



Science For A Better Life



Enhancing Clarity – Personal Remarks, a Short Review of the Estimands Concept, and a Current Discussion

4th Annual ASA New Jersey Chapter / Bayer Statistics Workshop

11 NOVEMBER 2016 / Dr. Michael Kunz, PD Dr. Christoph Gerlinger/ Whippany



Agenda

- An old discussion revisited.
- Estimands:
 - Motivating typical situation and problem statement
 - Review of the current discussion on estimands
- Case study: Thoughts on estimands in a chronic pain indication



Acknowledgements & Disclaimer

Part of this presentation is work in progress and the talk represents the current thinking of the authors and are not to be considered as official positions of BAYER.

The considerations on a chronic pain indication are joint work with Christoph Gerlinger.

Thanks to

Dr. Liying Dong

Dr. Vivian Lanius

Dr. Heinz Schmitz

for the many discussions so far



Agenda

- An old discussion revisited.
- Estimands:
 - Motivating typical situation and problem statement
 - Review of the current discussion on estimands
- Case study: Thoughts on estimands in a chronic pain indication

An old discussion revisited

Typical objectives of a diagnostic trial



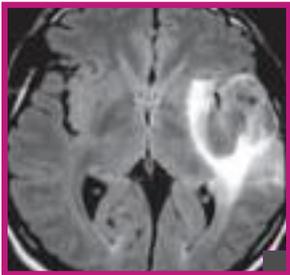
- Diagnostic trials aim to show that a certain procedure can be used to diagnose a subject correctly.
 - Subjects with the disease shall be identified as diseased (**sensitivity**).
 - Subjects without the disease shall be identified as undiseased (**specificity**)
- Example from a recent study for the diagnosis of Coronary Artery Disease (CAD):
 - Subjects being evaluated for suspected or known CAD based on signs and/or symptoms, will be invited to participate in the study.
 - One primary objective of this study is to demonstrate that sensitivity of contrast-enhanced Magnetic Resonance Imaging (MRI) exceeds the pre-specified minimum performance threshold of 60%.
- As contrast media are injected these trials are conducted under the same regulations as drug trials.

An old discussion revisited

Diagnostic (imaging) trials



- Images are obtained and subsequently assessed by experienced readers for presence or absence of disease.



An old discussion revisited

Diagnostic (imaging) trials



- Images are assessed centrally by (usually 3) blinded readers
 - Readers only get the images, but no additional information about the subject.



Blinded Read Facility



An old discussion revisited

Diagnostic (imaging) trials



Data structure for diseased (hypothetical) subjects:

Subject	Reader			Average	Majority
	R1	R2	R3		
1	1	0	0	1/3	0
20	1	0	1	2/3	1
25	0	0	0	0	0
53	1	1	1	1	1

1 = identified as diseased by the reader
0 = identified as undiseased by the reader

An old discussion revisited

Diagnostic (imaging) trials



Possible questions to ask:

Hypothesis – formally		Alternative – verbally
$H_{O,AVE}$:	$E(\text{Average}) \leq c\%$	'In the Average the 3 readers assess $> c\%$ of the diseased subjects as diseased.'
$H_{O,MAJ}$:	$E(\text{Majority}) \leq c\%$	'After the 3 readers have assessed and the assessments are condensed to the majority of assessments then this majority assessment identifies $> c\%$ of the diseased subjects as diseased.'
$H_{O,MS}$:	$E(R_i) \leq c\%$ for more than 1 index $1 \leq i \leq 3$	'For at least 2 readers it is established that they assess at least $c\%$ of the diseased subjects as diseased.'

An old discussion revisited

Diagnostic (imaging) trials



- All 3 hypotheses have been formulated in pivotal diagnostic imaging trials.
- Properties of hypotheses and related analysis procedures:
 - Average (reader) usually produces narrow confidence intervals, but does not account for potential heterogeneity.
 - Majority (reader) usually results in higher point estimates than the (average) reader.
 - Rejecting $H_{0,MS}$ ensures that the majority of the raters has a sufficient performance.
- Intensive and heated discussions (inside and outside of BAYER) about which hypothesis to choose.

