



# Faculty of Health Sciences

## Day 3 <sub>(part 2)</sub>: Adherence, Intention to treat & Per protocol analyses, estimand framework and Statistical Analysis Plan

Paul Blanche

Section of Biostatistics, University of Copenhagen

# Outline/Intended Learning Outcomes (ILOs)

## Adherence, Intention to treat & Per protocol analyses

ILO: exemplify the challenges caused by nonadherence

ILO: restate and contrast different approaches to analyze data with nonadherence

ILO: explain their strengths and limitations for both superiority and non-inferiority studies

## Estimand framework

ILO: restate and exemplify the benefit of using the estimand framework

ILO: restate and exemplify the core attributes of estimands

ILO: recognize the challenges posed by intercurrent events and discuss different strategies to handle them

## Statistical Analysis Plan (SAP)

ILO: restate the rationale for SAPs and their content

ILO: recall examples of SAPs and where to find resources SAPs



# Nonadherence

**Nonadherence** in a randomized clinical trial (RCT) occurs when study participants do not follow the randomly assigned treatment protocol. Reasons for nonadherence may include the study participant not taking trial medications, crossing over to the other intervention being studied, or accessing treatment outside of the trial. Nonadherence also may occur when the clinician is unable to complete the assigned therapy (eg, a surgical procedure) as intended.

Source: Smith et al. "Interpreting the results of intention-to-treat, per-protocol, and as-treated analyses of clinical trials." JAMA 326.5 (2021): 433-434.

- ▶ Nonadherence is often **difficult to avoid**.
- ▶ **Nonadherence should be documented in the clinical study report.**

*"The frequency and type of protocol violations, missing values, and other problems should be documented in the clinical study report and their potential influence on the trial results should be described (see ICH E3)." <sup>1</sup>*

<sup>1</sup> [EMA scientific guidelines: ICH Topic E9 Statistical Principles for Clinical Trials, 1998, https://www.ema.europa.eu/en/documents/scientific-guideline/ich-e-9-statistical-principles-clinical-trials-step-5\\_en.pdf](https://www.ema.europa.eu/en/documents/scientific-guideline/ich-e-9-statistical-principles-clinical-trials-step-5_en.pdf)



# Analysis sets & analyses approaches

*"Intention-to-treat (ITT), per- protocol (PP), and as-treated (AT) approaches to analysis differ in how the included patient population and treatment assignments are defined, with important implications for interpretation of treatment effects in clinical trials."*

- ▶ "With the ITT approach, **all randomized patients** are included in the analysis, based on the groups to which they were initially randomly assigned."
- ▶ "Per-protocol only analyzes data from **participants who follow the protocol**, excluding their data after they become nonadherent."
- ▶ "AT analyses **consider the treatment actually received by the participant**, without regard to adherence to their randomization assignment."

Source: Smith et al. "Interpreting the results of intention-to-treat, per-protocol, and as-treated analyses of clinical trials." JAMA 326.5 (2021): 433-434.



# Intention-to-treat (ITT) principle

*"The intention-to-treat (see Glossary) 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. In practice this ideal may be difficult to achieve, for reasons to be described. In this document the term 'full analysis set' is used to describe the analysis set which is as complete as possible and as close as possible to the intention-to-treat ideal of including all randomised subjects. Preservation of the initial randomisation in analysis is important in preventing bias and in providing a secure foundation for statistical tests. In many clinical trials the use of the full analysis set provides a conservative strategy."<sup>2</sup>*

---

<sup>2</sup> EMA scientific guidelines: ICH Topic E 9 Statistical Principles for Clinical Trials, 1998, [https://www.ema.europa.eu/en/documents/scientific-guideline/ich-e-9-statistical-principles-clinical-trials-step-5\\_en.pdf](https://www.ema.europa.eu/en/documents/scientific-guideline/ich-e-9-statistical-principles-clinical-trials-step-5_en.pdf)



# Appendix: ITT principle, other quotes

## Box 6. Intention-to-treat analysis

The special strength of the RCT is the avoidance of bias when allocating interventions to trial participants (see Box 1). That strength allows strong inferences about cause and effect that are not justified with other study designs. In order to preserve fully the huge benefit of randomisation we should include all randomised participants in the analysis, all retained in the group to which they were allocated. Those two conditions define an “intention-to-treat” analysis, which is widely recommended as the preferred analysis strategy [18,223]. Intention-to-treat analysis corresponds to analysing the groups exactly as randomised. Strict intention-to-treat analysis is often hard to achieve for two main reasons—missing outcomes for some participants and non-adherence to the trial protocol.

Source: Moher et al (2010). CONSORT 2010 Explanation and Elaboration: updated guidelines for reporting parallel group randomised trials. *Journal of Clinical Epidemiology*, 63(8), e1–e37

**Principle 4.2. Intention-to-treat (ITT)** *The primary comparison in a clinical trial should be of treatments to which patients were randomized, not of treatments actually received.*

Another way to phrase the ITT principle is that one should analyze according to how one intended to treat the patients regardless of whether they received treatment, hence the name, intention-to-treat.

