Implications of the ICH E9 estimand addendum on how we develop, run, and analyse clinical trials

Kaspar Rufibach
Methods, Collaboration, and Outreach Group, Roche Data and Statistical Sciences
ANVISA, 19th May 2021

Acknowledgments

Audrey Souverain (Roche) for working up YOSEMITE example.

Many colleagues for many discussions over the last five years.

Agenda

1. Endpoints
2. ICH E9 addendum
3. Examples
   - Diabetic Macular Edema
   - NALT in lymphoma
4. Covid-19
5. Scope
6. What does it mean for clinical and safety science
7. Impact and conclusions
8. Backup

Agenda

1. Endpoints
2. ICH E9 addendum
3. Examples
   - Diabetic Macular Edema
   - NALT in lymphoma
4. Covid-19
5. Scope
6. What does it mean for clinical and safety science
7. Impact and conclusions
8. Backup

Endpoints

Traditional time-to-event endpoints

Time-to-event endpoints

Interval between two **clearly** defined - potentially composite - events.

Why time-to-event endpoints? Measure of **clinical benefit** in late phases.

Life-threatening diseases, disease for which no treatment exists, accelerate approval $\Rightarrow$ **surrogate** endpoints:

- response proportion,
- pathological complete response (pCR) in breast cancer,
- minimal residual disease (MRD) in lymphoma.

Overall survival

Time-to-event endpoints

Overall survival (OS):

- Randomization to death of **any cause**.
- Obvious measure of clinical benefit.
- Objectively determined, i.e. no influence of investigator. Blinding sometimes not necessary.
- Easily measurable to the day.

Disadvantages OS:

- Includes deaths unrelated to cancer.
- Potentially **confounded** by subsequent therapies ("cross-over").
- Modest hazard ratio improvement $\Rightarrow$ high #events required.
- Large medians $\Rightarrow$ long trial duration.

Trial on OS might be unfeasible.

Alternative time-to-event endpoints

Alternative time-to-event endpoints

Reduce trial cost, duration, #patients ⇒ use **surrogates** of OS. See Bellera et al. (2013).

General remarks:

- Different level of established surrogacy of alternative endpoints for OS.
- Endpoint the more credible the better association with OS.
- Requires balanced timing of assessment among treatment arms.
- Typically investigator-influenced ⇒ **blinding**, **independent response review** to minimize bias.

Progression-free survival (PFS):

- For advanced disease.
- Randomization to PD or death.
- Not only surrogate for OS, carries value in itself.
- Not affected by cross-over after PD.

Alternative time-to-event endpoints

Time to progression (TTP):

- Randomization to PD. Death is censored.
- Acceptable if majority of deaths unrelated to cancer.
- Less related to OS than PFS.

Time to treatment failure (TTF):

- Randomization to discontinuation (PD, death, toxicity patient decision).
- Composite endpoint influenced by factors **unrelated to efficacy**.
- Useful if toxicity potentially as serious as PD (e.g. allogeneic stem cell transplant).

Event-free survival (EFS):

- Randomization to PD, death, discontinuation, new treatment without PD.
- Useful to evaluate highly toxic therapies.
- Initiation of new therapy subjective.

References time-to-event endpoints

Disease-free survival (DFS):

- Randomization to recurrence or death.
- Typically in setting where patient "disease-free": surgery (+ adjuvant).
- Clinical benefit declared if benefit outweighs toxicity of adjuvant treatment.

Bellera et al. (2013), Pazdur (2008), Johnson et al. (2003).

FDA guidance: `https://www.fda.gov/regulatory-information/search-fda-guidance-documents/clinical-trial-endpoints-approval-cancer-drugs-and-biologics`.

Do we really know our endpoints?

Some (counter)examples.

DFS in breast cancer

Lack of common definition: DFS in breast cancer

Primary for many large adjuvant breast cancer trials.

Table 1. Example of Inconsistent Definitions of Disease-Free Survival

Trial	Local/Regional Recurrence	Distant Metastasis	Death From Any Cause	Invasive Contralateral Breast Cancer	Second Primary Invasive Cancer (nonbreast)	Ipsilateral DCIS	Contralateral DCIS	Ipsilateral LCIS	Contralateral LCIS
BIG 1-98[4]	X	X	X	X	X				
MA-17[1]	X	X		X		X	X	X	X
ATAC[2]	X	X	X	X		X	X		
IES[3]	X	X	X	X					
ARNO[5]	X	X		X					

NOTE: Event-free survival used by ARNO.
Abbreviations: DCIS, ductal carcinoma in situ; LCIS, lobular carcinoma in situ; BIG, Breast International Group; MA, National Cancer Institute of Canada Clinical Trials Group MA-17; ATAC, Arimidex, Tamoxifen Alone, or in Combination; IES, Intergroup Exemestane 031; ARNO, Arimidex, Nolvadex 95 Study.

Five trials - no two have same endpoint definition!

Hudis et al. (2007)

New anti-lymphoma therapy and PFS

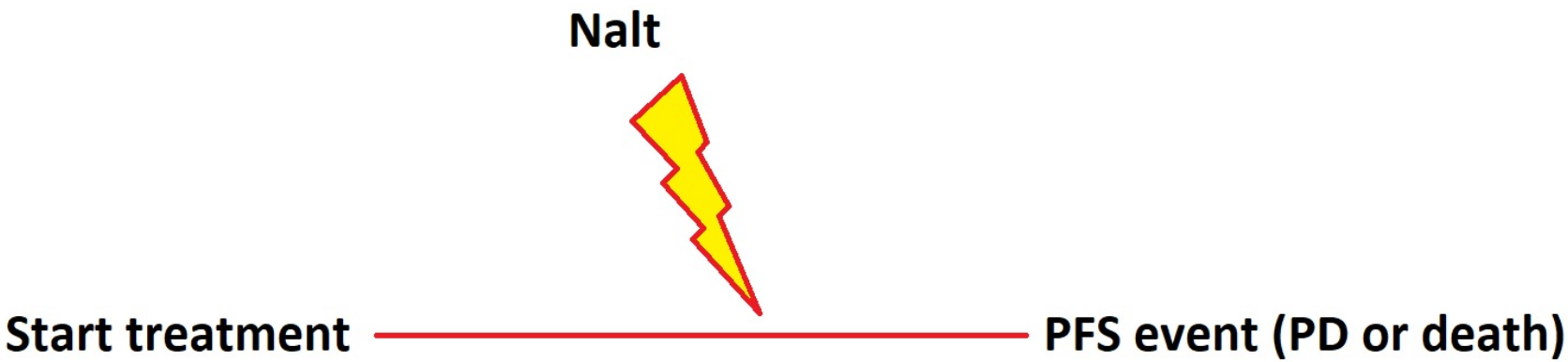
Endpoint definitions in GALLIUM: PFS

The primary end point was progression-free survival, as determined on the basis of the assessment of tumors at an independent review facility (independently assessed progression-free survival). Progression-free survival was defined as the time from randomization to the first documented radiographic evidence of progressive disease according to Response Evaluation Criteria in Solid Tumors (RECIST), version 1.0,[20] or death from any cause within 18 weeks after the last independent assessment of tumors. Data for patients who did

Marcus et al. (2017)

NALT?

