Estimands: implemented, but not fully embraced

Rob Hemmings and Frank Bretz*

EFSPI, Basel Sept 2024

* acknowledging Emmanuel Zuber, Hans-Jochen Weber, Angelika Caputo

History

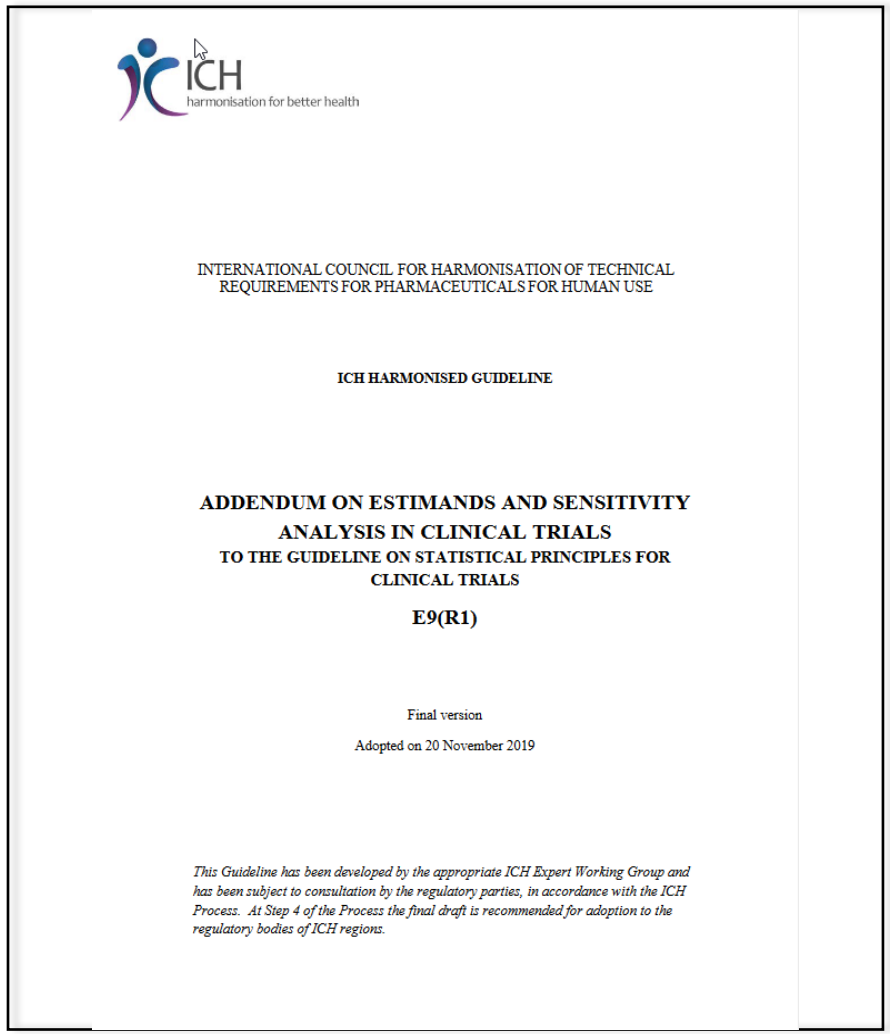
- ICH E9(R1) concept paper in October 2014 highlighted the issues arising from absence of guidance on the primary estimand for a confirmatory clinical trial
  - *"The property that is to be estimated in the context of a scientific question of interest."*

- Which data are relevant for estimation?
- How to handle missing data?

- Ultimately, what are we trying to learn from clinical trials about the investigational products being tested?

Implementation status of ICH E9(R1)

By regulatory ICH members



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)