**Remark 4.3.** *Two important consequences of the ITT principle are:*

1. *One should never stop following patients who do not comply with their randomized treatment.*
2. *Missing data threaten the integrity of a clinical trial by making it difficult to compare the original randomized groups.*

The ITT principle is a tough pill to swallow, so to speak. For instance, the ITT principle



# Why is ITT (often, but not always!) conservative?

- ▶ “We start by considering **placebo-controlled double-blind RCTs**. It is well known that if treatment A has a null effect on the outcome, then both the effect of assigned treatment Z and the effect of treatment A will be null.”
- ▶ “This is a key advantage of the ITT analysis: it correctly estimates the effect of treatment A under the null, regardless of the adherence pattern.”
- ▶ “It is also well known that if treatment A has a **non-null effect** (that is, either increases or decreases the risk of the outcome) and some participants do not adhere to their assigned treatment, then the **effect of assigned treatment Z will be closer to the null than the actual effect of treatment A.**”
- ▶ “This bias toward the null is due to **contamination** of the treatment groups: some subjects assigned to treatment ( $Z=1$ ) may not take it ( $A=0$ ) whereas some subjects assigned to placebo ( $Z=0$ ) may find a way to take treatment ( $A=1$ ). As long as the proportion of patients who end up taking treatment ( $A=1$ ) is greater in the group assigned to treatment ( $Z=1$ ) than in the group assigned to placebo ( $Z=0$ ), the **effect of assigned treatment Z will be in between the effect of treatment A and the null value.**”

Source: Hernán & Hernández-Díaz. Beyond the intention-to-treat in comparative effectiveness research. Clinical trials 9.1 (2012): 48-



## Example & Exercise 3.1. use R!

In Exercise 1.1 and 1.2, we have seen that we could include 1048 patients to have a power of 80%. The investigator expected 11% and 17% because they expected 40% of the patients to be in septic shock, with risk 30% vs 20%; and 60% not in septic shock, with risk 7.5% vs 5%.

In fact, the above expectations were under perfect adherence. But unfortunately, the investigator suspected that up to 10% of the patients in septic shock randomized to the new treatment will likely end up not receiving it, for unfortunate logistical reasons. Let's assume that for these patients, the risk is the same as in the control group (30%) and that for those who receive it the risk is the same as under perfect adherence (20%). Further assume that we plan to perform an ITT analysis.

### Questions:

1. What is the risk we can expect to estimate, in average, in the experimental treatment arm, with an ITT analysis (instead of 11%)?
2. What does the power become, if we indeed expect this risk instead of 11% (versus 17%)?
3. By how much could we inflate the sample size to compensate for nonadherence and still have 80% power?
4. Remember, we expect to be able to include 130 patients per year. By how long should we prolong the accrual period, and consequently, the trial duration?



# When is ITT NOT conservative? Why? (1/2)

## ITT is not conservative in non-inferiority trials.

- ▶ Contamination will usually make the two arms more similar than they should. Hence, an ITT analysis will produce bias towards “no difference between the two arms” and a treatment effect biased towards zero. In non-inferiority trials, this makes it is “easier” to conclude that the treatment effect is below the non-inferiority margin; and it facilitates concluding non-inferiority. This increases the risk of falsely concluding non-inferiority (i.e., the risk of type-I error might be inflated). See e.g., Scenario 1 in Hernán & Hernández-Díaz. Clinical trials 9.1 (2012): 48-55.

## Hence ITT analyses should be complemented.

- ▶ *“In a superiority trial the full analysis set, based on the ITT (intention-to-treat) principle, is the analysis set of choice, with appropriate support provided by the PP (per protocol) analysis set. In a non-inferiority trial, the full analysis set and the PP analysis set have equal importance and their use should lead to similar conclusions for a robust interpretation.”<sup>3</sup>*



<sup>3</sup> EMA scientific guidelines: Points to Consider on switching between superiority and non-inferiority, 200, [https://www.ema.europa.eu/en/documents/scientific-guideline/points-consider-switching-between-superiority-and-non-inferiority\\_en.pdf](https://www.ema.europa.eu/en/documents/scientific-guideline/points-consider-switching-between-superiority-and-non-inferiority_en.pdf)

## Example & Exercise 3.2: read and think

Now let us consider double-blind RCTs that compare two active treatments. These trials are often designed to show that a new treatment ( $A = 1$ ) is not inferior to a reference treatment ( $A = 0$ ) in terms of either benefits or harms. An example of a noninferiority trial would be one that compares the reduction in blood glucose between a new inhaled insulin and regular injectable insulin. The protocol of the trial would specify a noninferiority margin, that is, the maximum average difference in blood glucose that is considered equivalent (e.g., 10 mg/dL). Using an ITT comparison, the new insulin ( $A = 1$ ) will be declared not inferior to classical insulin ( $A = 0$ ) if the average reduction in blood glucose in the group assigned to the new treatment ( $Z = 1$ ) is within 10 mg/dL of the average reduction in blood glucose in the group assigned to the reference treatment ( $Z = 0$ ) plus/minus random variability. Such ITT analysis may be

### Scenario 1

The new treatment  $A = 1$  is actually inferior to the reference treatment  $A = 0$ , for example, the average reduction in blood glucose is 10 mg/dL under treatment  $A = 1$  and 22 mg/dL under treatment  $A = 0$ . The type and magnitude of adherence is equal in the two groups, for example 30% of subjects in each group decided not to take insulin. As a result, the average reduction is, say, 7 mg/dL in the group assigned to the new treatment ( $Z = 1$ ) and 15 mg/dL in the group assigned to the reference treatment ( $Z = 0$ ). An ITT analysis, which is biased toward the null in this scenario, may incorrectly suggest that the new treatment  $A = 1$  is not inferior to the reference treatment  $A = 0$ .

Source: Hernán & Hernández-Díaz. Clinical trials 9.1 (2012): 48-55.

### Questions:

1. Why 7 and 15 mg/dL?
2. What are the expected values of the mean difference in reduction under perfect adherence versus non-adherence, in this example, when using an ITT analysis?
3. What is the expected sign of the  $E(Z)$  in each case? What can we conclude on the type-I error? **Help:** type-I error =  $2\Phi\{E(Z) - 1.96\}$  for a usual two-sided test at the desired 5% level and  $\Phi(-1.96) = 2.5\%$ , with  $\Phi(z) = P(Z \leq z)$  when  $Z$  is a standard normal random variable.
4. Assume that the trial was planned to have a power of 80% under the assumption that the treatment were actually equally efficacious. Can we compute the type-I error of the ITT analysis? **Help:**  $E(Z) = 2.8$  under the alternative for which the trial is powered and  $E(Z) = (\delta - \delta_0)/SE$ . Also, if the treatments are equally efficacious, then  $\delta = 0$ .