NALT?



Treatment effect changes!

Interpretation of PFS different depending on whether patient received **new anti-lymphoma treatment** (NALT) or not!

How to deal with NALT?

Depends on scientific objective of trial!

Objective $\Rightarrow$ estimand $\Rightarrow$ estimator.

Hemophilia

Hemophelia A

Primary endpoint: **Annualized bleeding rate**, recurrent event type.

Study	Publication	Primary endpoint	High level design	Primary test
Advate	Valentino, 2011	ABR	Randomized comp of 2 prophy + within pt on demand vs prophy	
Advate	Mahlangu, 2013	ABR	Randomized comp of 2 prophy, 1 on demand	
SPINART, Kogenate	Manco-Johnson, 2013	ABR	Randomized comp of on demand vs prophy	
Pro-FEIBA	Leissinger, 2011	ABR	Crossover on demand vs prophy	
Hemophilia B	Powell, 2013	ABR	Non randomized 4 arm comparison	
FEIBA	Antunes, 2014	ABR	Randomized comp of on demand vs prophy	

Hemophelia A

Primary endpoint: **Annualized bleeding rate**, recurrent event type.

Study	Publication	Primary endpoint	High level design	Primary test
Advate	Valentino, 2011	ABR	Randomized comp of 2 prophy + within pt on demand vs prophy	t-test, ITT and PP
Advate	Mahlangu, 2013	ABR	Randomized comp of 2 prophy, 1 on demand	Neg bin mixed?, «PP»
SPINART, Kogenate	Manco-Johnson, 2013	ABR	Randomized comp of on demand vs prophy	Neg bin offset, ITT
Pro-FEIBA	Leissinger, 2011	ABR	Crossover on demand vs prophy	Wilcoxon
Hemophilia B	Powell, 2013	ABR	Non randomized 4 arm comparison	Test for rates, ABR est from neg bin, PP
FEIBA	Antunes, 2014	ABR	Randomized comp of on demand vs prophy	t-test of transformed ABR, ITT and PP

Observation: heterogeneity in endpoint definitions

Time-to-event endpoints **not well defined** in clinical trials, persistently over time:

- Peto et al. (1977): recommendations how to define and report time-to-event endpoints.
- Altman et al. (1995): finds **frequent failure** to specify whether non-cancer deaths treated as events or censored or how deaths without relapse were considered. Re-iterate need for clear definition of *time origin, the event of interest and the circumstances where survival times are censored*.
- Mathoulin-Pelissier et al. (2008): $\approx 50\%$ of reviewed articles in major clinical journals **failed** to provide clear endpoint definition. 68% reported insufficient information on survival analysis.
- Bellera et al. (2013): *Most of these time-to-event endpoints currently* **lack standardised definition** *enabling a cross comparison of results from different clinical trials.*
- Endpoint definition decides on **significant or non-significant**, PETACC-03, colon cancer, DFS as primary endpoint, Van Cutsem et al. (2005).

Applies to **all endpoint types and indications**.

Agenda

1. Endpoints
2. ICH E9 addendum
3. Examples
   - Diabetic Macular Edema
   - NALT in lymphoma
4. Covid-19
5. Scope
6. What does it mean for clinical and safety science
7. Impact and conclusions
8. Backup

ICH E9 addendum

ICH E9 addendum

ICH E9: "Statistical principles for Clinical Trials."

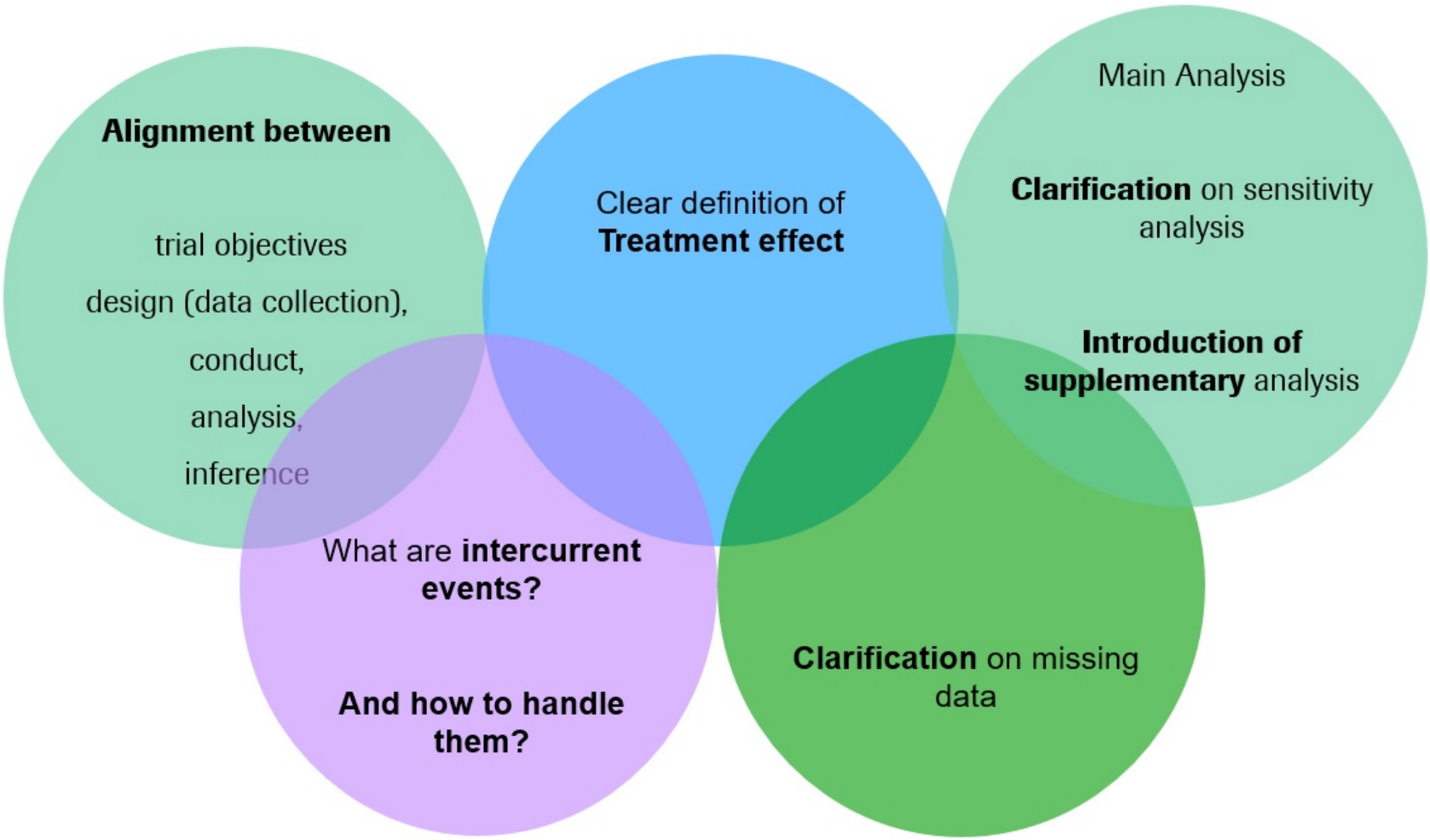
1998.

