Medicines & Healthcare products Regulatory Agency

MHRA
Regulating Medicines and Medical Devices

Use of Clear Estimands – Beyond Hypothesis Testing

Yolanda Barbachano

Outline

- My own ‘estimand journey’
- An example with tolerability data
- An example with QoL data
- Uses in safety data
- Conclusions

The views expressed in this presentation are those of the speaker, and are not necessarily those of MHRA or EMA

From concept paper to draft addendum… My own journey

- As a regulator I welcomed the idea of an addendum on estimands
- Most of my assessments and scientific advices involved discussions of estimands even before the addendum
- But we didn’t call them estimands and discussions on trial design, endpoints, data collection and missing data were often disjointed
- And often problems were spotted too late. So we used sensitivity analysis to try and ‘fix’ them
- Which did not always work

From concept paper to draft addendum… My own journey

- I helped organise talks and workshops on estimands
- I tried to learn from the discussions
- And noticed the idea of estimands was like marmite. Some people loved it and some people hated it
- Also noticed most discussions concentrated on the primary outcome of the study

- So today I will move away from the primary outcome
- And hopefully convince you that the estimands framework is useful for all outcomes (so don’t hate it!)

Descriptive Endpoints

Tolerability Example - Background

- New tool developed to measure symptomatic adverse events during cancer treatment (PRO-CTCAE)
- Includes 78 symptomatic toxicities
- Recall period is the past 7 days
- Often measured at several time points during the trial – sometimes weekly
- Workshop organised to investigate the use of this tool to assess tolerability in cancer trials
- One session focused on analytical approaches – all the speakers were statisticians

Questions

- There are many symptoms that could affect tolerability, do we analyse them all the same?
- What if some symptoms are easily resolved with medication (e.g. diarrhoea)?
- Do we look at incidence or duration of the symptom?
- Symptoms are graded. Is incidence at any grade important? Or do we just care about the worst grade? Or grade 3 and 4?
- Is the mean grade any use?
- How do we take baseline grade into account?

More Questions

- Patients that die have no tolerability issues after death. Do we only want to know symptoms whilst on treatment?
- Do we treat their ‘missing data’ the same as those that withdrew from treatment? They also have no tolerability issues any more.
- But what if lack of tolerability led to discontinuation from treatment? Is it OK to ignore this at later time points?
- And what do we do with patients that withdrew due to lack of efficacy?
- Do we want to look at tolerability of patients that withdraw separately from those that do not?

Solution

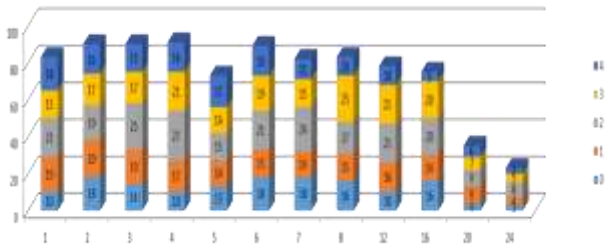
Should we talk about estimands?

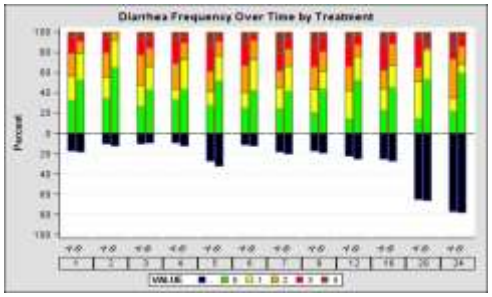
- Idea rejected as we were not testing a hypothesis. We were just meant to describe the tolerability on each arm

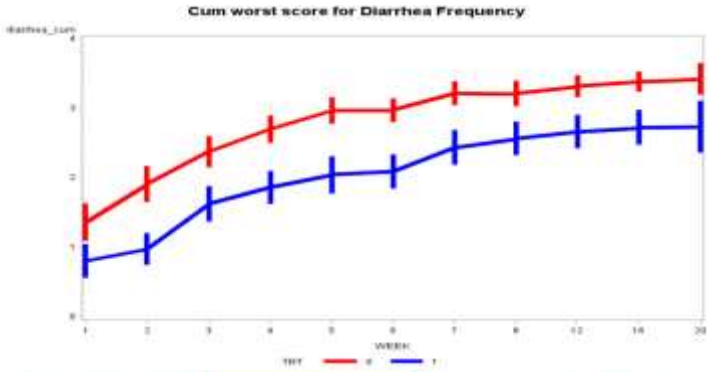
The estimand framework is useful regardless of hypothesis testing!