# When is ITT NOT conservative? Why? (2/2)

## Other (rare/specific) cases exist:

*"Finally, if neither treatment has any effect on the outcome, the ITT is generally considered unbiased. However, there is the potential for bias away from the null compared with the average causal effect of treatment when 1) a treatment is beneficial for some participants and harmful for others, and 2) the proportion of adherence is related to the outcome, such as when some of the side effects causing nonadherence are also markers for the outcome."*<sup>4</sup>

---

<sup>4</sup>Shrier et al. *The intention-to-treat analysis is not always the conservative approach.* The American journal of medicine 130.7 (2017): 867-871.



# ITT analyses can be very conservative

**3. An excessive conservatism.** Imagine a double-blind crossover placebo-controlled study of a new drug for the prophylaxis of angina. Fig. 2 shows both the placebo and active drug data on one hypothetical patient. Here the intended dose of the new drug was one per day, and if the patient had taken the new drug as intended, there would be no reason to think it was superior to placebo on the basis of the similar frequency of anginal attacks on both treatments.

But what if a medication monitor had been used,<sup>4</sup> and the pattern of dosage was the one listed as *Actual* in Fig. 2? It seems the patient did not take any drug at all for the first 2 days and had anginal attacks on days 2 and 3. He then took the drug for 3 days without angina, stopped taking the drug, and had an attack the next day. The same thing happened again after 1 week of full compliance without angina. Surely, if there were multiple records like this one and the placebo records did not show similar patterns, we would be justified in concluding that the drug is effective, would we not? Apparently not: Many statisticians would tell us that we must do something called an “intention to treat” analysis. In this case such an analysis

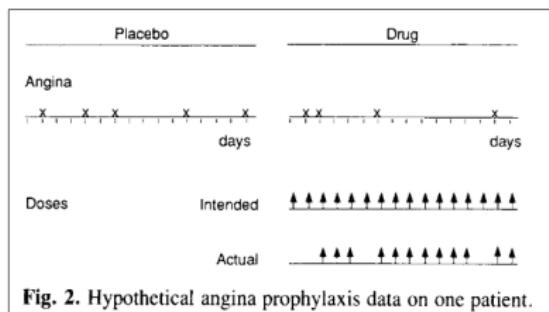


Fig. 2. Hypothetical angina prophylaxis data on one patient.

Source: Sheiner, L.B. (1991). The intellectual health of clinical drug evaluation. Clinical Pharmacology and Therapeutics 50 (1): 4–9



# ITT analysis: effectiveness instead of efficacy?

## Definitions:

**Effectiveness** is usually defined as 'how well a treatment works in everyday practice,' and **efficacy** as 'how well a treatment works under perfect adherence and highly controlled conditions.'<sup>5</sup>

Does ITT analysis evaluate effectiveness instead of efficacy? Not necessarily.

- ▶ "Arguments in favor of the ITT principle include [...] (2) analysis by randomized groups irrespective of compliance tells what will happen in the real world if we attempt to treat patients [i.e., it estimates **effectiveness**] [...] Opponents of the ITT principle counter the second point above with the argument that compliance during the trial and outside the trial are two different things. If the trial shows that the intervention works, that will likely lead to better compliance in the real world."<sup>6</sup>
- ▶ "The result of an intention-to-treat analysis is affected by the trial-specific pattern of adherence to the treatment strategies under study and therefore may not be directly relevant for guiding decisions in clinical settings with different adherence patterns. In fact, the publication and dissemination of the intention-to-treat result may change adherence in clinical settings, possibly rendering the result outdated."<sup>7</sup>

<sup>5</sup> Hernán & Hernández-Díaz. Beyond the intention-to-treat in comparative effectiveness research. Clinical trials 9.1 (2012): 48-55.

<sup>6</sup> page 36, in *Statistical Thinking in Clinical Trials*, by Michael Proschan (2022)

<sup>7</sup> Hernán & M. Robins. Per-protocol analyses of pragmatic trials. N Engl J Med 377.14 (2017): 1391-1398.



# Per-Protocol analysis: what and why?

## ► What?

*"Study participants who do not take the assigned therapy are excluded in the analyses and may also be censored from the analyses when treatment is discontinued. PP analyses can also be defined where study participants who do not adhere to protocol (e.g., schedule of assigned therapy) or have protocol violations (e.g., took prohibited medications or the wrong therapy) might be excluded from the analyses. [...] The exact definition of the PP analysis set is not unique and frequently varies depending on the goals of the analyses.*

## ► Why?

*The PP analyses may be viewed as an evaluation of the treatment under near ideal use of the therapy* (e.g., good adherence), and thus may be informative when adherence can be influenced by clinician intervention. A PP analyses can intuitively be an attractive alternative as it is more consistent with the theory that "treatment not taken cannot affect the outcome" than ITT.<sup>8</sup>

## ► Common recommendation:

*"In general, it is advantageous to demonstrate a lack of sensitivity of the principal trial results to alternative choices of the set of subjects analysed. In confirmatory trials it is usually appropriate to plan to conduct both an analysis of the full analysis set and a per protocol analysis, so that any differences between them can be the subject of explicit discussion and interpretation."*<sup>9</sup>

---

<sup>8</sup>

Source: page 2028, in *Fundamental Concepts for New Clinical Trialists*, Scott Evans & Naitee Ting (2015).

