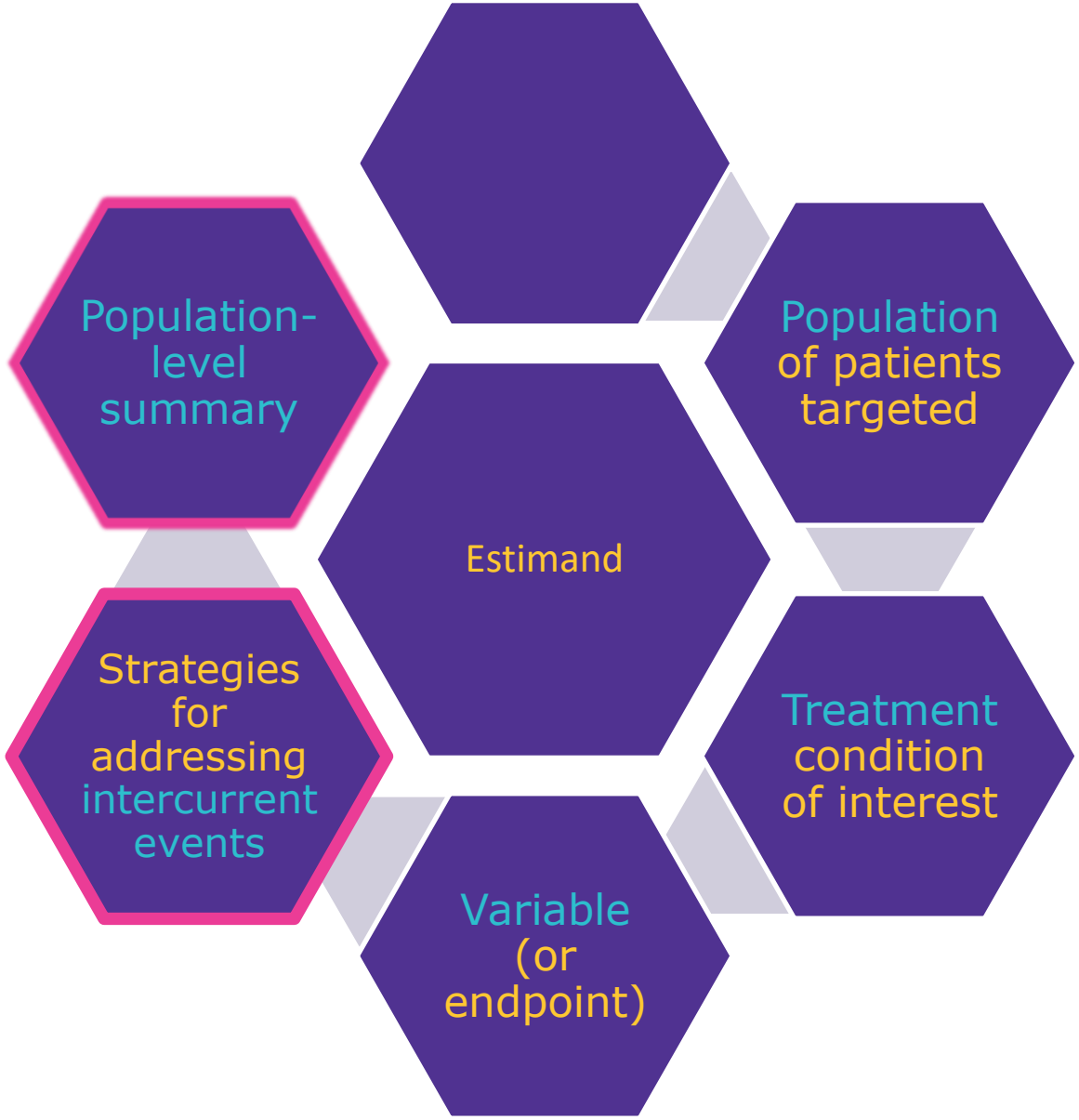


FIGURE 1 Study designs for S-TRAC, PROTECT, and ATLAS. Abbreviations: BICR, blinded independent central review; DFS, disease-free survival; ECOG PS, Eastern Cooperative Oncology Group performance status; G, Fuhrman Grade; N, node stage; OS, overall survival; PO, by mouth; RCC, renal cell carcinoma; T, tumor stage. [a]Denotes primary endpoint. [b]Initial starting dose was 800 mg; this was lowered to 600 mg to address toxicity attrition with the 800-mg starting dose (Casey, 2020)