An old discussion revisited

Diagnostic (imaging) trials - Conclusions



- Majority reader:
 - Is this democratic process really what is to be expected in subsequent clinical practice?
- Lacking in the past:
 - Clarity: A clear formulation which question these analysis procedures/ statistical hypotheses address.
- Benefits of more clarity:
 - Would have had abbreviated many discussions.
 - Could potentially have resulted in alternative proposals that combine some aspects and results in smaller sample sizes.



Agenda

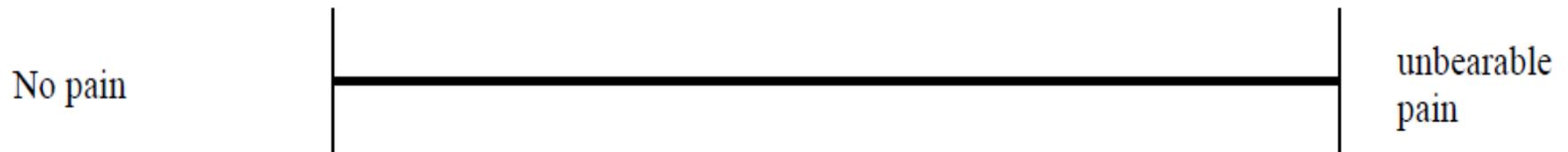
- An old discussion revisited.
- Estimands:
 - Motivating typical situation and problem statement
 - Review of the current discussion on estimands
- Case Study: Thoughts on estimands in a chronic pain indication

Estimands

Motivating typical situation



Visual Analogue Scale (VAS):



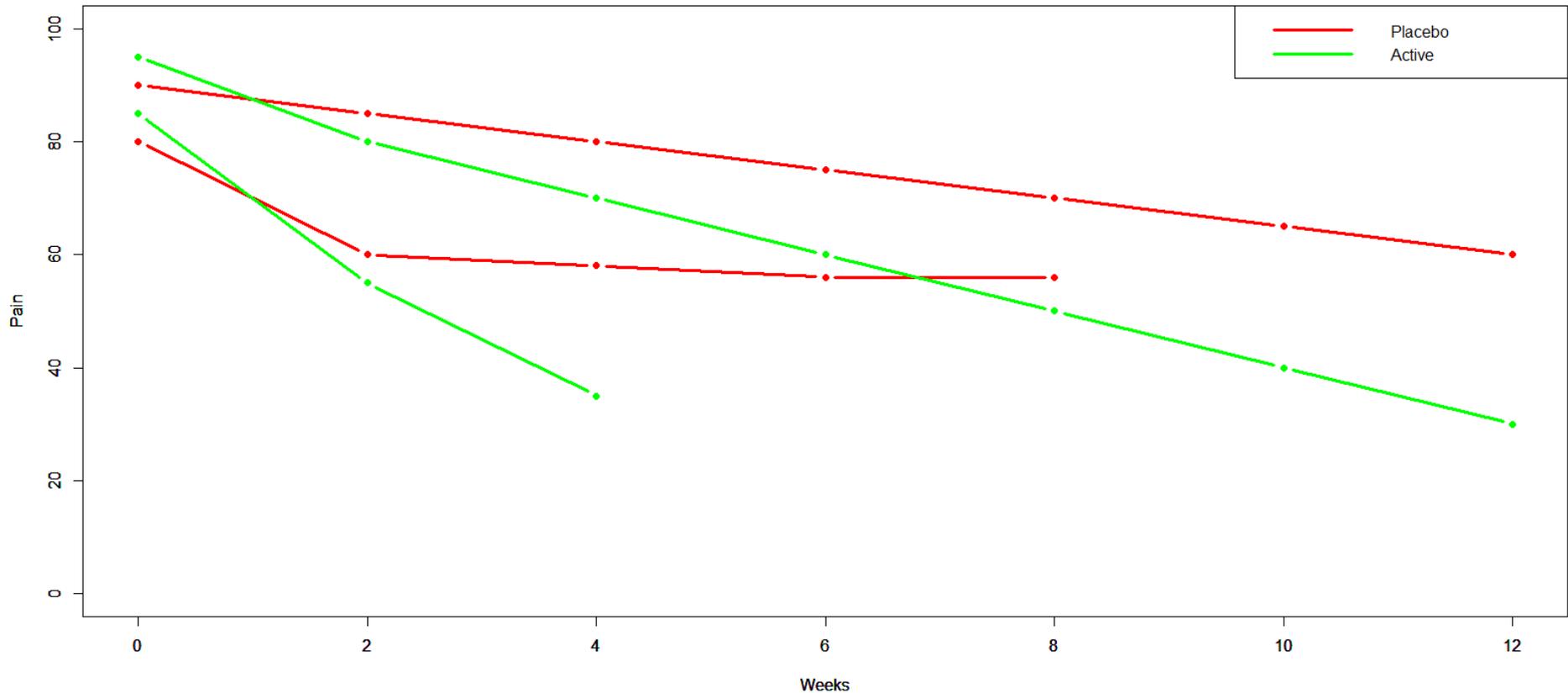
- Two-arm trial
- Chronic pain measured after 12 weeks of treatment on the VAS (0 mm = absence of pain, 100 mm = unbearable pain)
- Study objective: Show that test drug is better than control