Why amend E9?

Trial objectives:

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

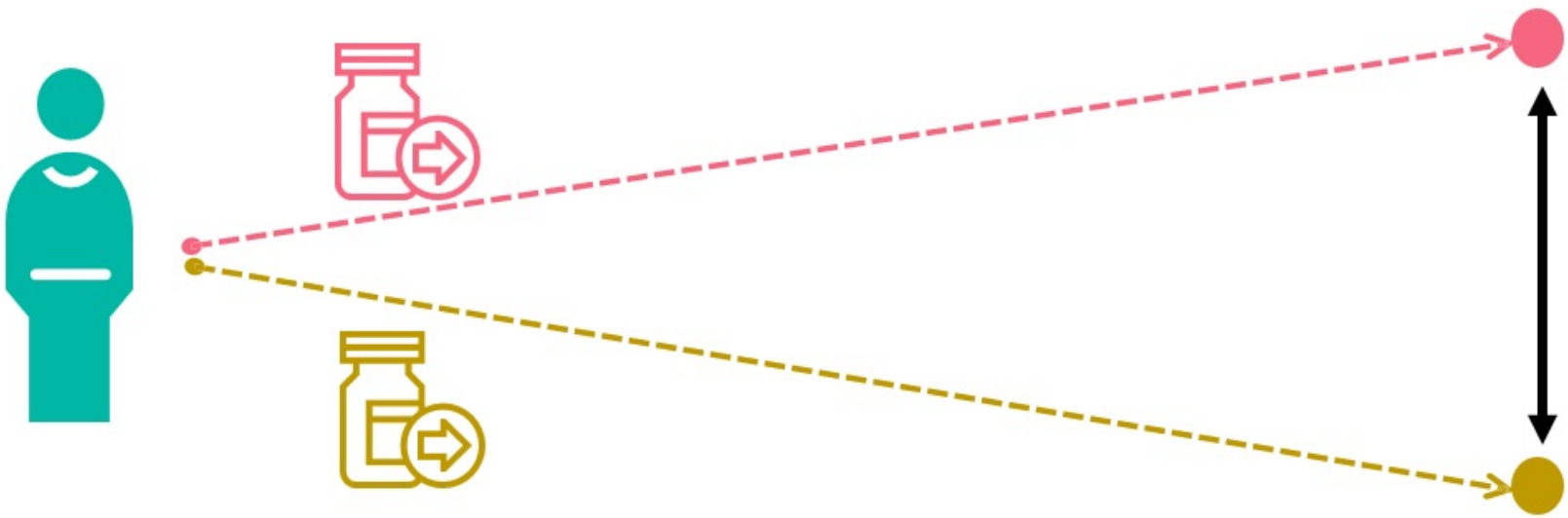
Estimand framework



What is a “treatment effect”?

Treatment effect

Not defined in original E9!



Intention-To-Treat (ITT) principle

Effect of **treatment policy**:

- Best assessed using planned treatment regimen rather than actual treatment given.
- Analysis based on all subjects.
- Subjects to be included in the analysis as randomized.
- Subjects are **followed, assessed and analyzed** irrespective of completion of planned course of treatment.

What ITT does not tell us!

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

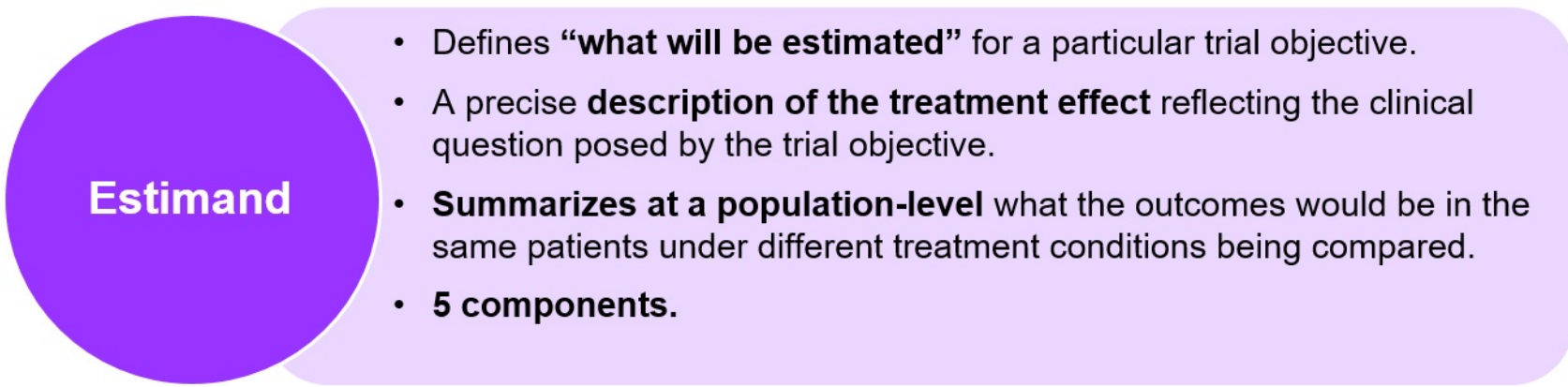
Lack of clarity in the definition of treatment effect

May have very serious implications!