Let’s link the intercurrent event subsequent therapy with potential clinical questions



3 different clinical questions
Intercurrent event subsequent therapy

Clinical Question	Analysis	Estimand
Does the drug delay recurrence if no patient had received subsequent therapy?	Censor at start of subsequent therapy	**Hypothetical** A scenario is envisaged in which the intercurrent event would not occur

3 different clinical questions
Intercurrent event subsequent therapy

Clinical Question	Analysis	Estimand
Does the drug delay recurrence if no patient had received subsequent therapy?	Censor at start of subsequent therapy	**Hypothetical** A scenario is envisaged in which the intercurrent event would not occur
Does the drug delay recurrence with or without subsequent therapy?	Use all events, even after start of subsequent therapy	**Treatment policy** The occurrence of the intercurrent event is considered irrelevant

3 different clinical questions
Intercurrent event subsequent therapy

Clinical Question	Analysis	Estimand
Does the drug delay recurrence if no patient had received subsequent therapy?	Censor at start of subsequent therapy	**Hypothetical** A scenario is envisaged in which the intercurrent event would not occur
Does the drug delay recurrence with or without subsequent therapy?	Use all events, even after start of subsequent therapy	**Treatment policy** The occurrence of the intercurrent event is considered irrelevant
Does the drug delay recurrence or start of subsequent therapy?	Start of subsequent therapy is an event itself like recurrence or death	**Composite Strategy** An intercurrent event is considered in itself to be informative about the patient's outcome and is therefore incorporated into the definition of the variable.

(Casey, 2020)

Mapping to the estimands framework
Differences between S-TRAC, PROTECT, ATLAS

	S-TRACT	Protect	ATLAS
Population	Stage: ≥T3 and/or N+	Stage: **T2, G3 or G4, N0** ≥T3 and/or N+	Stage: **T2, any Fuhrman grade, N0,** ≥T3 and/or N+
Endpoint (DFS)	**Recurrence, second primary cancer**, death from any cause	**Local recurrence, metastasis,** death from any cause	**Recurrence, second primary cancer**, death from any cause
Intercurrent events handling	Composite: Including deaths and **second primary malignancy** in the endpoint Treatment policy: Ignoring treatment discontinuation or dose modifications due to tolerability Hypothetical: In the situation where subsequent therapy had not been administered	Composite: Including deaths in the endpoint Treatment policy: Ignoring treatment discontinuation or dose modifications due to tolerability and **second primary malignancy** Hypothetical: In the situation where subsequent therapy had not been administered	Composite: Including deaths and **second primary malignancy** in the endpoint Treatment policy: Ignoring treatment discontinuations or dose modifications due to tolerability Hypothetical: In the situation where subsequent therapy had not been administered

(Casey, 2020)

MERCK

Mapping to the estimands framework
Differences between S-TRAC, PROTECT, ATLAS

	S-TRACT	Protect	ATLAS
Scan date	**Latest date** used for equivocal findings	**Earliest date** used for equivocal findings	**Earliest date** used for equivocal findings
Assessment Schedule	Every 12 weeks during the first 3 years, then every 6 months thereafter	Weeks 20, 36, and 52 during year 1, every 6 months during years 2-5, and yearly thereafter	Every 16 weeks during first 3 years, every 6 months thereafter
Censoring in case of missing assessments	Patients were censored in case of **two or more missed assessments** prior to an event	Patients were censored in case of **two or more missed assessments** prior to an event	**No censoring**, all events were counted

(Casey, 2020)

Why Estimands?