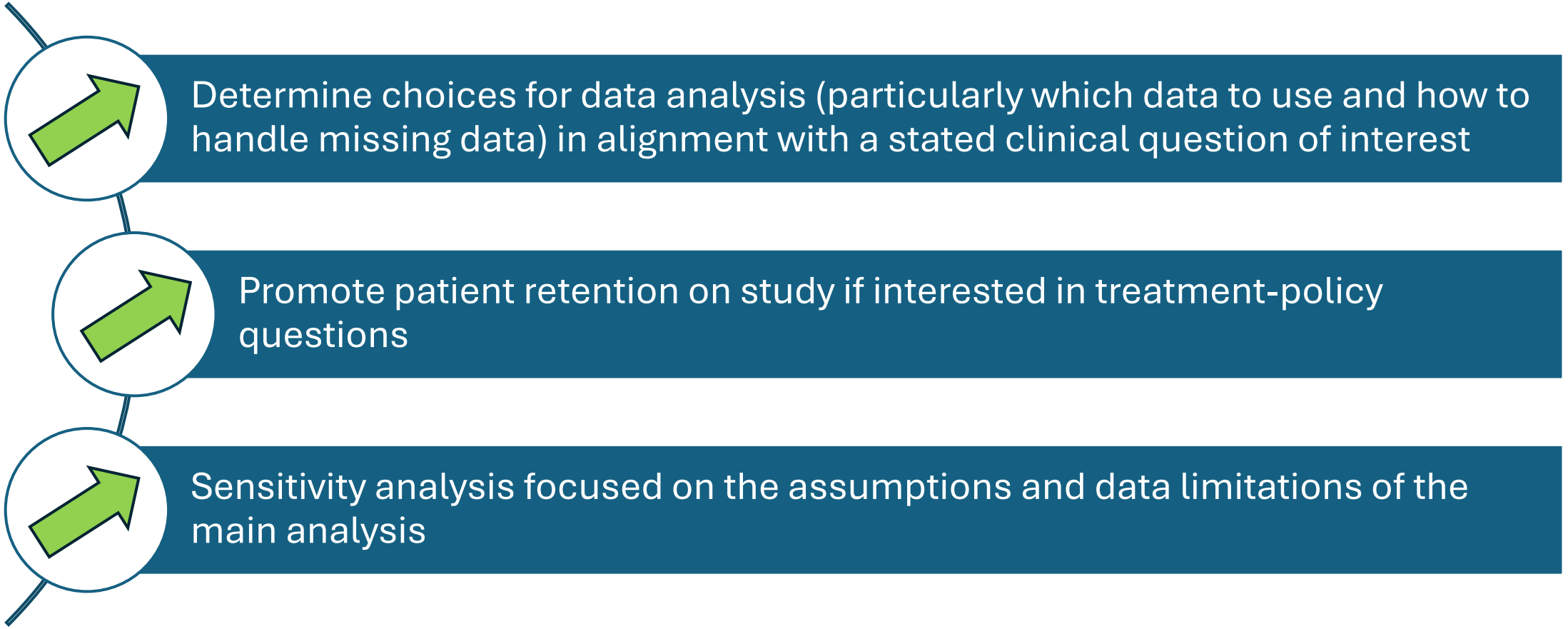
Final version

Adopted on 20 November 2019

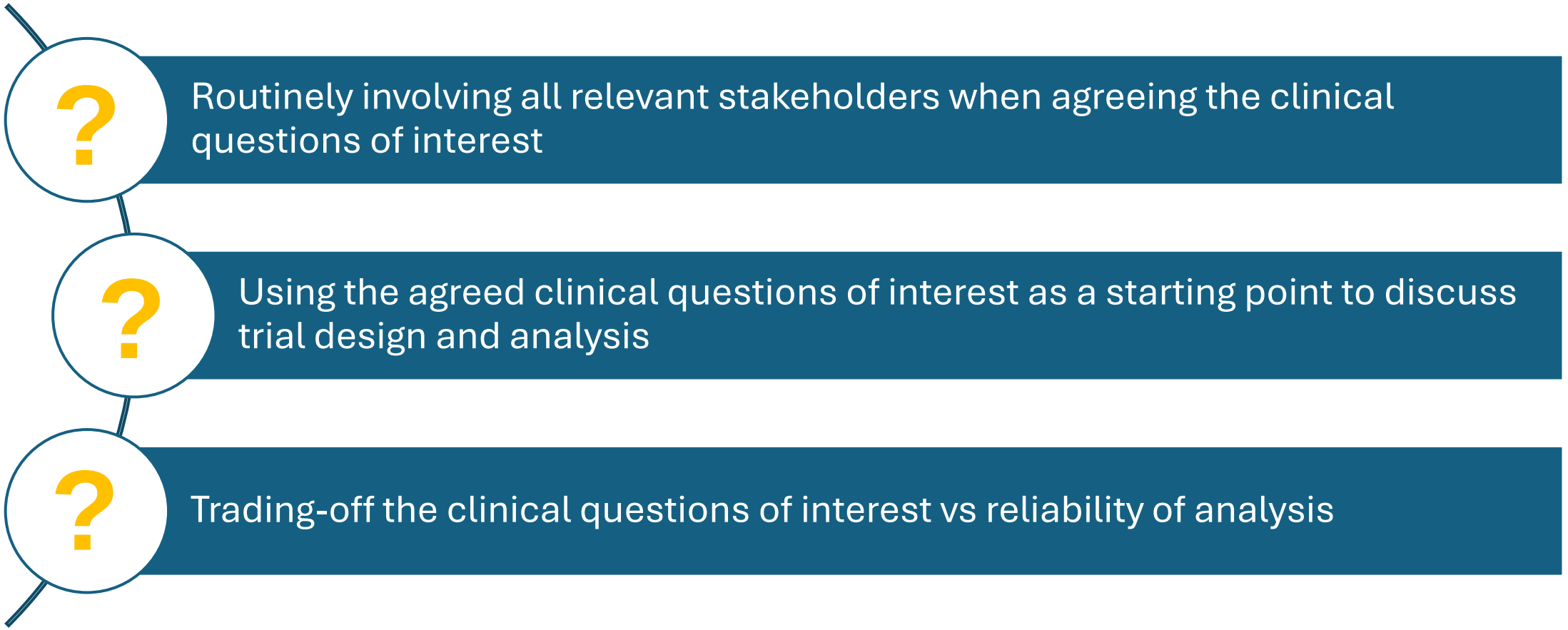
This Guideline has been developed by the appropriate ICH Expert Working Group and has been subject to consultation by the regulatory parties, in accordance with the ICH Process. At Step 4 of the Process the final draft is recommended for adoption to the regulatory bodies of ICH regions.