Estimands

Motivating typical situation - individual responses



- Efficacy measured at the end of the trial and in between.



Estimands

Problem Statement



- What do we as statisticians produce?

p-values

Quantification of treatment effects

Most analysis techniques are adequate for testing the global null, 'No treatment effect at all'

What is the treatment effect

- in case of death
- in case of drop-out/ non-adherence
- if rescue medication is taken

Estimands – Problem statement



Some questions to ask on the 'Typical example'

- What is the drug's effect
 - when taken as directed?
 - in those patients that actually took drug until week 12?
 - if all patients would have taken drug until week 12?
 - until patients stop taking drug?
- What is the (prescription) effect regardless of patient adherence at week 12?

- Non-adherers/drop outs:
 - Which information is contained in these patients?
 - Can the information obtained from them be used to gauge the performance of the drug via a modified definition of an endpoint?
 - Is non-adherence alone sufficient to conclude on efficacy?

- These questions yield different answers or are present because patients do not adhere.
- Solutions may require techniques to adjust for missing data, but may go far beyond that.



Agenda

- An old discussion revisited.
- Estimands:
 - Motivating typical situation and problem statement
 - Review of the current discussion on estimands
- Case study: Thoughts on estimands in a chronic pain indication

Review of the discussion on estimands



General definition Estimand:

- The estimand is that what we want to estimate.
- Estimands address the usual lack of clarity in the study objectives.

Holzhauer, Akacha, Bermann, 2015:

The choice of an estimand needs to involve:

- Population of interest.
- Endpoint of interest.
- Measures of intervention effect.
 - Confounding factors obtained **before** randomization.
 - Confounding factors occurring **post** randomization, eg early discontinuation, rescue medication, death.

Review of the discussion on estimands

'De Jure' and 'De Facto'



De jure estimand

- expected treatment effect if treatment was taken as directed in the protocol
 - no need to collect data after discontinuation of assigned therapy as for estimating this estimand values after non-adherence will be imputed (using certain assumptions)

De facto estimand

- expected effect of prescribing/ initiating treatment in all patients regardless of adherence or rescue
 - logical to try to collect data after discontinuation of assigned therapy

Key difficulty of many of the present analyses we do on our data:

- What precise estimand do they address?
- What imputations ('certain assumptions') are useful for decision making?