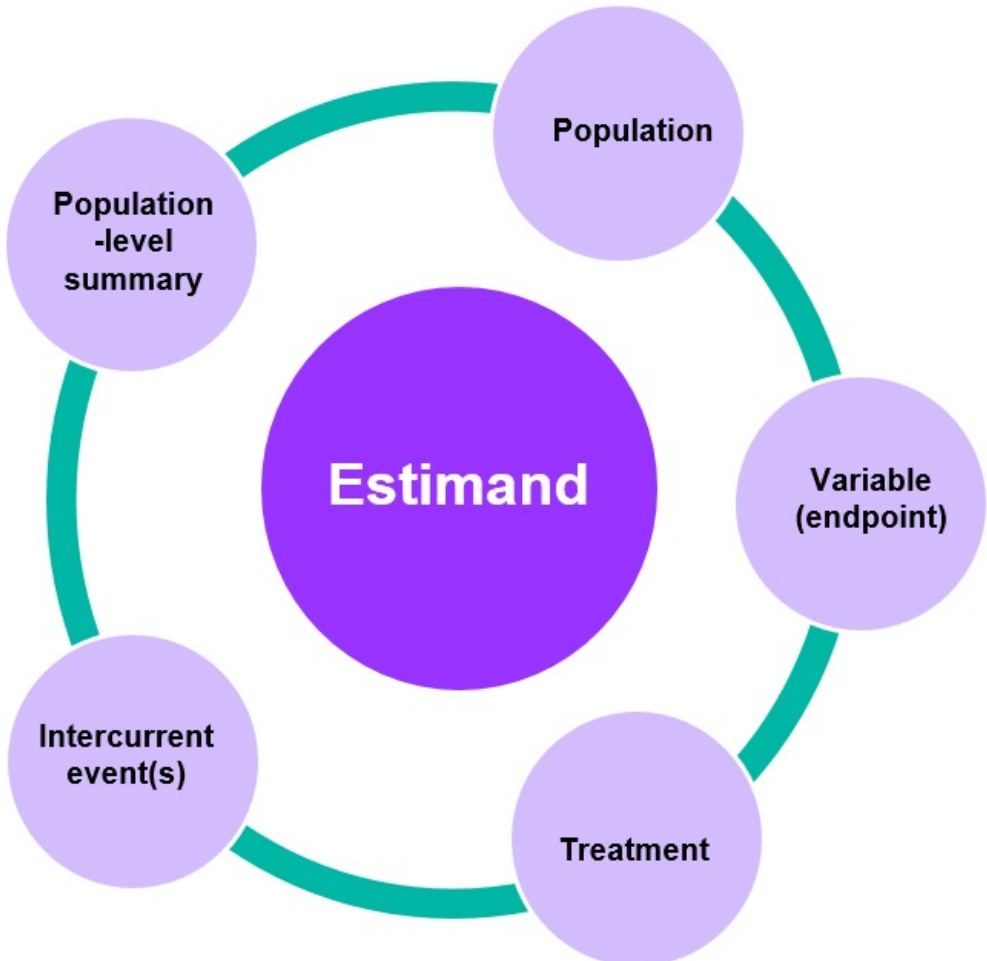
Bad drugs might appear good!

Good drugs might appear bad!

Definition of an estimand



Estimand attributes



Agenda

1. Endpoints
2. ICH E9 addendum
3. Examples
   - Diabetic Macular Edema
   - NALT in lymphoma
4. Covid-19
5. Scope
6. What does it mean for clinical and safety science
7. Impact and conclusions
8. Backup

Examples

Diabetic Macular Edema

Ambiguity!

Primary endpoint YOSEMITE trial:

> *Difference between faricimab and Aflibercept in change from baseline in best-corrected visual acuity (BCVA) at 1 year.*

Ambiguity!

Some patients will tolerate faricimab and **adhere** to its administration schedule, others will not.

Ambiguity!

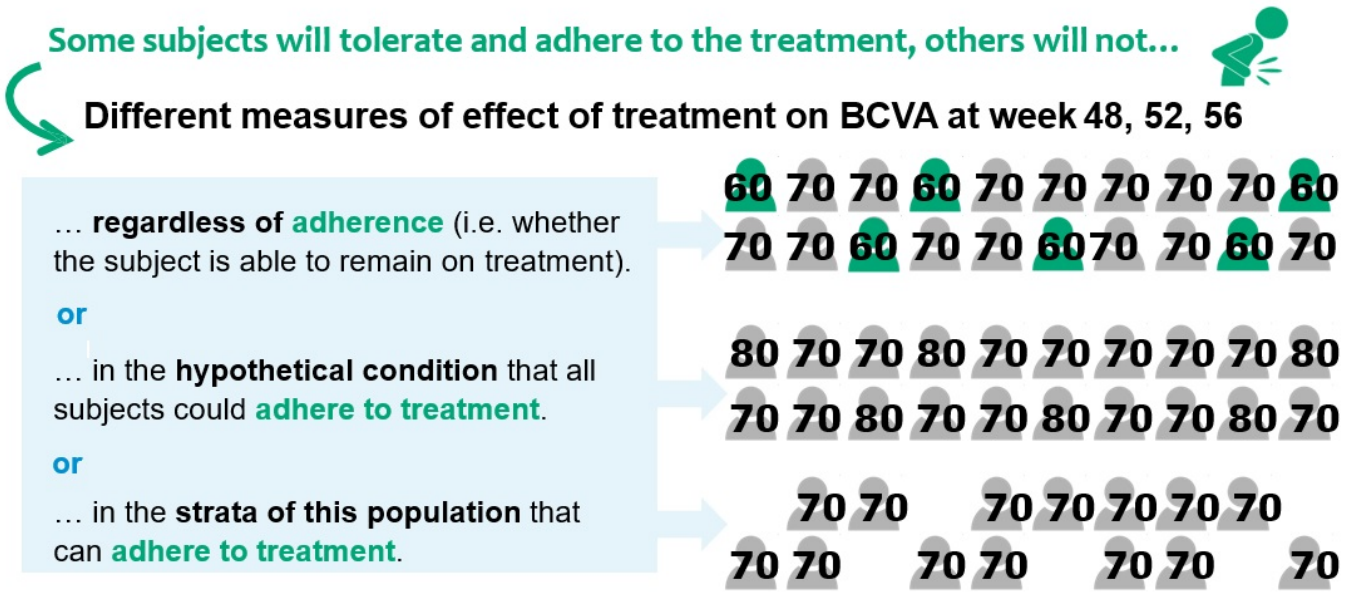
Primary endpoint YOSEMITE trial:

> *Difference between faricimab and Aflibercept in change from baseline in best-corrected visual acuity (BCVA) at 1 year.*