<sup>9</sup> **EMA scientific guidelines:** ICH Topic E 9 Statistical Principles for Clinical Trials, 1998, [https://www.ema.europa.eu/en/documents/scientific-guideline/ich-e-9-statistical-principles-clinical-trials-step-5\\_en.pdf](https://www.ema.europa.eu/en/documents/scientific-guideline/ich-e-9-statistical-principles-clinical-trials-step-5_en.pdf)



# PP and AT analyses: limitations

## Randomization is broken, biases can arise:

*"As-treated and PP analyses are not simple to interpret because of the potential loss of an important benefit of randomization: the elimination of systematic bias in treatment assignment. Selection bias [in PP analysis] and confounding [in a AT analysis] in the treatment effect estimate arises if patients who are more adherent with the assigned treatment differ in ways that also influence their outcomes compared with those who are less adherent."<sup>10</sup>*

## Adjustment for confounding might be needed:

*"In summary, 'as treated' and 'per protocol' analyses transform RCTs into observational studies for all practical purposes. The estimates from these analyses can only be interpreted as the effect of treatment A if the analysis is appropriately adjusted for the confounders L. If the intended analysis of the RCT is 'as treated' or 'per protocol,' then the protocol of the trial should describe the potential confounders and how they will be measured, just like the protocol of an observational study would do."<sup>11</sup>*

- ▶ Simple baseline covariate adjustments might be sufficient in some case, but not always.
- ▶ Sensitivity analyses may be needed.
- ▶ Advanced methods to deal with time-dependent confounding might be necessary. In that case, to properly adjust for confounding "it will be necessary to collect and analyze postrandomization data on adherence to the assigned treatment strategies and on the evolution of prognostic factors."<sup>12</sup>

---

<sup>10</sup> Smith et al. JAMA 326.5 (2021): 433-434.

<sup>11</sup> Hernán & Hernández-Díaz. Clinical trials 9.1 (2012): 48-55.

<sup>12</sup> Hernán & M. Robins. Per-protocol analyses of pragmatic trials. N Engl J Med 377.14 (2017): 1391-1398.



# Outline/Intended Learning Outcomes (ILOs)

## Adherence, Intention to treat & Per protocol analyses

ILO: exemplify the challenges caused by nonadherence

ILO: restate and contrast different approaches to analyze data with nonadherence

ILO: explain their strengths and limitations for both superiority and non-inferiority studies

## Estimand framework

ILO: restate and exemplify the benefit of using the estimand framework

ILO: restate and exemplify the core attributes of estimands

ILO: recognize the challenges posed by intercurrent events and discuss different strategies to handle them

## Statistical Analysis Plan (SAP)

ILO: restate the rationale for SAPs and their content

ILO: recall examples of SAPs and where to find resources SAPs



# The need for estimands

Received: 25 July 2019 | Revised: 11 October 2019 | Accepted: 20 October 2019  
DOI: 10.3351/bcp.14195

**EMA GUIDELINES**

 EMA European Medicines Agency

Beyond "Intent-to-treat" and "Per protocol": Improving assessment of treatment effects in clinical trials through the specification of an estimand

Surprisingly, even a careful examination of the trial protocol or of the methods section in a scientific publication is often insufficient to find fundamental information: what exact research question does the trial seek to answer? This can be articulated into several subquestions that are not answered explicitly in many trial protocols or scientific publications: what is the medical outcome of interest and when and how is this measured for each individual patient? To which patients should the trial results exactly apply: to all those who could have entered the trial or all those who could tolerate the treatment long enough? Which effect is of interest in a setting where patients can, eg, switch to other treatments or have access to rescue therapy? Is it the effect of starting the new treatment, irrespective of subsequent treatment strategies? Is the study designed and is data collection planned based on these choices (see Section A.4. Impact on Trial Design and Conduct)? Stating the targeted analysis set or principle, eg intention to treat or per protocol, does not fully capture these choices and their consequences for design, data collection and analysis (see Section A.5. Impact on trial analysis). The way trial objectives are usu-

Source: Pétavy et al. British Journal of Clinical Pharmacology 86.7 (2020): 1235-1239.

## What's an estimand? (short answer)

**Estimand.** A term used to cover a quantity or probability one would like to estimate through analysis. Thus, an estimand is some ways analogous to a parameter in a model whereas an estimator is a statistic of which an estimate is a given value. However, in some circumstances it may be that the estimand has not been expressed in terms of a parameter and it may be better to think of it in terms of a goal or end with the estimator as the means of approaching it.

Source: page 581, in *Statistical Issues in Drug Development* (3<sup>rd</sup> Edition), by Stephen Senn.

*"Estimands are rapidly becoming an essential part of medical research. They serve to remind us of this important truth: the research question should drive our study methods, and not the other way around."* (Kahan et al. BMJ 2024;384:q173.)





17 February 2020  
EMA/CHMP/ICH/436221/2017  
Committee for Medicinal Products for Human Use

## ICH E9 (R1) addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials

Step 5

•  
•

### Glossary

#### **Estimand:**

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

#### **Estimate:**

A numerical value computed by an estimator.

#### **Estimator:**

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

#### **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. It is necessary to address intercurrent events when describing the clinical question of interest in order to precisely define the treatment effect that is to be estimated.

(EMA scientific guidelines:

<https://www.ema.europa.eu/en/ich-e9-statistical-principles-clinical-trials-scientific-guideline>)

- ▶ About 19 pages guidelines, first draft released in 2017 (for comments). Updated in 2019, after receiving many constructive comments. See e.g., Scharfstein. Clinical Trials 16.4 (2019): 375-380.
- ▶ Numerous papers to describe and discuss it (mostly positively, but not always).
- ▶ A little revolution in statistical analyses in the pharmaceutical industry (coming to academia, with delay).
- ▶ **Main focus is how to deal with "Intercurrent Events" and consequently better defining the **treatment regimens** that we compare.**
- ▶ (Probably overly simplified) One could say that the estimand framework should (1) help us to clarify and communicate what we would actually estimate with a "standard" per-protocol analysis, intention to treat analysis (or other) and (2) once we understand that, help us to either modify the analysis as relevant to get closer to what we really want (e.g., by changing analysis sets and statistical methods used for analysis) or to update the research question, to the best match with a doable analysis.
- ▶ Fundamental ideas trace back to earlier important discussions about how to handle missing data (National Research Council. The prevention and treatment of missing data in clinical trials. Washington, DC: National Academies Press, 2010 [http://www.nap.edu/catalog.php?record\\_id=12955](http://www.nap.edu/catalog.php?record_id=12955).)