Structured framework to strengthen the dialogues between disciplines involved in the formulation of clinical study objectives, design, conduct, analysis and interpretation

- Precision in describing a treatment effect of interest is facilitated by constructing the estimand
- **Promote discussion** of different strategies to address intercurrent events and missing data
- Inform appropriate choices about the trial design and data collection
- **Increase transparency** with respect to data analysis and inference
- Clarify the role of sensitivity analysis to explore robustness of conclusions from the main statistical analysis

➢ It's like with a recipe to cook a soup. It clearly describes the ingredients and the way to combine them

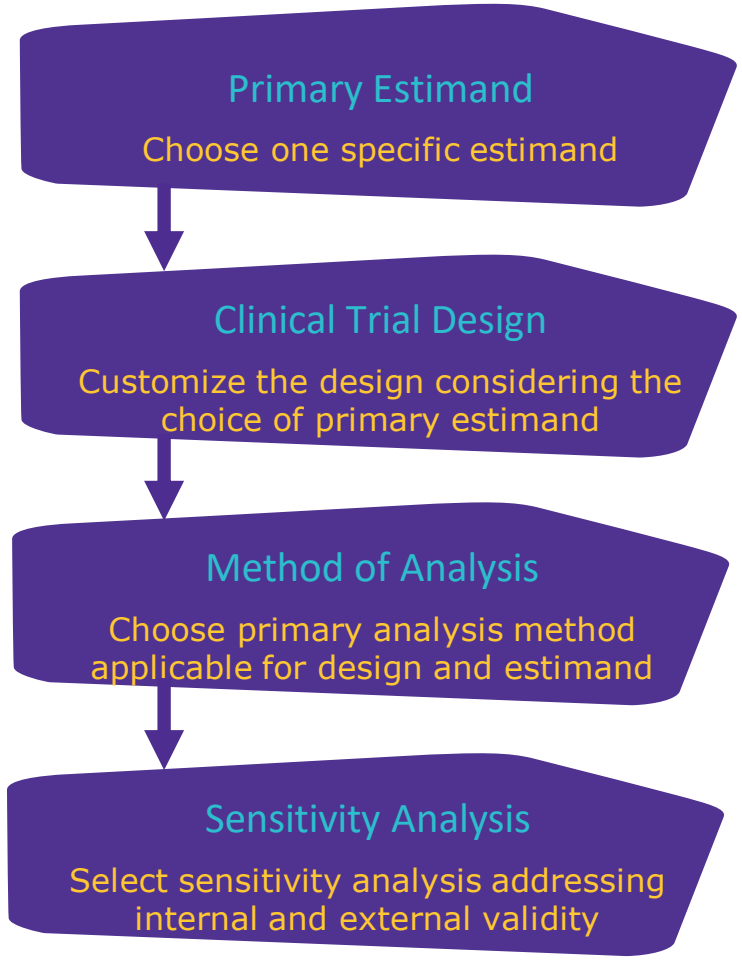


Figure adapted from Leuchs et al (2015): Choosing Appropriate Estimands in Clinical Trial, doi: 10.1177/2168479014567317

MERCK

"The estimand framework does not change common analysis methods within oncology. Rather, it facilitates discussions during clinical trial design to ensure alignment between the key question of interest; the analysis, including harmonization of the individual elements of composite endpoints; and interpretation"
(Casey, 2020)

Better alignment between the clinical question and the associated analyses will lead to a more systematic approach to design clinicals trials. A better ability to answer the clinical question will ultimately lead to a higher probability of success.

Take home messages

Motivation:

International Council for Harmonisation (ICH) has released an addendum to E9 on estimands and sensitivity analysis in clinical trials.

What is an Estimand?

A precise description of the treatment effect reflecting the clinical question posed by the trial objective, and is described by five attributes.

Why?

Provides a structured framework to strengthen the dialogues between disciplines to improve precision in describing a treatment effect of interest.

When?

Always

Required for primary and key secondary endpoints in studies with registrational intention.


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

This Guideline has been developed by the appropriate ICH Expert Working Group and has been subject to consultation by the regulatory parties, in accordance with the ICH Process. At Step 4 of the Process the final draft is recommended for adoption to the regulatory bodies of ICH regions.