- ANVISA, Brazil - In the process of implementation
- COFEPRIS, Mexico - Not yet implemented
- EC, Europe - Implemented; Date: 30 July 2020
- EDA, Egypt - Implemented; Date: 1 January 2019
- FDA, United States - Implemented; Date: 11 May 2021
- HSA, Singapore - Implemented; Date: 1 November 2019
- Health Canada, Canada - Implemented; Date: 21 July 2020
- MFDS, Republic of Korea - Implemented; Date: 15 Dec 2023
- MHLW/PMDA, Japan - Implemented; Date: 20 June 2024
- MHRA, UK - Implemented; Date: 1 July 2020
- NMPA, China - Implemented; Date: 25 January 2022
- SFDA, Saudi Arabia - Implemented; Date: 10 August 2023
- Swissmedic, Switzerland - Implemented; Date: 30 Nov 2019
- TFDA, Chinese Taipei - Implemented; Date: 9 February 2021

Implementation



Implementation



Implementation

ICEs and strategies are the means, not the end…

Attribute	Strategy
Population	*[some text]*
Treatment conditions	*[some text]*
Variable	*[some text]*
Intercurrent events Intercurrent event #1 Intercurrent event #2 …	Strategies Strategy #1 Strategy #2 …
Summary measure	*[some text]*