Review of the discussion on estimands



'De Jure' and 'De Facto'

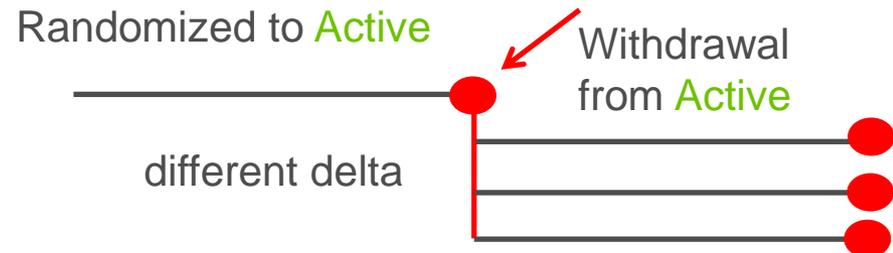
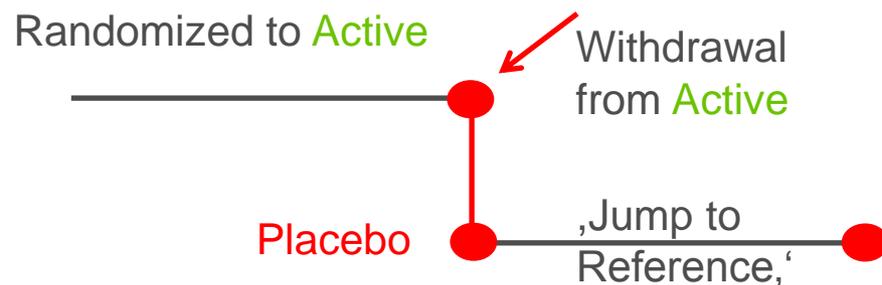
- What do these estimands mean to the patient?

De facto estimand:

- Is a mixture of those who adhered until the end and those who did not, but the patient is either a 'completer' or a 'non-completer'.

De facto estimand and no complete collection of data after stop of treatment:

- Also requires imputation. Options:



De jure estimand

- How useful is it at all to impute eg in a QoL-trial values for patients who died?

Review of the discussion on estimands

'ICH – E9 addendum on the horizon



Final Concept Paper
E9(R1): Addendum to Statistical Principles for Clinical Trials
on
Choosing Appropriate Estimands and Defining Sensitivity Analyses in Clinical Trials
dated 22 October 2014
Endorsed by the ICH Steering Committee on 23 October 2014

Draft expected by the end of 2016.

Review of the discussion on estimands



The 'Better-Half-Estimand'

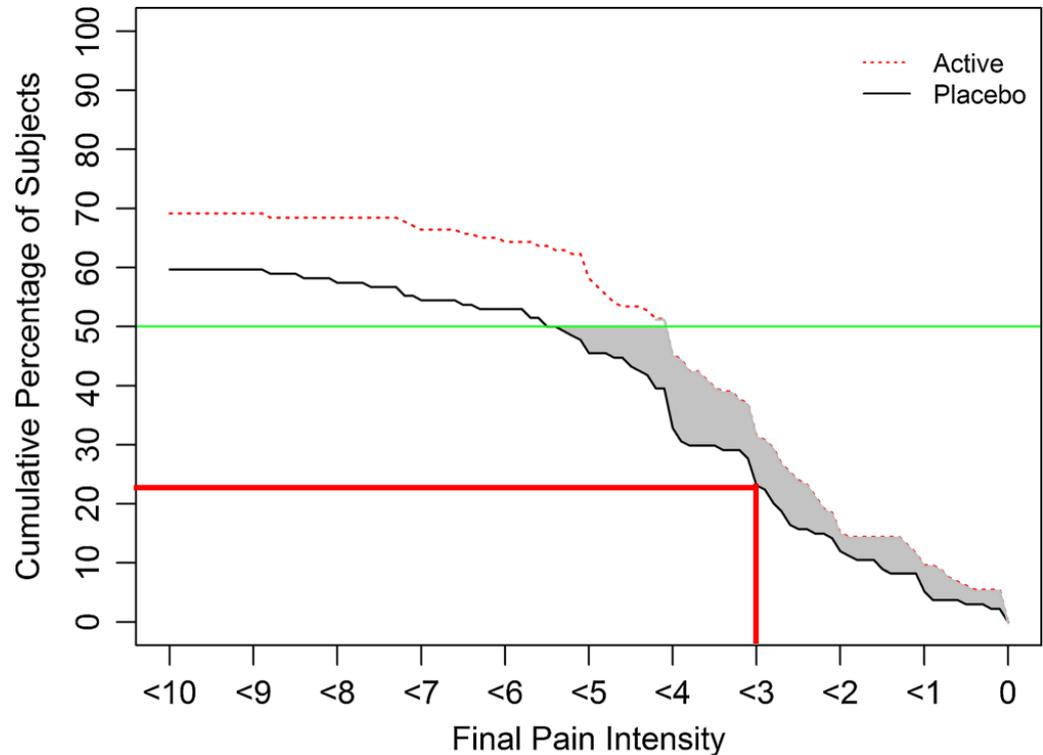
- The 'Better-Half-Estimand' was eg advocated by Permutt, 2015 and Permutt and Li 2016.
- Idea presented in the framework of Causality, but can be described as follows:
 - Assign subjects who dropped out the worst value (ie 10cm ↔ 100 mm).
 - Choose as measure of effect size the difference in means only of the best 50% in each treatment group.