# Example of Intercurrent Events

## Box 1: Importance of intercurrent events

### Example 1

In a study of dupilumab versus placebo for uncontrolled asthma, patients in the placebo arm might receive rescue treatment more often than patients in the dupilumab arm.<sup>6</sup> Where does interest lie: in the effect of dupilumab versus placebo when rescue forms part of the two treatment strategies, or in the effect of dupilumab if patients had not received rescue?

### Example 2

In a study comparing two different fluid delivery methods in patients undergoing emergency bowel surgery, patients could have their surgery cancelled after enrolment.<sup>1</sup> Do researchers want to compare the two fluid delivery methods only in those patients who actually undergo surgery, or in all patients regardless of whether they undergo surgery?

### Example 3

In a study evaluating a music intervention delivered by caregivers for people with dementia on symptom reduction at 90 days, some participants could die before day 90.<sup>7</sup> Should researchers use their final symptom score before they died to evaluate the intervention effect while they still lived, or assign their 90 day score a low value, to reflect that death is a poor outcome?

### Example 4

In a study of triamcinolone versus usual care in patients undergoing eye surgery, some patients might take additional non-study treatments.<sup>8</sup> Should researchers evaluate the effect of triamcinolone alongside these additional non-study treatments, or its effect if patients had not taken any additional treatments?

Source: Kahan, Brennan C., et al. "The estimands framework: a primer on the ICH E9 (R1) addendum." BMJ 384 (2024).



# Estimands and research question

## Box 2: How estimands can clarify research questions

- It is important to understand which type of treatment effect a study sets out to estimate. Historically, two types of studies have been considered<sup>9</sup>: pragmatic studies that seek to estimate an intervention's real world effect, and explanatory studies that seek to estimate an intervention's effect under ideal conditions.
- However, these two paradigms are not sufficient to precisely define the exact research question, because within these broad definitions exist multiple versions of a pragmatic or explanatory effect that could be estimated. Thus, international guidelines have called for greater clarity.<sup>5</sup>
- Estimands extend the commonly used PICO (population, intervention, comparator, outcome) framework for defining research questions by adding two additional attributes: the summary measure, which defines how outcomes are summarised and compared between treatments; and the strategies used to handle each type of intercurrent event, which define how things such as treatment switching or treatment discontinuation are handled in the treatment effect definition.
- Estimands are now required in some reporting guidelines,<sup>10-12</sup> and medicine regulators in Europe, US, Canada, Singapore, China, Switzerland, and Chinese Taipei now require regulatory applications to include estimands, while regulators in Brazil, the Republic of Korea, and Japan are currently in the process of implementing the inclusion of estimands.<sup>13</sup>



# Estimands and research question

## Box 2: How estimands can clarify research questions

- It is important to understand which type of treatment effect a study sets out to estimate. Historically, two types of studies have been considered<sup>2</sup>: pragmatic studies that seek to estimate an intervention's real world effect, and explanatory studies that seek to estimate an intervention's effect under ideal conditions.
- However, these two paradigms are not sufficient to precisely define the exact research question, because within these broad definitions exist multiple versions of a pragmatic or explanatory effect that could be estimated. Thus, international guidelines have called for greater clarity.<sup>5</sup>
- Estimands extend the commonly used PICO (population, intervention, comparator, outcome) framework for defining research questions by adding two additional attributes: the summary measure, which defines how outcomes are summarised and compared between treatments; and the strategies used to handle each type of intercurrent event, which define how things such as treatment switching or treatment discontinuation are handled in the treatment effect definition.
- Estimands are now required in some reporting guidelines,<sup>10-12</sup> and medicine regulators in Europe, US, Canada, Singapore, China, Switzerland, and Chinese Taipei now require regulatory applications to include estimands, while regulators in Brazil, the Republic of Korea, and Japan are currently in the process of implementing the inclusion of estimands.<sup>13</sup>

**Table 1 | Example of how estimands can help researchers understand the research question**

Study description	Statistical methods	Problems understanding the research question	How estimands explain the research question
A trial compared dupilumab with placebo on forced expiratory volume (FEV <sub>1</sub> ) at week 12 in patients with uncontrolled persistent asthma. Some patients stopped dupilumab early or received rescue treatments for exacerbations.	Data were analysed on an intention-to-treat basis. Outcome data after receipt of rescue treatment or discontinuation of dupilumab was treated as missing, and a mixed model for repeated measures was used to estimate the treatment effect.*	Because the statistical methods do not make explicit how the research question handles early stopping of dupilumab or receipt of rescue treatment, readers must infer this.* Since the analysis was by intention to treat, they might incorrectly assume that interest lies in the effect of dupilumab regardless of the early stopping or use of rescue treatment.	The estimand explicitly describes how early stopping and receipt of rescue treatment are handled in the research question: "The estimand is the difference in the mean FEV <sub>1</sub> at week 12 between dupilumab plus standard of care versus placebo plus standard of care, in patients with uncontrolled persistent asthma, if they were to continue using dupilumab over the entire trial period without the use of rescue treatment."