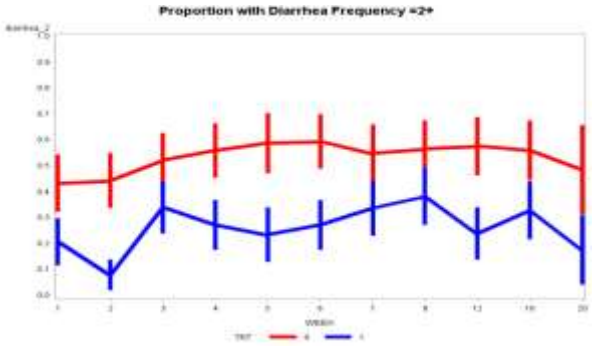
- But we did not have the view from the doctors or the patients on what estimand is of interest anyway

Graphical Outputs

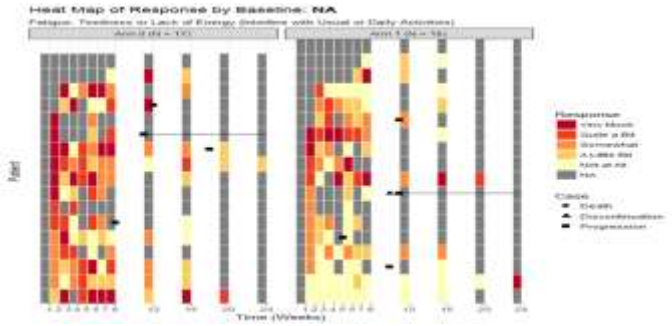








Patient ID	W1 (Baseline)	W2	W3	W4	Maximum Score at Post-Baseline	Baseline Adjustment on Any Adverse Event? Yes or No (Score)	Baseline Adjustment on Severe Adverse Event? Yes or No (Score)
1	0	0	1	3	3	Yes (3)	Yes (3)
2	1	1	4	3	4	Yes (4)	Yes (4)
3	0	1	2	1	2	Yes (2)	No (2)
4	2	1	0	1	1	No (0)	No (0)
5	3	2	3	3	3	No (0)	No (0)



Estimand Description - Tolerability

- Population
  - Everyone that started treatment
- Variable
  - PRO-CTCAE score (over all time points? Maximum score? Change from baseline? Same for every symptom?)
- Accounting for intercurrent events
  - Tolerability not relevant after death? Or could treatment cause death? Do we want to know what the tolerability would have been for patients that withdrew from treatment had they not withdrawn?
- Summary measure
  - Agreed to just describe both arms and not formally compare - Proportion of patients with each score at each time point? With score > 2? Mean score on each arm? Ignore missing data? (Linked to decisions above)

Secondary Endpoints

Example - QoL in advanced cancer

- New product for palliative treatment (e.g. for pain or fatigue)
- Does not treat the cancer
- But patients more likely to continue cancer treatment if symptoms are bearable
- Primary endpoint focused on symptom improvement
- Secondary endpoint is QoL

- Death likely in many patients
- Overall survival presented (for safety)

Analysis

- To account for the fact that the primary endpoint may have been unobserved due to death, patients who died prior to week 10 were given ranks according to their survival time and any efficacy data collected up to their death was not used;
  - But this treatment is not supposed to prevent or postpone death
- For Qol the analysis estimated the treatment effect that would have been observed if all patients had survived to week 10.
  - But is this what we are interested in?

Interaction with regulators (before estimands)

- Company took death as worst outcome for primary
- Why? Probably because this is the usual approach in cancer
- Regulators asked for further analyses to check the treatment effect on symptoms before death

- For QoL Company assumed MAR after death
- Regulators asked for further analyses with death as worst outcome
- Why? Probably because this is the usual approach in cancer

Change of mindset

- What is really the main estimand of interest in this situation?
- Should death always be counted as the worst possible outcome regardless of the endpoint?
- If the treatment is not targeting the cancer, and all patients are terminally ill, is it possible that we only want to know whether the QoL was improved whilst alive?

- Consider possible strategies mentioned in the addendum to ICH E9 separately for each main objective (symptom control and QoL)

QoL estimand – Treatment policy

Population: All treated (defined by inclusion/exclusion criteria)

Variable: Change from baseline to week 10 in QoL score

Intercurrent events: Regardless of death

Summary measure: Difference in means

QoL estimand – Composite strategy

Population: All treated (defined by inclusion/exclusion criteria)

Variable: Alive and with QoL above a certain threshold at week 10 (binary)

Intercurrent events: captured in variable

Summary measure: Difference in response rates

QoL estimand – Hypothetical strategy

Population: All treated (defined by inclusion/exclusion criteria)