Review of the discussion on estimands



The 'Better-Half-Estimand'

- Alternative view on the empirical Cumulative Distribution Function (eCDF)
 - A subject is considered a **RESPONDER** if the final value on the VAS is smaller than a threshold **c** and the subject did not drop out.
 - eCDF provides all responder analyses simultaneously.
- Reasonable, if eg
 - not much difference between bad and very bad outcomes.

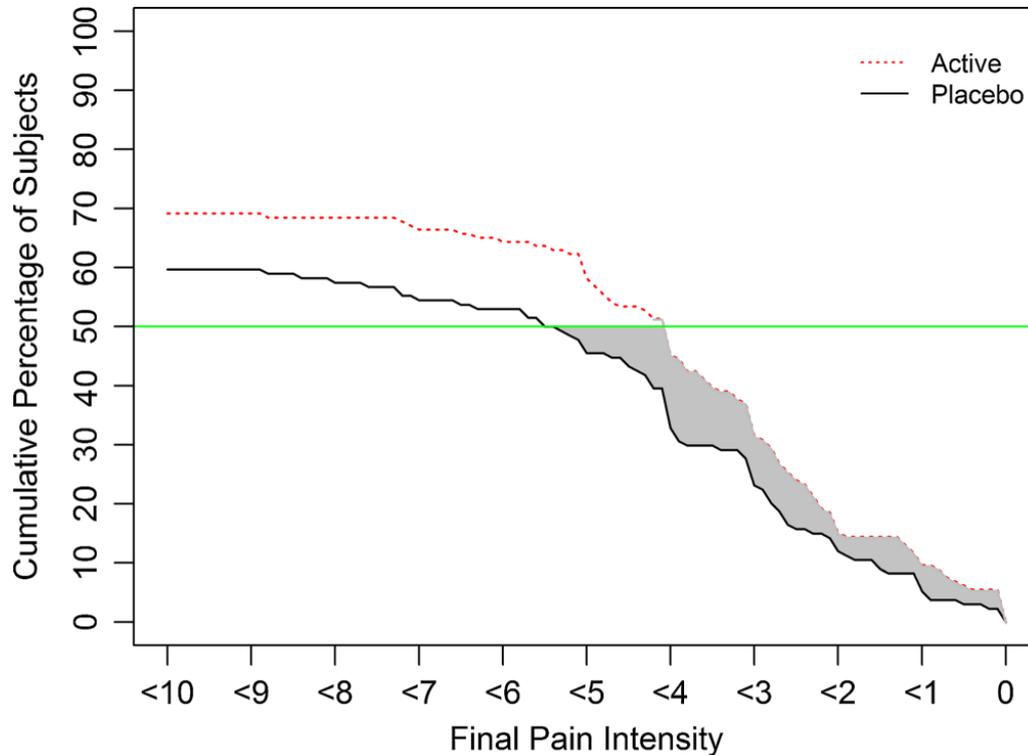


Example: Ca. 23% responders under placebo with a threshold of $c=3$ cm.

Review of the discussion on estimands



The ‚Better-Half-Estimand‘



Grey area:

~ to the mean difference of

- ‚better half‘ of Active treated subjects, and
- ‚better half‘ of Placebo treated subjects.

‚Better-Half-Estimand‘.

Corresponds to differences in (one-sided) trimmed means.