\*In this setting, the mixed model for repeated measures estimates dupilumab's hypothetical effect if patients were to continue taking dupilumab and did not receive rescue treatment, because investigators treated outcome data after receipt of rescue treatment or discontinuation of dupilumab as missing. Here, the mixed model served to implicitly impute what the outcome data would have been had participants not received rescue treatment or discontinued. Here, deciphering the research question requires an in-depth understanding of the mechanics underlying mixed models for repeated measures, which not all readers will have.

Source: Kahan, Brennan C., et al. "The estimands framework: a primer on the ICH E9 (R1) addendum." BMJ 384 (2024).



# Digression: explanatory versus pragmatic trials

- The estimand framework relates to earlier discussions about **explanatory** versus **pragmatic** trials.

## 1. Definition of the Treatments

### 1.1. "Equalized" or "optimal" conditions

Consider a trial of anti-cancer treatments in which radiotherapy alone is to be compared with radiotherapy preceded by the administration of a drug which has no effect by itself but which may sensitise the patient to the effects of radiation. Suppose the drug is to be administered over a 30-day period. The "radiotherapy alone" group may then be handled in two different ways (Fig. 1):

- (a) radiotherapy may be preceded by a blank period of 30 days, so that it is instituted at the same time in each group;
- (b) radiotherapy may be instituted at once, thereby carrying it out at what is most probably the optimal time.

Neither procedure can be said to be "better" than the other. The first allows us to compare two groups which are alike from the radiotherapy point of view and which differ solely in the presence or absence of the drug. It therefore provides an assessment of the sensitising effect of the drug and gives valuable information at a **biological level**. The second procedure enables us to compare two treatments under the conditions in which they would be applied in practice. We distinguish the two procedures as stemming from two different approaches to the trial, the first **explanatory**, the second **pragmatic**.

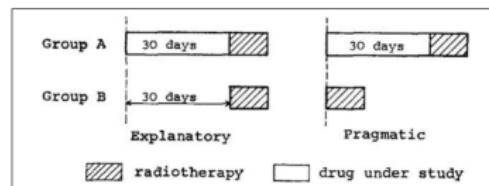
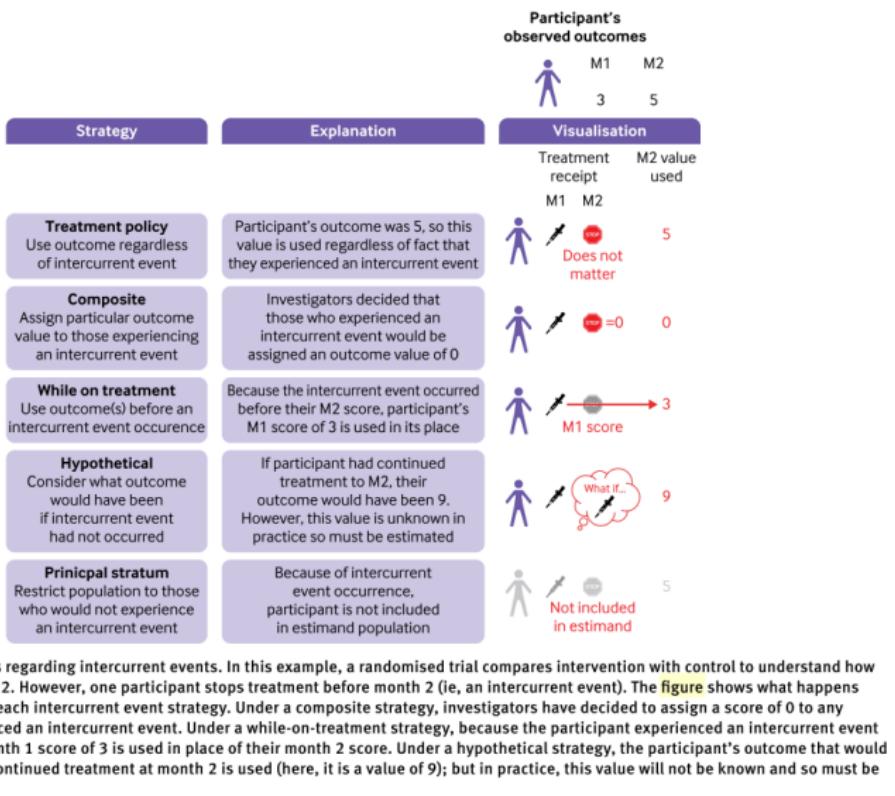


Fig. 1. Explanatory and pragmatic approaches in the first example.





**Fig 1 |** Different strategies regarding intercurrent events. In this example, a randomised trial compares intervention with control to understand how outcomes differ at month 2. However, one participant stops treatment before month 2 (ie, an intercurrent event). The figure shows what happens to this participant under each intercurrent event strategy. Under a composite strategy, investigators have decided to assign a score of 0 to any participant who experienced an intercurrent event. Under a while-on-treatment strategy, because the participant experienced an intercurrent event before month 2, their month 1 score of 3 is used in place of their month 2 score. Under a hypothetical strategy, the participant's outcome that would have occurred had they continued treatment at month 2 is used (here, it is a value of 9); but in practice, this value will not be known and so must be estimated. M=month

Source: Kahan, Brennan C., et al. "The estimands framework: a primer on the ICH E9 (R1) addendum." BMJ 384 (2024).



## General considerations:

### Box 5: Intercurrent events, protocol deviations, and missing data

The definition of an intercurrent event is broad, encompassing several distinct concepts (eg, treatment-modifying events, truncating events). Owing to some overlap with other common concepts, understanding what is (and what is not) an intercurrent event can be challenging. We summarise below how intercurrent events differ from protocol deviations and missing data.

#### *Protocol deviations*

Some but not all protocol deviations can also be intercurrent events. Intercurrent event status depends on whether the protocol deviation affects assigned treatment. If it does affect assigned treatment (eg, receipt of prohibited drug treatment), the deviation is also an intercurrent event; if it does not (eg, failure to take proper informed consent), the deviation usually is not an intercurrent event.