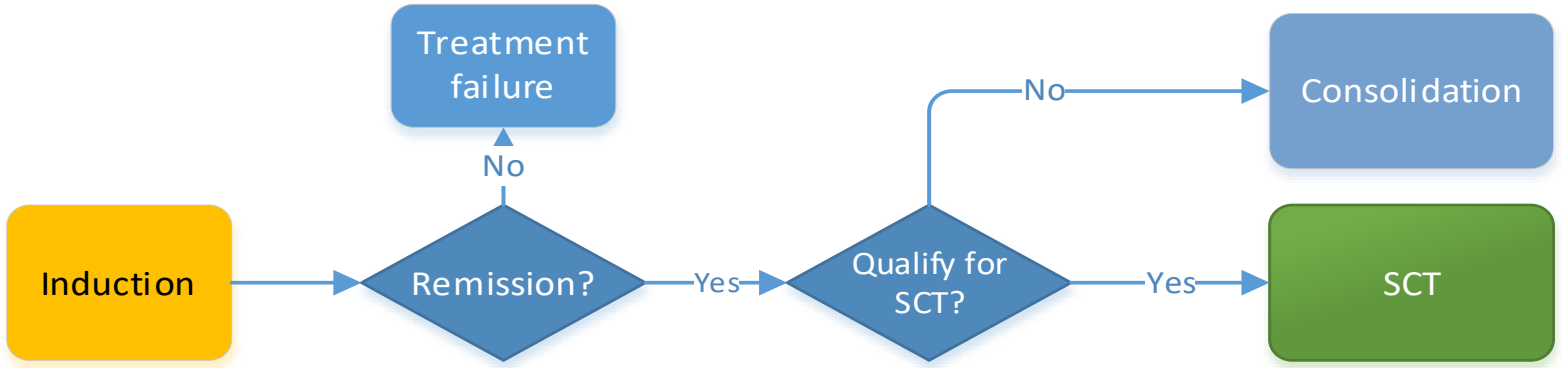
Objectives and flow

- Through use of examples (a non-exhaustive list of interesting cases!), highlight points to consider when contemplating whether a ‘strategy’ truly reflects a clinical question of interest
  - Our starting point is that regulatory advice often defaults to the ‘treatment-policy’ strategy* (e.g., for discontinuation of treatment, use of additional treatments) based on statistical rather than clinical considerations.

    **The occurrence of the intercurrent event is considered irrelevant in defining the treatment effect of interest*” (ICH E9(R1))
- To promote discussion between the panel and the audience on:
  - The reason behind the predominance of the treatment-policy strategy, and whether / when it truly reflects a clinical question of interest.
  - Promoting involvement of other disciplines in the estimand conversation.
  - The reason we undertake clinical trials: what is it that we want to learn about new medicines?

1. ‘Treatment-policy’ seems uncontroversial: we learn about the effect of a treatment sequence

- AML is characterized by the rapid growth of abnormal cells in the bone marrow and blood, interfering with normal blood cell production
- Standard of care (SOC) consists of a treatment sequence including intensive induction and consolidation chemotherapy
- Patients with a complete remission during induction / consolidation are eligible for SCT, given with a curative intention



AML: Acute myeloid leukemia
SCT: Hematopoietic stem cell transplantation

1. ‘Treatment-policy’ seems uncontroversial: we learn about the effect of a treatment sequence

- Phase III trial to assess the efficacy and safety of midostaurin, added to SOC and continued as maintenance therapy, compared to SOC alone.[1]
  - Interested in survival times.
- Midostaurin given during induction therapy might have impact on the eligibility of patients to receive SCT as well as on its outcome.
- Since SCT is part of clinical practice to treat newly diagnosed AML patients, the treatment effect of main interest is that of midostaurin on the overall treatment strategy, which includes the option of SCT.
  - Results in a ‘treatment-policy’ question in respect of SCT use.

[1] *N Engl J Med* 2017; 377(5):454–464 (link)

1. ‘Treatment-policy’ seems uncontroversial: we learn about the effect of a treatment sequence

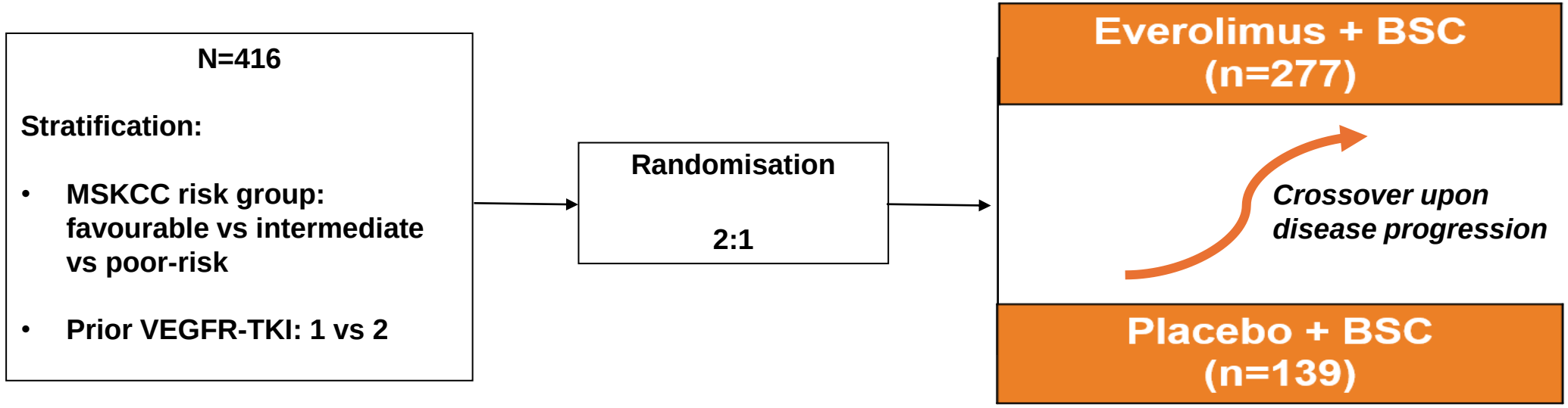
- The estimand does not tell us directly / only about the effect of midostaurin, but about the effect of a well-defined treatment sequence that includes midostaurin (in comparison to one that does not).
  - Treatment sequence in the control arm reflects SOC.
- Seems to be an uncontroversial basis to conclude on evidence of ‘therapeutic efficacy’ to approve midostaurin.
  - A new well-defined treatment sequence that improves survival times.
  - Analysis understood to fully respect / utilize randomization.
- Seems important to have clarity
  - …on the ‘treatment condition’ of interest for the estimand description and for the label,
  - …that effects of midostaurin in induction, in consolidation and in maintenance (a new therapeutic principle), as well as the indirect effect mediated by a possible increase in SCT frequency, are not isolated.