Variable: Change from baseline to week 10 in QoL score

Intercurrent events: if patients had not died

Summary measure: Difference in means

QoL estimand – Principal stratum strategy

Population: Patients which would not die in 10 weeks on either treatment

Variable: Change from baseline to week 10 in QoL score

Intercurrent events: Not applicable in this population

Summary measure: Difference in means

QoL estimand – While on treatment strategy

Population: All treated (defined by inclusion/exclusion criteria)

Variable: AUC till death/week 10

Intercurrent events: Captured in variable

Summary measure: Difference in means

QoL estimand

- I have focused on death but in reality need to also consider other intercurrent events e.g. treatment withdrawal
- Some estimands considered may be impossible to calculate e.g. effect at week 10 regardless of death
- Study design and data collection will depend on chosen estimand
- Need communication between clinicians and statisticians
- May require an iterative process

Assessment of Safety

Anticancer guideline – new draft

- Changes focus on safety
- Shift from conventional cytotoxics to targeted agents and immune modulators has changed the tolerability and toxicity profiles of anti-cancer drugs.
- But data collection and analysis has not changed accordingly
- Usually a cumulative table of AEs is presented
- New draft suggests many other possible analyses

- Would be useful to think of the estimand of interest in each case

Anticancer guideline – new draft

"Cumulative incidences by toxicity grade are not sufficient to characterize the toxicity profile. The impact of an adverse drug reaction (ADR) on the benefit-risk balance may for example differ importantly depending on how the incidence, prevalence and severity change with time on treatment, and on the possibility to alleviate the ADR by dose reduction." (Intercurrent event)

"It is not uncommon that certain adverse drug reactions are most prominent during the first to second treatment cycle(s), following which tolerance appears to develop. On the other hand there is cumulative toxicity, of consequence mainly to those who have long-term benefit of the drug." (how to handle early withdrawals?)

Anticancer guideline – new draft

“A common problem is when the experimental drug shows substantially improved efficacy and patients therefore stay longer than on the comparator. This introduces a bias by observation time if the collection of AEs is stopped at the time of study drug discontinuation or shortly thereafter. Furthermore, the “real-life” safety consequences of the comparator arm will be underestimated”

“Extended safety data collection, including off-therapy and on-new therapy, may therefore be included in the study design”

Treatment policy estimand

Anticancer guideline – new draft

“Cytotoxic drugs are typically given at weekly or longer intervals and are characterised by major acute but transient toxicity, followed by recuperation before the next treatment cycle. Thus the safety profile of cytotoxic drugs presents different challenges compared with other treatments that are administered continuously, either until progression or for a limited treatment period, such as targeted drugs or immune modulators. For some products tolerability could be the major issue, while for others it can be potentially life-threatening adverse reactions. Both types of toxicity should be comprehensively investigated. The frequent co-administration of drugs from these major pharmacological groups further add to the complexity and demands on the safety collection and analysis.”

Anticancer guideline – new draft

“For key events, i.e. events that are common and affect tolerability, safety by treatment cycle is often of value. For example, fatigue or diarrhoea grade 3 for limited periods of time may not affect tolerability to a great degree, while long-term fatigue or diarrhoea grade 2 may be a major issue to the benefit-risk balance”

Different estimands may be needed for different AEs

Anticancer guideline – new draft

"All market applications should include cumulative adverse event rates from the pivotal study(ies) at the specified time points 3 months, 6 months and 1 year, in order to facilitate regulatory safety assessment."

- If estimand is clearly defined, comparison across trials is easier
- Particularly useful for single arm trials

Estimand Framework – Conclusions

- Encourages us to think about what we really want to measure
- and why
- Encourages us to think about the possible intercurrent events that will occur during the trial
- and how we want to handle them
- It enables us to plan the trial and data collection so that missing data can be minimised
- and the results are clear to both the company and the regulator

Estimand Framework - Conclusions

- It is not only relevant for the primary objective
- It is helpful to bear it in mind for any objective
- Regardless of hypothesis testing
- It requires close collaboration between clinicians and statisticians

Acknowledgements

Some of the output presented on tolerability was produced by other presenters and panellists in my session at the COAs workshop. These included:

Diane L. Fairclough, Corneel Coens, Joseph Cappelleri and Mallorie H. Fiero

References:

- ICH Concept Paper E9(R1): Addendum to Statistical Principles for Clinical Trials on Choosing Appropriate Estimands and Defining Sensitivity Analyses in Clinical Trials (2014)

- ICH E9 (R1) addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials (EMA/CHMP/ICH/436221/2017)

- Draft guideline on evaluation of anticancer medicinal products in man (EMA/CHMP/205/95 Rev.5)