Similarly, some but not all intercurrent events can also be protocol deviations. Protocol deviation status will depend on whether the intercurrent event is allowed by the protocol. For instance, if the protocol allows patients to modify or stop treatment in response to an adverse event, this event is not a deviation. However, if a participant receives drug treatment prohibited by the protocol, this event is a deviation.

#### *Missing data*

Loss to follow-up, study withdrawal, and missing data frequently occur alongside certain intercurrent events, but they are not themselves intercurrent events.<sup>5</sup> For instance, participants who stop treatment early might also withdraw from the study. However, it is the treatment discontinuation that affects our interpretation of outcome data, and not the withdrawal from the study (which simply poses a missing data issue that needs to be handled as part of the statistical analysis, but not as part of the estimand definition).

Source: Kahan, Brennan C., et al. "The estimands framework: a primer on the ICH E9 (R1) addendum." BMJ 384 (2024).



## General considerations:

### Box 5: Intercurrent events, protocol deviations, and missing data

The definition of an intercurrent event is broad, encompassing several distinct concepts (eg, treatment-modifying events, truncating events). Owing to some overlap with other common concepts, understanding what is (and what is not) an intercurrent event can be challenging. We summarise below how intercurrent events differ from protocol deviations and missing data.

#### Protocol deviations

Some but not all protocol deviations can also be intercurrent events. Intercurrent event status depends on whether the protocol deviation affects assigned treatment. If it does affect assigned treatment (eg, receipt of prohibited drug treatment), the deviation is also an intercurrent event; if it does not (eg, failure to take proper informed consent), the deviation usually is not an intercurrent event.

Similarly, some but not all intercurrent events can also be protocol deviations. Protocol deviation status will depend on whether the intercurrent event is allowed by the protocol. For instance, if the protocol allows patients to modify or stop treatment in response to an adverse event, this event is not a deviation. However, if a participant receives drug treatment prohibited by the protocol, this event is a deviation.

#### Missing data

Loss to follow-up, study withdrawal, and missing data frequently occur alongside certain intercurrent events, but they are not themselves intercurrent events.<sup>5</sup> For instance, participants who stop treatment early might also withdraw from the study. However, it is the treatment discontinuation that affects our interpretation of outcome data, and not the withdrawal from the study (which simply poses a missing data issue that needs to be handled as part of the statistical analysis, but not as part of the estimand definition).

Source: Kahan, Brennan C., et al. "The estimands framework: a primer on the ICH E9 (R1) addendum." BMJ 384 (2024).

## A specific remark about non-inferiority studies:

*"We argue the main threat to validity of NI trials is the occurrence of 'trial-specific' intercurrent events (IEs), that is, IEs which occur in a trial setting, but would not occur in practice."* (Morgan et al. Statistics in Medicine 44.5 (2025): e10348.)



# Core attributes of estimands

**Table 3 | Core attributes of estimands**

Attribute	Definition	Example from the FLO-ELA trial <sup>1</sup>
Population	Patients for whom researchers want to estimate the treatment effect	Patients >50 years old who would undergo emergency bowel surgery under any treatment assignment
Treatment conditions	Different intervention strategies being compared in the treatment effect definition	Intervention group: assignment to protocolised, cardiac output guided, haemodynamic treatment during surgery and for six hours after, regardless of whether cardiac output monitor is followed correctly; usual care group: assignment to intravenous fluid use without cardiac output monitoring or protocol during surgery, and for six hours after
Endpoint	Outcome for each participant that is used in the treatment effect definition	Number of days alive and out of hospital within 90 days of randomisation
Summary measure	Method used to summarise and compare the endpoint between treatment conditions (eg, risk ratio, odds ratio)	Ratio of means
Handling of intercurrent events	Strategies used to handle each intercurrent event* in the treatment effect definition; different strategies could be used for different types of intercurrent events	Surgery cancelled after randomisation (applies to both treatment groups); principal stratum (subpopulation of patients who would undergo surgery under either treatment assignment); receipt of cardiac output monitoring (usual care group); treatment policy; failure to initiate cardiac output monitoring (intervention group); treatment policy; cardiac output monitoring algorithm not followed (intervention group); treatment policy

\*Intercurrent events are post-baseline events (or post-randomisation events in randomised trials) that affect the interpretation or existence of outcome data. These events frequently affect receipt of treatment (eg, treatment switching or treatment discontinuation) or preclude existence of the outcome (eg, death, if it is not defined as part of the outcome).

Source: Kahan, Brennan C., et al. "The estimands framework: a primer on the ICH E9 (R1) addendum." BMJ 384 (2024).



### Example & Exercise 3.3: Read and think (if time allows)

1. Read the interesting short paper of Keene et al. BMC medicine 21.1 (2023): 276. <https://doi.org/10.1186/s12916-023-02969-6>
2. Present and discuss the two examples (SYNAPSE and PIONEER-1). Especially, do they illustrate well the benefit of using the estimand framework?

### Example & discussion: students case studies

#### Example from SAP of Nielsen et al. (2024).

[https://cdn.clinicaltrials.gov/large-docs/25/NCT04563325/SAP\\_000.pdf](https://cdn.clinicaltrials.gov/large-docs/25/NCT04563325/SAP_000.pdf)



# Still new...

*"The centrality of estimands to clinical trials is currently not reflected in methods recommended by the Cochrane group or the CONSORT statement, the current standard for reporting clinical trials in medical journals. We encourage revisions to these guidelines."<sup>13</sup>*

- ▶ To the best of my knowledge, this is still the case.
- ▶ The estimand framework is still new. Most of us are still learning how to best use it and write about it in protocols, SAPs and papers.