Review of the discussion on estimands



The 'Better-Half-Estimand'

- Considerations apply
 - to situations where drop out is considered a bad outcome - whether due to lack of efficacy or due to side effects.
 - (especially) where differential dropout is present, ie proportion of drop outs differ between the two treatments.
 - if it is medically justified to handle all bad outcomes as equivalent (here eg treating all pain above 8 cm as the same) and when it makes sense to investigate only those who profited from the drug.
- 'Better-Half-Estimand':
 - not straightforward to interpret.
 - has elements of an appropriately chosen endpoint that incorporates the information on drop-out into the endpoint.
- Difference of medians an alternative option to base conclusions on?

Review of the discussion on estimands



The Patient Perspective

- The concepts presented so far tried to incorporate the fact that subjects did not adhere/ dropped out prematurely.
- However do these approaches provide the patient with the necessary information to decide?
- Tripartite framework as suggested by Akacha, Bretz, and Ruberg, 2016:
 - What percentage of patients was unable to tolerate treatment and stopped taking it (**non-adherence to treatment due to safety**).
 - What percentage of patients stopped taking the treatment due to lack of efficacy (**non-adherence to treatment due to lack of efficacy**).
 - What is the efficacy and safety profile for those who adhere (**Effects in adherers**).
- Would be a move away from having only one endpoint.
 - Multiplicity?
 - How many endpoints would have to be significant? Conflicting results?



Agenda

- An old discussion revisited.
- Estimands:
 - Motivating typical situation and problem statement
 - Review of the current discussion
- Case study: Thoughts on estimands in a chronic pain indication

Case study: Thoughts on estimands in a chronic pain indication

Medical background: Endometriosis

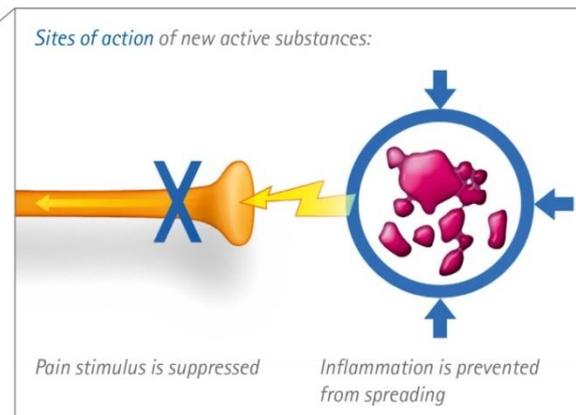
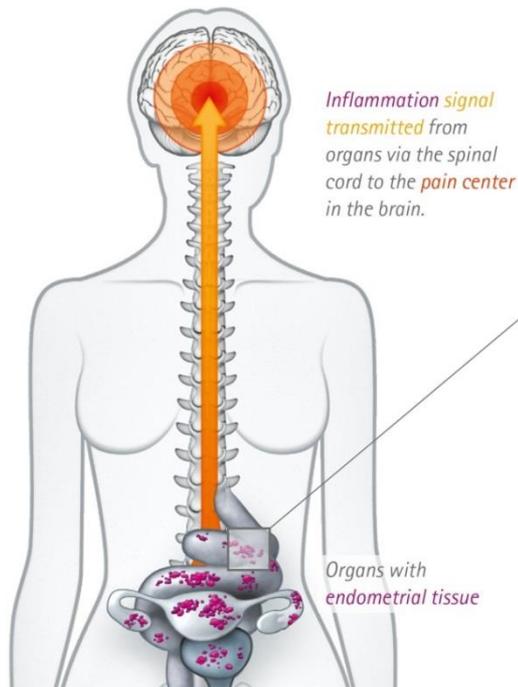


Relieving pain

Endometrial tissues can migrate to various organs in the abdominal cavity and cause inflammation there. Women who are affected by this condition generally suffer extreme pain. Bayer scientists are now taking a two-fold approach to the disease: they are developing active substances that both inhibit inflammation and influence the onset of pain signals in the endometrial lesions and the peripheral nervous system.



A crisis in the abdomen: endometriosis causes extreme pain that can seriously restrict the sufferer's daily life.