2. ‘Treatment-policy’ seems more controversial, but may still be preferred

- Late-line oncology, want to quantify impact on overall survival (OS) in addition to progression-free survival (PFS).
- Many new drugs have plausible mechanism of action and promising data from early phase development.
- Randomised controlled trial of new drug vs best supportive care (BSC), but with crossover at progression.
  - Common arguments for this are a clinical / ethical imperative for crossover, or that investigators/patients are only willing to participate in trials where patients are guaranteed to receive new drug.

2. ‘Treatment-policy’ seems more controversial, but may still be preferred

Metastatic renal cell carcinoma patients with disease progression on or within 6 months of prior VEGFR-TKI therapy (sunitinib, sorafenib, or both).[1]



MSKCC: Memorial Sloan Kettering Cancer Center
VEGFR-TKI 1/2: Vascular endothelial growth factor-tyrosine kinase inhibitors Type 1/2

[1] *Lancet* 2008; 372:449–456 (link)

2. ‘Treatment-policy’ seems more controversial, but may still be preferred

- Would want to estimate the effect of new drug vs BSC on OS – but cannot do the experiment.
- Trial design should not dictate the choice of estimand.
- Should consider hypothetical in respect of crossover: ‘… as if option for crossover were not available.’
- Might estimate using, e.g., RPSFT model, IPCW, etc. but
  - …methods rely on untestable assumptions,
  - …utility depends on proportion of, and reasons for, crossover.
- Hence, regulatory preference for treatment-policy – based on properties of estimation, arguably despite the clinical question of interest.

RPSFT: Rank-preserving structural failure time
IPCW: Inverse-probability-of-censoring weighting

2. ‘Treatment-policy’ seems more controversial, but may still be preferred

- EMA,[1] Question and answer on adjustment for cross-over in estimating effects in oncology trials:
  - “*Given that the underlying assumptions of the adjustment methods for cross-over described above can in principle not be proven to be true, a positive result from an analysis adjusted for cross-over cannot be used to rescue a trial that is negative as per other evidence, or to ascertain that a treatment confers an OS advantage when this is not apparent in an analysis that does not (strongly) depend on unverifiable assumptions, such as an ‘ITT-analysis’ that uses the observed OS outcome for each patient. For these reasons, these analyses may only be useful for regulatory purposes as supportive or sensitivity analyses with (as outlined above) a clearly demonstrated robustness against deviations from the underlying assumptions.*”
- Afinitor (everolimus), EMA[2] SmPC Section 5.1:
  - “*No statistically significant treatment-related difference in overall survival was noted (hazard ratio 0.87; confidence interval: 0.65-1.17; p=0.177). Crossover to open-label afinitor following disease progression for patients allocated to placebo confounded the detection of any treatment-related difference in overall survival.*”

[1] https://www.ema.europa.eu/en/documents/scientific-guideline/question-and-answer-adjustment-cross-over-estimating-effects-oncology-trials_en.pdf
[2] https://www.ema.europa.eu/en/documents/product-information/afinitor-epar-product-information_en.pdf

An aside…

- Common practice in oncology for a new drug to be approved for later line settings, and then move to earlier lines.
- In that case, ‘crossover’ can reflect comparison of a new treatment sequence (e.g., drug used in 2nd line) with an approved treatment sequence (e.g., drug used in 3rd line).
- Treatment-policy question in respect of crossover then reflects comparison of a new treatment sequence against an indicated (3rd line) use of drug and SOC in clinical practice.
  - Importantly different setting than on the previous slides and seems ‘uncontroversial’ per Example 1.
  - Associated estimate compares treatment sequences that both include drug.