# Estimand vs Estimator

- ▶ Estimand framework is about what we would like to estimate.
- ▶ It is not about how to estimate what we would like (at least not directly).
- ▶ For some estimand choices, it might be challenging to come up with a valid statistical analysis (unbiased) that do not rely on questionable assumptions.
- ▶ Unmeasured confounding and/or questionable non-testable modeling or structural assumptions might be needed.
- ▶ Sensitivity analyses are very much encouraged when questionable assumptions are used.
- ▶ Some statistical methods for handling missing data can be used with some estimands. Interesting literature reviews about methods to handle missing data in clinical trials exist.<sup>14</sup>

---

<sup>14</sup> See e.g., National Research Council. 2010. *The Prevention and Treatment of Missing Data in Clinical Trials*. Washington, DC: The National Academies Press. <https://doi.org/10.17226/12955>



# Outline/Intended Learning Outcomes (ILOs)

## Adherence, Intention to treat & Per protocol analyses

ILO: exemplify the challenges caused by nonadherence

ILO: restate and contrast different approaches to analyze data with nonadherence

ILO: explain their strengths and limitations for both superiority and non-inferiority studies

## Estimand framework

ILO: restate and exemplify the benefit of using the estimand framework

ILO: restate and exemplify the core attributes of estimands

ILO: recognize the challenges posed by intercurrent events and discuss different strategies to handle them

## Statistical Analysis Plan (SAP)

ILO: restate the rationale for SAPs and their content

ILO: recall examples of SAPs and where to find resources SAPs



# Statistical Analysis Plan (SAP)

## What is it?

### Statistical Analysis Plan

A statistical analysis plan is a document that contains a more technical and detailed elaboration of the principal features of the analysis described in the protocol, and includes detailed procedures for executing the statistical analysis of the primary and secondary variables and other data.

**EMA scientific guidelines:** ICH Topic E9 Statistical Principles for Clinical Trials, 1998, [https://www.ema.europa.eu/en/documents/scientific-guideline/ich-e-9-statistical-principles-clinical-trials-step-5\\_en.pdf](https://www.ema.europa.eu/en/documents/scientific-guideline/ich-e-9-statistical-principles-clinical-trials-step-5_en.pdf)

## How to write it?

- ▶ “Readers should not have to infer what was probably done; they should be told explicitly.”<sup>15</sup>  
**The higher level of evidence you seek and the more detailed/rigorous the SAP should be.**
- ▶ Precise **definitions of outcomes** (especially primary and key secondary), **analysis sets** and **statistical methods** used are essential to describe. Writing about estimands is increasingly promoted.
- ▶ **Prespecification** matters most for **confirmatory analyses**. Consequently, being transparent and precise about the primary and secondary outcomes analyses is essential. It is less important to write about e.g., standard descriptive analyses (although it does not hurt and often helps).
- ▶ It helps to look at SAPs from others (often available from the supplementary material of publications) and guidelines/tutorials (see next slides).



# Resources for SAP writing

- ▶ **TransCelerate template:** extensively used in the pharmaceutical industry.  
Maybe too detailed for most academic trials, but provides good inspiration, especially about how to write about **estimands**

<https://www.transceleratebiopharmainc.com/assets/clinical-content-reuse-solutions/> (Accessed: 2025-10-11)

- ▶ **Template from Stevens et al:**<sup>16</sup> focused on academic trials, provides many examples to include in suggested SAP sections, but nothing about estimands.
- ▶ **Guidelines from Gamble et al:**<sup>17</sup> (mostly) focused on academic trials. Easy to read, valuable, rather consensual, but nothing about estimands.
- ▶ **Chapter in Evans & Ting's book:**<sup>18</sup> clear, easy to read, provides example of a SAP outline and rationale. Refer to ICH guidelines and provides historical remarks. But nothing about estimands.

---

<sup>16</sup> Stevens et al. "A template for the authoring of statistical analysis plans." Contemporary Clinical Trials Communications 34 (2011): 101100.

<sup>17</sup> Gamble et al. "Guidelines for the content of statistical analysis plans in clinical trials." JAMA 318.23 (2017): 2337-2343.

<sup>18</sup> Fundamental Concepts for New Clinical Trialists, Scott Evans & Naitee Ting (2015)



# SAP examples

- ▶ A few I know well:
  - ▶ Nielsen et al (2024): [https://cdn.clinicaltrials.gov/large-docs/25/NCT04563325/SAP\\_000.pdf](https://cdn.clinicaltrials.gov/large-docs/25/NCT04563325/SAP_000.pdf)
  - ▶ Mortensen et al (2023): [https://cdn.clinicaltrials.gov/large-docs/40/NCT03396640/SAP\\_000.pdf](https://cdn.clinicaltrials.gov/large-docs/40/NCT03396640/SAP_000.pdf)
  - ▶ TRACTION trial (2024): [https://cdn.clinicaltrials.gov/large-docs/62/NCT06101862/SAP\\_000.pdf](https://cdn.clinicaltrials.gov/large-docs/62/NCT06101862/SAP_000.pdf)
  - ▶ “How long” study (2025): [https://cdn.clinicaltrials.gov/large-docs/64/NCT04637464/SAP\\_000.pdf](https://cdn.clinicaltrials.gov/large-docs/64/NCT04637464/SAP_000.pdf)
- ▶ A recent scandinavian example from an important study: (See pages 216–241)  
[https://www.nejm.org/doi/suppl/10.1056/NEJMoa2505985/suppl\\_file/nejmoa2505985\\_protocol.pdf](https://www.nejm.org/doi/suppl/10.1056/NEJMoa2505985/suppl_file/nejmoa2505985_protocol.pdf)
- ▶ For more examples, just browse supplementary material of papers published in your favorite journals (or “prestigious” journals).