Case study: Thoughts on estimands in a chronic pain indication



Problem statement

- Pain killers (rescue medication) can be taken in addition to study drug, may affect the primary outcome. Patient decides.
- Use of rescue pain medication
 - Is by itself an outcome
 - Needs to be accounted for
- Huge daily variations within patients
- Pain correlated with menstrual cycle (in a subset of patients)

Case Study: Thoughts on estimands in a chronic pain indication

Pain measurement (ePRO)



11-point Numerical Rating Scale (NRS) with daily recall:

Daily diary

The following questions ask about pain in the target area. This means **endometriosis pain in the pelvic area**, which is shown in the diagram below:

Daily diary

Please rate your pain in the target area, at its **WORST**, in the past 24 hours.

0 1 2 3 4 5 6 7 8 9 10

No pain ↑ Pain as bad as you can imagine ↑

Weekly diary

During the past seven days, did pain in the target area make it difficult for you to walk?

Not at all

Slightly

Moderately

A lot

Extremely

Number of rescue tablets taken (daily)

Case Study: Thoughts on estimands in a chronic pain indication

Two typical endometriosis pain patterns

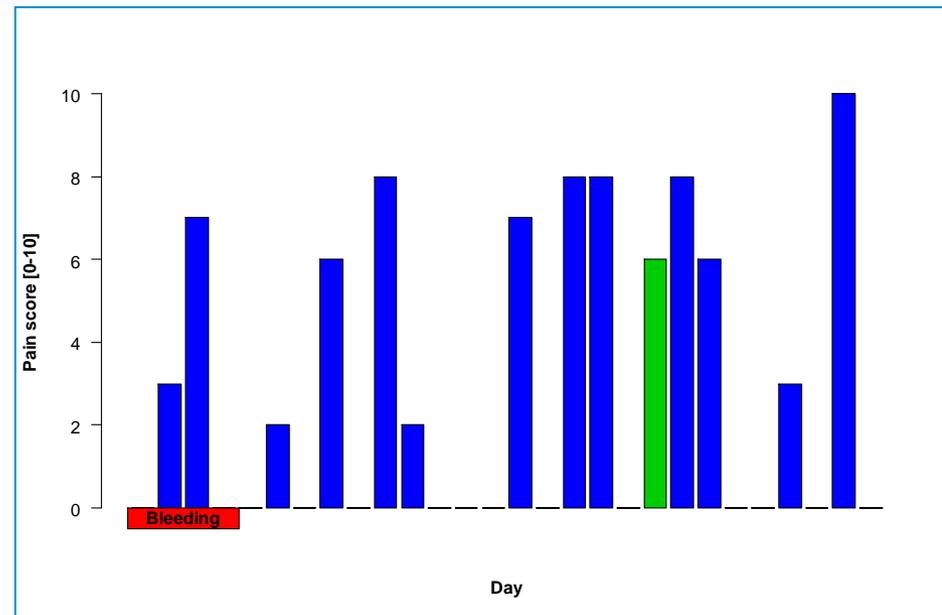
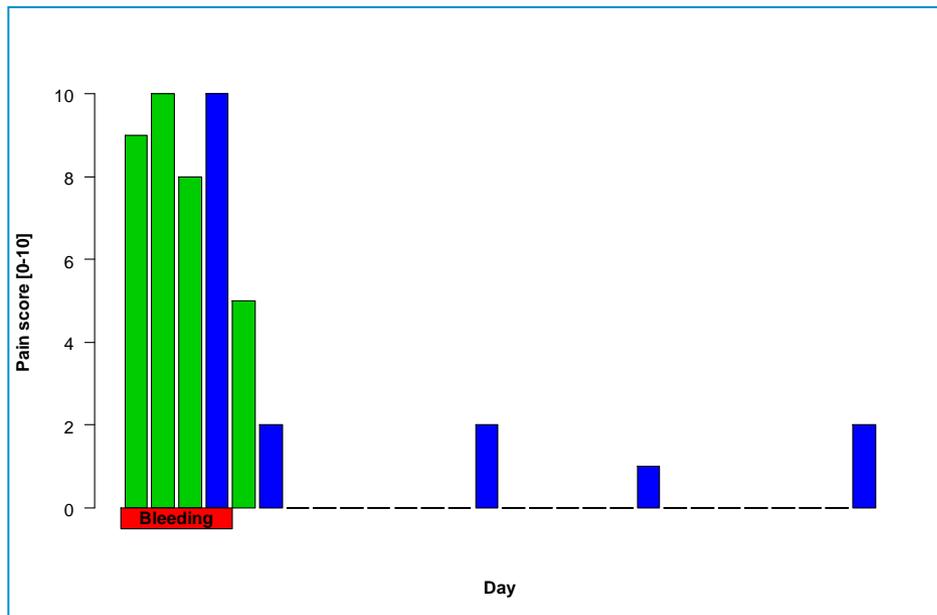