3. Treatment-policy question complicated by changes in SOC during the study

- Clinical trials in many settings can last for years (liver disease, cardiovascular outcome trials, etc.).
- AD is a neurodegenerative disease characterized by memory loss and progressive impairment in patients' cognition and ability to function.
- Next 'wave of innovation' after amyloid monoclonal antibodies (mAbs) expected for different validated targets.
- Competitive landscape is complex and evolving with approved amyloid mAbs
  - Current SOC: Symptomatic treatments.
  - Amyloid mAbs may become SOC in the coming years where approved and reimbursed in different geographical regions.

AD: Alzheimer's disease

3. Treatment-policy question complicated by changes in SOC during the study

- Phase III trial in AD patients to assess the efficacy and safety of DRUG X (non-amyloid mAb), added to SOC, compared to SOC alone.
- Investigators and patients may decide to stop taking study medication (DRUG X or placebo) and initiate treatment with commercially available amyloid mAbs.
  - Amyloid mAbs are specified as prohibited concomitant medication.
  - Unethical to prevent access to such treatments when indicated and available.
  - More likely to happen on SOC alone (‘placebo’).
- What is the primary clinical question of interest, particularly with regard to the treatment strategies being compared?

3. Treatment-policy question complicated by changes in SOC during the study

- What guides the choice of a relevant and interpretable treatment effect (‘estimand’) in settings of changing SOC?
  - Importantly different setting than in Example 1 due to changing SOC.
- How should respective treatment strategies be defined upfront in a protocol?
- Which control strategy is of interest for comparison to judge efficacy and safety and to communicate to prescribers?
  - ‘Control’ treatment strategy is in flux, moving from none to potentially widespread use of amyloid mAbs for certain types of patients in some regions.
  - The one at the trial start? The one at the trial end? The ‘average’ of SOCs across the trial?

More asides…

- As more drugs in the ‘new wave’ arrive, their trials are increasingly affected by use of amyloid mAbs in the control arm. Even if similarly effective, treatment effect estimates under treatment-policy will likely decrease in magnitude.
- How to apply regulatory consistency?
  - Consistency in use of ‘treatment-policy’ questions leads to inconsistency in the comparisons being made (‘treatment condition’ in the control arm changes) and hence in the estimates being produced.
- How to retain meaningfulness of estimates for prescribers?

4. Treatment-policy question complicated by rescue medication

- Previous examples are based on estimating treatment effects on long-term outcomes impacted by a treatment sequence.
- There are also challenges in studies assessing impact of treatment on short-term outcomes (e.g., symptomatic treatments).
  - Sufficient (preferable?) to understand effects of the drug rather than the treatment sequence?
- In clinical trials, comparing ‘drug’ vs ‘no drug’ can be attractive.
  - Superiority trials can be preferred to non-inferiority trials for proof of efficacy as there is no concern over (lack of) assay sensitivity, derivation of a non-inferiority margin or constancy assumptions.
- Comparison to ‘no drug’ / placebo (with rescue medication if patients deteriorate) does not reflect standard clinical practice
  - Patients suffering from pain, hypertension, Type II diabetes, migraine – and many more – would still be treated in practice.
- In the following, Drug Z is for acute pain. A placebo-controlled trial might be preferred. Rescue medication is available.

4. Treatment-policy question complicated by rescue medication – Scenario A

- Experimental treatment is not well tolerated.
- Patients on experimental arm:
  - 50% switch to (early and effective) rescue, 50% remain on Drug Z
- Patients on control arm:
  - 10% switch to effective rescue, 90% remain on placebo
- Superiority is shown for the treatment-policy question in respect of rescue medication
- The comparison to treatment sequences seems uninteresting in respect of approving Drug Z or communicating effects to prescribers

4. Treatment-policy question complicated by rescue medication – Scenario B

- Experimental treatment is tolerated.
- Patients on experimental arm:
  - 10% switch to (early and effective) rescue, 90% remain on Drug Z
- Patients on control arm:
  - 50% switch to effective rescue, 50% remain on placebo
- Superiority is shown for the treatment-policy question in respect of rescue medication
- Is the treatment-policy question of interest for approval?
- Is the associated estimate interpretable to prescribers?