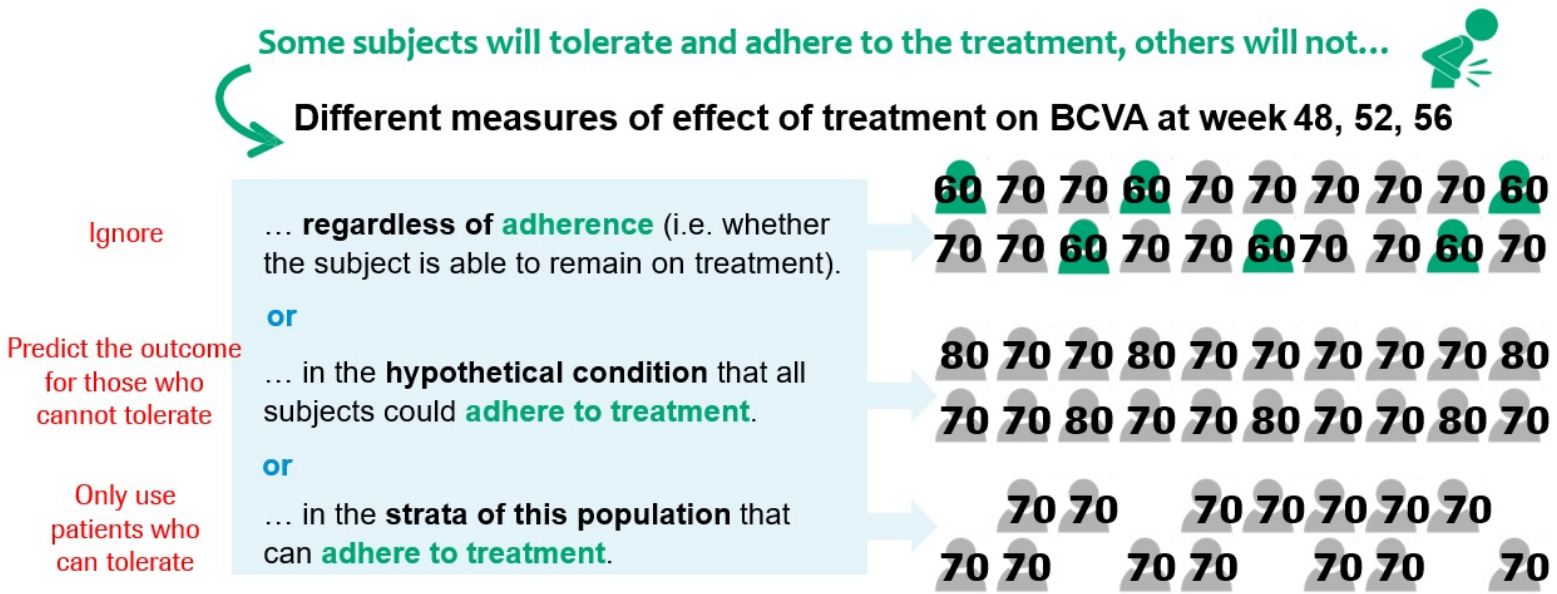
Ambiguity!

Some patients will **require changes in dose or administration** of additional medication (e.g. concomitant or rescue medication, treatment switch, etc.), others will not.

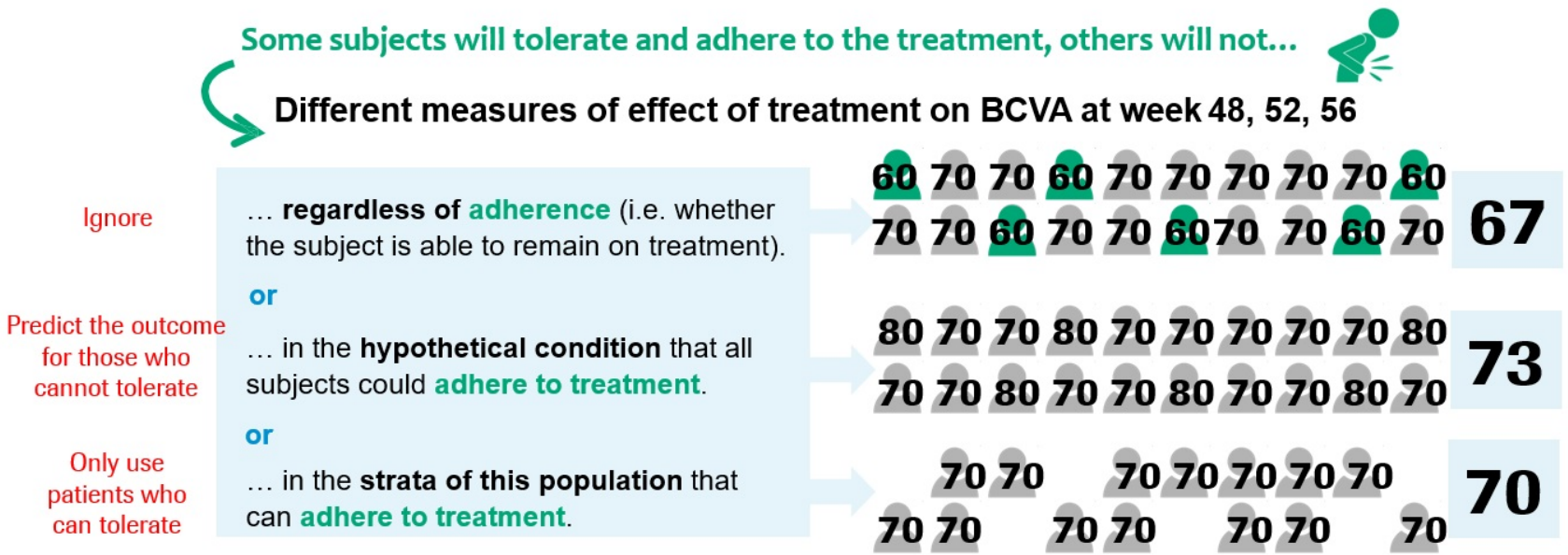
Multiple definitions of treatment effect



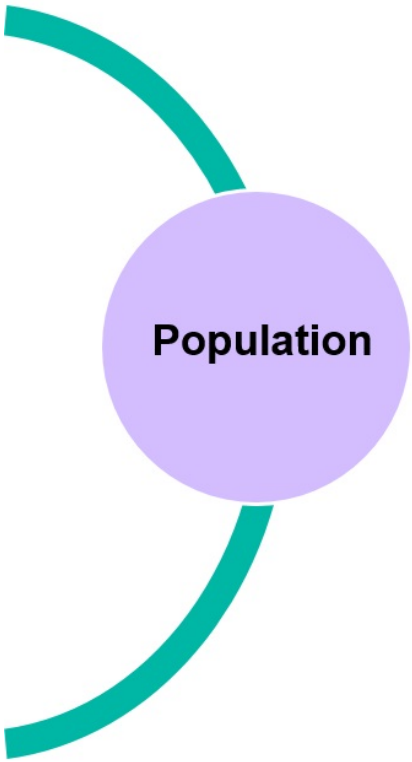
Multiple definitions of treatment effect



Multiple definitions of treatment effect



Attribute: population

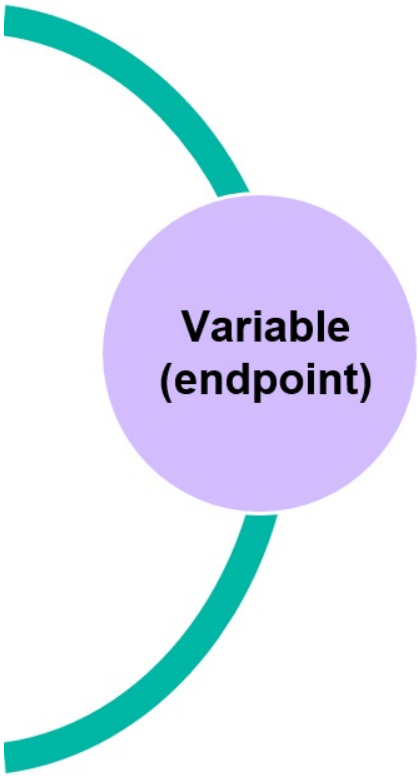


Subjects targeted by the scientific question.

YOSEMITE primary endpoint population:

Adult patients with DME, either treatment-naive or prior IVT anti-VEGF treated, as defined by the inclusion / exclusion criteria.