Days with rescue medication

Days w/o rescue medication

Pain correlated with bleeding

Pain independent of bleeding



Case Study: Thoughts on estimands in a chronic pain indication

Physicians' wish list for the label



I would be interested in the

effect in patients not responding to drug class "X"

effect in all patients

effect attributable to the drug

proportion of patients who discontinue due to AE **or** lack of efficacy

I would prefer to

separate efficacy and non-adherence as much as possible

Case Study: Thoughts on estimands in a chronic pain indication

From the wish list to one possible endpoint definition



Number of successful days (in 28 days/4 weeks), defined as

- pain below a certain threshold c **and**
- no intake of pain medication

Discussion points:

Threshold c as an element of a responder analysis

- Averaging over several values of c ?
- One c , but chosen individually for each patient? E.g. 50% reduction from baseline?
- No pain medication too strict?
 - E.g. no more pain medication than at baseline?

Case Study: Thoughts on estimands in a chronic pain indication

From the wish list to measure of intervention effect



Number of successful days:

- Straightforward, given the threshold definitions

Proportion of patients not discontinuing due to AE or lack of efficacy:

- Data are (almost) complete by definition
- A non-inferiority setting
- Challenge: Defining the non-inferiority margin

Summary



- Estimands framework will introduce more clarity in the reporting of trials.
- Presently lots of discussions/ contributions on estimands in the scientific literature.
- Proposed estimands often assume that it is reasonable to assume that also non-adhering subject have a value → analysis consists of techniques to model the outcome of unobserved data.
- Considerations that focus on the patient appeared rather late in the statistical literature.
- Presently there does not seem to be a consensus how data from patients shall be handled that are not missing, but are by definition unobservable (like eg QoL in dead subjects).

Summary – Case study



- The estimands framework greatly facilitated the discussions between clinicians and statisticians.
- Accounting for rescue medication (or other pain medication) in chronic pain not obvious
- Defining ‘successful days’ for the patient might be a solution
 - But, no generic definition of ‘successful days’
 - Not yet discussed with regulatory agencies
- Work in progress.



- M. Kunz 'A comparison of analysis procedures for correlated binary data in dedicated multi-rater imaging trials', Pharmaceutical Statistics, 2015
- C. Gerlinger and M. Kunz 'Some Thoughts on Estimands in a Chronic Pain Indication', 1st EFSPI-Workshop on Regulatory Statistics, September 2016
- B. Holzhauer, M. Akacha, G. Bermann 'Choice of estimands and analysis methods in diabetes trials with rescue medication', Pharmaceutical Statistics, 2015
- A. Phillips et al 'Estimands: Discussion points from the PSI estimands and sensitivity expert group', Pharmaceutical Statistics, 2016
- T. Permutt and F. Li, 'Trimmed means for symptom trials with drop outs', Pharmaceutical Statistics, 2016
- M. Akacha, Bretz, 'Estimands in clinical trials – broadening the perspective', Statistics in Medicine, 2016
- C. Mallinckrodt, G. Molenberghs, S. Rathman "Choosing estimands in clinical trials with missing data, Pharmaceutical Statistics, 2016
- T. Permutt 'A taxonomy of estimands for regulatory clinical trials with discontinuations', Statistics in Medicine, 2015



Science For A Better Life



Thank you!