4. Treatment-policy question complicated by rescue medication – Scenario B

Is the treatment-policy question of interest for approval?

Yes, because it is conservative.

- But … 'conservative' compared with what?
- If some 'pure' effect (effect of Drug Z without impact of rescue medication use) is what we want to know – consistent with the choice of placebo control – then there is an argument for that to be reflect as the estimand!

4. Treatment-policy question complicated by rescue medication – Scenario B

Is the associated estimate interpretable to prescribers?

- Arguably, only if one knows the rule for use of rescue, how many patients/group used rescue and when.
- Under a different protocol with a different rule for use of rescue, the estimate for the 'treatment effect regardless of the use of rescue' therapy is not X points.
  - X points is not 'the effect' of Drug Z.
- Does preference for treatment-policy question and associated analysis promote sponsors to run studies where the treatment effect estimate is maximised through choice of patients and experimental conditions, impacting external validity?
- The estimate for the 'hypothetical' treatment effect '…as if rescue medication were not available' is not similarly impacted and might in some sense give 'the' effect of Drug Z.
  - Estimate is potentially more 'stable' in respect of changing treatment paradigms

4. Treatment-policy question complicated by rescue medication – Scenario B

But hypotheticals do not reflect clinical practice…

- Perhaps they do for intercurrent events that would occur in clinical trials but not in clinical practice.
- It should not be assumed that the ‘treatment-policy’ of a clinical trial control arm reflects clinical practice.
  - The control arm of a placebo-controlled trial can, but rarely does, reflect clinical practice.
  - Hence, the estimates of treatment effect from a placebo-controlled trial rarely ‘mirror those observed in subsequent practice.’

Some regulatory guidance

- ICH E9(R1) highlights that treatment-policy questions might be "*more generally acceptable to support regulatory decision making*"

- ICH E9, Section 5.2.1, Full Analysis Set

  "*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. Under many circumstances it may also provide estimates of treatment effects which are more likely to mirror those observed in subsequent practice.*"

  - Conservative for what?
  - Under which circumstances?

- EMA[1] Assessment of SmPC section 5.1, A Guide for Assessors (draft)

[1] https://www.ema.europa.eu/en/documents/scientific-guideline/draft-assessment-smpc-section-51-guide-assessors-centralised-applications_en.pdf

Lots more to reflect on…

- Examples above focused on uses of other medications as the intercurrent events, but many other intercurrent events exists that come with their own challenges
- The clinical question of interest in respect of terminal events in different settings still presents a difficult discussion (e.g., when having to integrate mortality into a functional variable such as for Amyotrophic Lateral Sclerosis)
- Lots to discuss in respect of treatment discontinuation; widespread agreement that there is no interest in 'bad hypotheticals' (e.g., pretending a patient can stay on treatment despite a serious adverse event), but…
  - Multiple disease areas where a drug that works in tolerators would benefit public health, but requirement is to demonstrate a statistically significant treatment effect based on all patients – regardless of whether the drug is taken or not.
  - Increasingly common to use more granular intercurrent event specifications based on reason for discontinuation, with 'treatment-policy' questions of interest for discontinuations related to drug or disease, and 'hypothetical' questions otherwise. When does this approach reflect a question of interest? Why?
  - Analytical approaches for treatment-policy question and their limitations (e.g., is control-based imputation a good analytical approach where the control group comprises 'survivors'?)

Summary

- A clinical trial might aim to compare treatment sequences or might aim to learn about the effects of an individual drug. Both might be valid, though these represent importantly different clinical questions of interest.
- We are not arguing against treatment-policy questions but 'defaulting' to this is not trivial and we wish to highlight points for consideration and encourage greater reflection and discussion.
- In the trade-off between uncertainties in estimation and the true clinical question of interest, reliability of estimation currently dominates.
- Consistent with this, estimand discussions occur too often only among statisticians, overlooking perspectives from other disciplines.
- The estimand 'most valuable for regulatory decision making' and the estimand 'most informative for prescribers' might differ.
- We hope to have illustrated that the estimand framework gives a basis to discuss fundamental aspects of drug development, relating to the objectives as well as the design and analysis of clinical trials. We hope to have stimulated your interest to think about these questions!