Attribute: variable (endpoint)

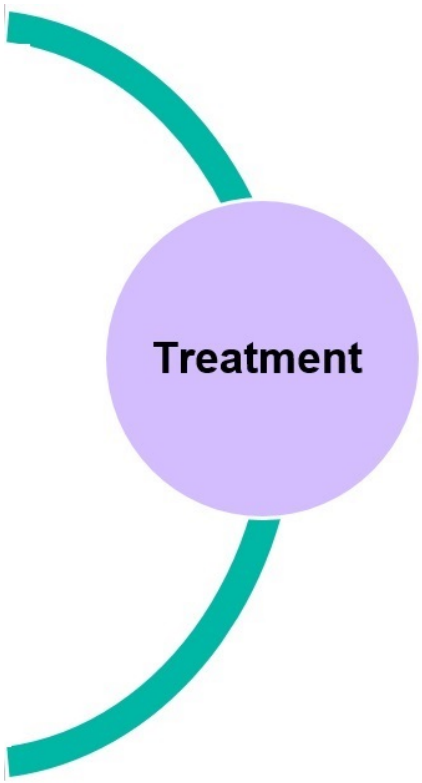


Measure(s) required to address the scientific question (to be obtained for each subject).

YOSEMITE primary endpoint:

Change in BCVA score from baseline averaged over Weeks 48, 52, and 56.

Attribute: treatment

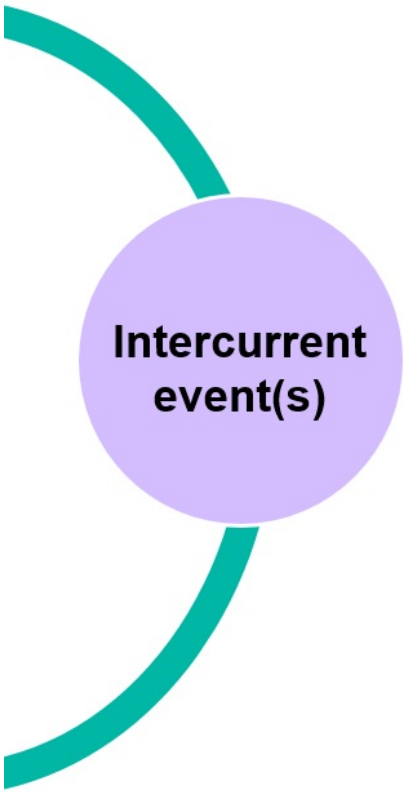


The **treatment of interest**, and if appropriate, the **alternative treatment** to which **comparison** will be made.

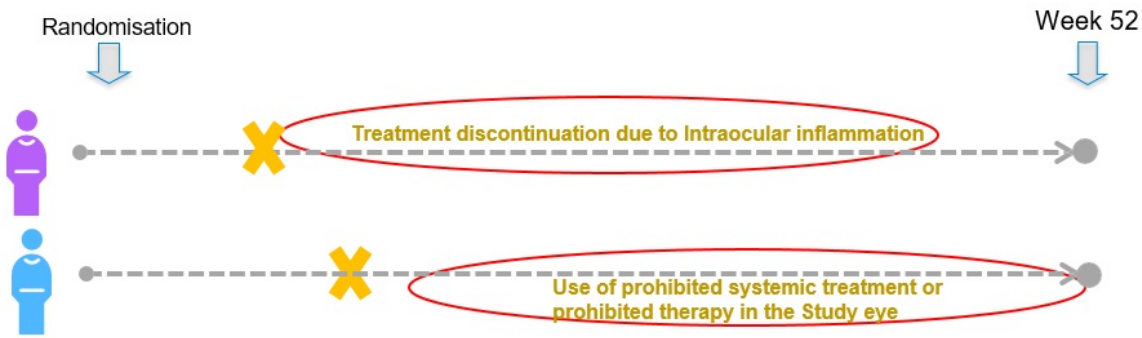
YOSEMITE:

The primary comparisons will be the pairwise comparisons between the active comparator (aflibercept Q8W) and each of the faricimab arms (Q8W and PTI).

Attribute: intercurrent events



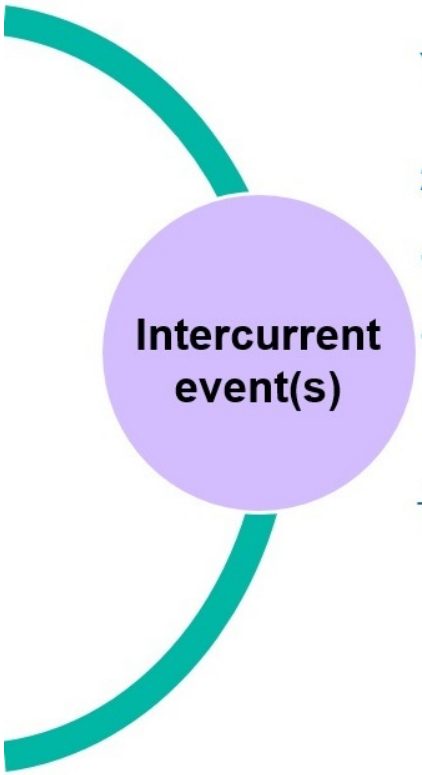
Events that **occur after treatment initiation** and either **preclude observation** of the variable or **affect** its **interpretation.**



Analysis plan to **specify** on **how to account for intercurrent events**

- **Different strategies** available (5 strategies described in addendum).
- Strategies to be **aligned** with **clinical question** of interest.
- Not necessary to use the same strategy to address all intercurrent events.

Attribute: intercurrent events



YOSEMITE:

2 intercurrent events:

- Discontinuation of study treatment due to adverse events or lack of efficacy.
- Use of any prohibited systemic treatment or prohibited therapy in the Study eye.

Intercurrent events handled with **treatment policy** strategy

- Intercurrent event is a part of the treatment effect
- **Intercurrent** event will be **ignored**
  - **All values** for BCVA will be **used** in the analysis regardless the occurrence of above intercurrent events.

Attribute: population-level summary

Provides a basis for a **comparison between treatment conditions**.

YOSEMITE:

Difference in **adjusted mean** between each of the two faricimab and aflibercept.

Adjusted mean

- Controlled for **baseline covariates.**
- **Covariates:** variables that **may be prognostic** for outcome.

NALT in lymphoma

What do we want to show?

Objective?

- PFS of initially randomized therapy if no NALT given? **Hypothetical strategy** for intercurrent event of NALT. Estimated through censoring at NALT.
- PFS irrespective of whether NALT received? **Treatment policy strategy**. Estimated through simply taking PFS event as observed.

Clinicians may disagree what the real objective is!

Follicular lymphoma (FL):

- Slowly progressing, incurable.
- New anti-lymphoma therapy (NALT) generally administered **after** PD.
- PFS with hypothetical?

Diffuse large B-cell lymphoma (DLBCL):

- Potentially curable. Patients failing to achieve response have dismal prognosis.
- NALT often initiated **prior to PD** when failing to achieve response.
- Rather count NALT an event, i.e. consider EFS?

Agenda

1. Endpoints
2. ICH E9 addendum
3. Examples
   - Diabetic Macular Edema
   - NALT in lymphoma
4. Covid-19
5. Scope
6. What does it mean for clinical and safety science
7. Impact and conclusions
8. Backup

Covid-19

Impact of COVID-19

- **Healthcare system disruptions**: missed visits, missed doses.
- **COVID-19 associated events**: use of prohibited medications, COVID-19 infection, death due to COVID-19.

Impact of COVID-19

YOSEMITE primary analysis: four intercurrent events related to COVID-19:

- **Discontinuation of study treatment** due to COVID-19.
- Use of any **prohibited systemic treatment** or **prohibited therapy in study eye** due to COVID-19.
- **Missed dose(s)** with potentially major impact on efficacy due to COVID-19.
- **Death** due to COVID-19.

Pandemic will end one day $\Rightarrow$ trial objective and primary estimand remain unchanged.

Hypothetical strategy for all the above intercurrent events.

Oncology context: Degtyarev et al. (2020).

Agenda

Scope

Scope

Endpoints:

- Thinking process useful to think about **any endpoint**.
- Addendum: endpoints that pertain to **regulatory decision-making** $\Leftrightarrow$ those for which you specify type I error protection.

Trial types:

- Whenever treatment effect is estimated or hypothesis related to a treatment effect is tested, whether related to efficacy or safety.
- Main focus on RCTs.
- Principles applicable for single arm trials and observational studies.

Data type: Binary, continuous, longitudinal, time-to-first event, recurrent event.

We have not talked about:

- Missing data.
- Sensitivity and supplementary analyses.

Agenda

What does it mean for clinical and safety science

Clinical and safety science

Clear trial objectives needed to identify potential estimands to target.

Defining estimands to target requires:

- Discussion between **sponsor and regulators**.
- Expertise from **multiple disciplines**.
- Additional time and **effort at design stage**.
- Discussion of relevant estimands in HA interactions.
- Adoption of estimands framework in all documentation.

Agenda

Impact and conclusions

Impact on data collection

- Estimand **dictates data that need to be collected**:
  - Treatment policy strategy: Data after treatment discontinuation!
  - Identification of missing data, reasons for missingness.
  - Types and timing of intercurrent events.
  - Depending on strategy: additional data needed, e.g. baseline characteristics or at switch.
- Each trial likely to have **multiple estimands** $\Rightarrow$ different estimands might require different data!
- Requires **multi-disciplinary** involvement from **earliest stages** of trial development.
- Impacts **design of eCRF** or other data collection tools and monitoring strategy.

Novo Nordisk, `https://www.dsbs.dk/moder/Estimands/HLynggaard.pdf`:

- Focussing on retention, keeping subjects in trial even after discontinuing trial drug.
- Increased completion rates from **90% to 98%** in type 1 diabetes studies and from **70% to over 90%** in obesity studies.

Might need to reflect this in **sample size computation**!

Impact on documentation

Protocols	Study population	Derive population from estimand definition
	Study intervention	Derive intervention from estimand definition, including rescue medicine
	Discontinuation	Derive discontinuation actions from intercurrent event strategies in estimand definition
	Statistical considerations	Hypothesis, analysis sets, sample size, endpoints follow from estimand definition Separate sensitivity from supplementary analyses.
Additonally for SAPs	Sample Size	Optionally provide (even) more details how intercurrent events are taken into account in sample size computation
Additonally for CSRs	Discontinuation	Tabulate observed intercurrent events.
	Changes in Planned Analyses Prior to Unblinding or DB lock	Discuss how intercurrent events that were not foreseen at the design stage, or identified during the conduct of the trial, were handled. Discuss not only the choices made for the analysis, but the effect on the estimand.

Broader impact

Aligning drug developers and regulatory bodies' expectations for target treatment effect **upfront** has potential to give:

- More **meaningful** descriptions of treatment effects for licensing and prescribing decisions.
- Clinical trials with designs that are **aligned to agreed objectives**.
- Increased **transparency** with respect to data analysis and inference.
- More **predictable** regulatory assessment procedures.
- More **flexibility** from regulators.
- **Reduction in total number of analyses** (primary + secondary + sensitivity).
- Clear language to describe and discuss different estimands required by different stakeholders.
- **Shift of resources** from analysis / filing to design.

Conclusions

Estimand ICH E9(R1) addendum:

- Turned out to be **highly useful for assessing impact of COVID-19** on trials.
- Requires **mindset change**.
- Area continues to evolve.
- Everyone is **still learning**!

All stakeholders need to work together!

Resources

Official ICH E9 training material: `www.database.ich.org/sites/default/files/E9-R1_EWG_Step2_TrainingMaterial.pdf`

3min intro video: `www.youtube.com/watch?v=oeoTOO1x37c`.

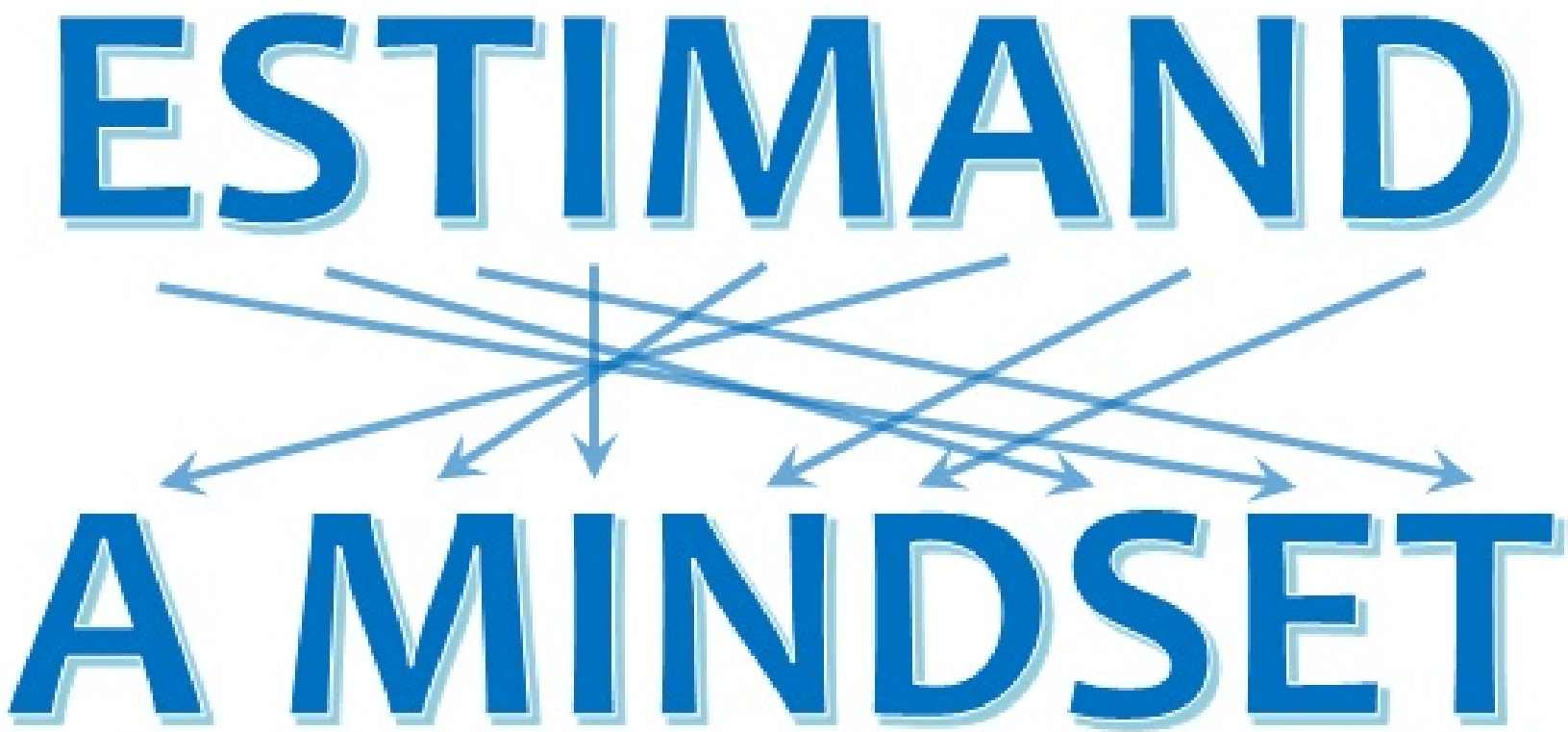
Webinar targeted at non-statisticians: `www.psiweb.org/vod/item/psi-eiwg-webinar-pioneering-estimands-in-clinical-research`.

X-industry working groups:

- **Oncology**: `www.oncoestimand.org`.
- **Neuroscience**: `www.psiweb.org/sigs-special-interest-groups/neuroscience-estimands`.

Putting together your estimand is a thinking process!

Putting together your estimand is a thinking process!



Thank you for your attention.

kaspar.rufibach@roche.com

http://www.kasparrufibach.ch

numbersman77

numbersman77

References I

- Altman, D. G., De Stavola, B. L., Love, S. B. and Stepniewska, K. A. (1995). Review of survival analyses published in cancer journals. *Br. J. Cancer* **72** 511–518.
- Bellera, C. A., Pulido, M. and Gourgou, S. e. a. (2013). Protocol of the Definition for the Assessment of Time-to-event Endpoints in CANcer trials (DATECAN) project: formal consensus method for the development of guidelines for standardised time-to-event endpoints' definitions in cancer clinical trials. *Eur. J. Cancer* **49** 769–781.
- Birgisson, H., Wallin, U., Holmberg, L. and Glimelius, B. (2011). Survival endpoints in colorectal cancer and the effect of second primary other cancer on disease free survival. *BMC Cancer* **11** 438.
- 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** 427–437.
- Hudis, C. A., Barlow, W. E., Costantino, J. P., Gray, R. J., Pritchard, K. I., Chapman, J. A., Sparano, J. A., Hunsberger, S., Enos, R. A., Gelber, R. D. and Zujewski, J. A. (2007). Proposal for standardized definitions for efficacy end points in adjuvant breast cancer trials: the STEEP system. *J. Clin. Oncol.* **25** 2127–2132.
- Johnson, J. R., Williams, G. and Pazdur, R. (2003). End points and United States Food and Drug Administration approval of oncology drugs. *J. Clin. Oncol.* **21** 1404–1411.
- 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** 4761–4771.
- Marcus, R., Davies, A., Ando, K., Klapper, W., Opat, S., Owen, C., Phillips, E., Sangha, R., Schlag, R., Seymour, J. F., Townsend, W., Trneny, M., Wenger, M., Fingerle-Rowson, G., Rufibach, K., Moore, T., Herold, M. and Hiddemann, W. (2017). Obinutuzumab for the First-Line Treatment of Follicular Lymphoma. *N. Engl. J. Med.* **377** 1331–1344.
- Mathoulin-Pelissier, S., Gourgou-Bourgade, S., Bonnetain, F. and Kramar, A. (2008). Survival end point reporting in randomized cancer clinical trials: a review of major journals. *J. Clin. Oncol.* **26** 3721–3726.
- Pazdur, R. (2008). Endpoints for assessing drug activity in clinical trials. *Oncologist* **13 Suppl 2** 19–21.

References II

- Peto, R., Pike, M. C., Armitage, P., Breslow, N. E., Cox, D. R., Howard, S. V., Mantel, N., McPherson, K., Peto, J. and Smith, P. G. (1977). Design and analysis of randomized clinical trials requiring prolonged observation of each patient. II. analysis and examples. *Br. J. Cancer* **35** 1–39.
- Sargent, D., Shi, Q., Yothers, G., Van Cutsem, E., Cassidy, J., Saltz, L., Wolmark, N., Bot, B., Grothey, A., Buyse, M., de Gramont, A., Sargent, D. J., Green, E., Grothey, A., Alberts, S. R., Bot, B., Campbell, M., Shi, Q., Yothers, G., O'Connell, M. J., Wolmark, N., de Gramont, A., Gray, R., Kerr, D., Haller, D. G., Benedetti, J., Buyse, M., Labianca, R., Seitz, J. F., O'Callaghan, C. J., Francini, G., Catalano, P. J., Blanke, C. D., Andre, T., Goldberg, R. M., Benson, A., Twelves, C., Cassidy, J., Sirzen, F., Cisar, L., Van Cutsem, E. and Saltz, L. (2011). Two or three year disease-free survival (DFS) as a primary end-point in stage III adjuvant colon cancer trials with fluoropyrimidines with or without oxaliplatin or irinotecan: data from 12,676 patients from MOSAIC, X-ACT, PETACC-3, C-06, C-07 and C89803. *Eur. J. Cancer* **47** 990–996.
- Van Cutsem, E., Labianca, R., Hossfeld, D. and et al (2005). Randomized phase iii trial comparing biweekly infusional fluorouracil/leucovorin alone or with irinotecan in the adjuvant treatment of stage iii colon cancer: Petacc-3. vol. 41st Annual Meeting of the American Society of Clinical Oncology.

Agenda

Backup

Estimand strategies

Six strategies for addressing intercurrent events

- **Composite strategies**, impacting the variable definition
- **Principal stratum strategies**, impacting the population definition
- **Treatment strategies**, impacting the treatment definition*

- **Treatment policy strategies**, disregarding ICEs
- **Hypothetical strategies**, assuming ICE hadn't happened
- **While on treatment strategies**, considering data until ICE

Doing now what patients need next

R version and packages used to generate these slides:

R version: R version 4.0.5 (2021-03-31)

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

Other packages:

This document was generated on 2021-05-19 at 17:05:23.