Bundesinstitut für Impfstoffe und biomedizinische Arzneimittel
Federal Institute for Vaccines and Biomedicines

Paul-Ehrlich-Institut

www.pei.de

Safety Analyses: The Cinderella of Biostatistics?

\\ Regulatory perspective \\

EFSPI Statistics Leaders Meeting
23.07.2020

PD Dr. **Benjamin Hofner**
Head of Section Biostatistics

Disclaimer

Views are my own and do not necessarily represent the views of the Paul-Ehrlich-Institut (PEI), the European Medicines Agency (EMA) or any other European regulatory agency (NCA).

Setting the Scene

- Of note, **I am by no means a safety expert!**

- Focus of the talk:
  - MAA, not PhV [PSUR, …]
  - Mainly on statistical issues
  - Some clinical aspects as well
  - Oncological examples

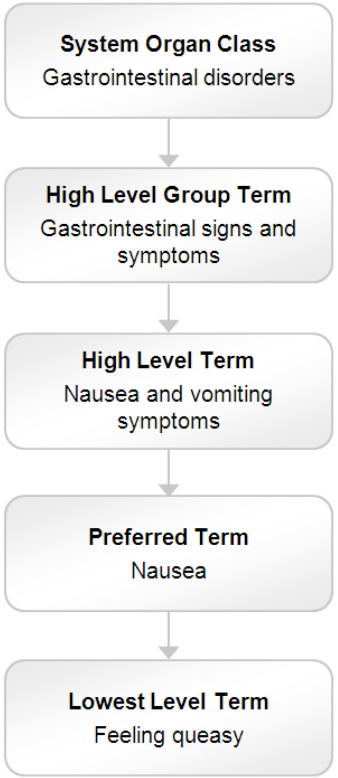
Background

- Usually no or little statistical involvement in safety assessments
  - Involvement only upon request in more complicated situations
- Mostly descriptive analyses are discussed based on clinical reasoning

- Assessment of efficacy considered „easier“ than safety
  - Adequately pre-planned
  - Clear hypotheses

Basics on Safety Analyses

- Different safety events such as
  - AE (adverse event)
  - SAE (serious adverse event ≠ severe AE)
  - AESI (AE of special interest)
  - Treatment emergent AEs
  - Treatment related AEs
  - …
- AEs are grouped by MedDRA terms
  - For signal detection preference for SOC, as PT usually to granular
- AEs are grouped by maximum severity (grade), e.g.,
  - „Any AE“
  - „Grade 3-4 AE“ (= severe or life-threatening)



Source: https://www.meddra.org/how-to-use/basics/hierarchy

Generic Grading based on NCI CTCAE

Common Terminology Criteria for Adverse Events

- **Grade 1** Mild
  asymptomatic or mild symptoms; clinical or diagnostic observations only; intervention not indicated.
- **Grade 2** Moderate
  minimal, local or noninvasive intervention indicated; limiting age-appropriate instrumental ADL*.
- **Grade 3** Severe or medically significant but not immediately life-threatening
  hospitalization or prolongation of hospitalization indicated; disabling; limiting self care ADL**.
- **Grade 4** Life-threatening consequences
  urgent intervention indicated.
- **Grade 5** Death related to AE.

(A Semi-colon indicates ‘or’ within the description of the grade.)

Standard Presentation

- AEs per arm are **usually** presented as ***frequency tables***
  - Usually in groups (SOC/PT)
  - Any AE, Grade 3-4, SAE, TRAE, Death, AESI, AE leading to dose modification, AE leading to discontinuation
  - **Sometimes *filtered*** based on absolute frequencies, size of difference between arms, p-values, …
- **Sometimes** more ***advanced methods*** such as incidence proportions and time to first event
- Additionally immunogenicity, safety in special populations, …

Examples of Standard Safety Tables

Overview of Safety Profile*

	Arm A (N=)	Arm B (N=)
Total number of patients with at least one adverse event	(94.7%)	(90.2%)
Total number of events		
Total number of patients with at least one		
Treatment-related AE	(85.2%)	(60.5%)
Grade 3-4 AE	(52.5%)	(30.1%)
Treatment-related Grade 3-4 AE	(44.1%)	(12.9%)
Grade 5 AE	(4.2%)	(3.8%)
Treatment-related Grade 5 AE	(0.4%)	
Serious Adverse Event	(28.5%)	(28.3%)
Treatment-Related Serious Adverse Event	(15.6%)	(8.4%)
AE leading to any treatment withdrawal	(16.3%)	(6.3%)
– Component 1		(6.3%)
– Component 2	(3.8%)	
– Component 3	(5.7%)	
– Component 4	(10.6%)	
– Component 5	(4.9%)	
AE leading to dose modification/interruption	(44.1%)	(25.9%)
– Component 1		(25.5%)
– Component 2	(10.3%)	
– Component 3	(24.3%)	(0.3%)
– Component 4	(29.3%)	
– Component 5	(14.1%)	(0.3%)

* Edited to remove product names and absolute frequencies

Examples of Standard Safety Tables

Any AE with ≥ 10% frequency by SOC / PT*

MedDRA System Organ Class MedDRA Preferred Term	Arm A (N=)	Arm B (N=)
Total number of patients with at least one adverse event	(83.3%)	(67.8%)
Gastrointestinal disorders		
Total number of patients with at least one adverse event	(51.0%)	(31.8%)
Nausea	(33.8%)	(13.6%)
Constipation	(21.7%)	(12.2%)
Diarrhoea	(11.8%)	(11.2%)
Vomiting	(12.9%)	(6.3%)
Blood and lymphatic system disorders		
Total number of patients with at least one adverse event	(61.2%)	(17.5%)
Anaemia	(47.5%)	(15.4%)
Neutropenia	(28.1%)	(1.4%)
Thrombocytopenia	(16.7%)	(2.4%)
General disorders and administration site conditions		
Total number of patients with at least one adverse event	(38.8%)	(34.3%)
Asthenia	(17.5%)	(12.9%)
Fatigue	(17.5%)	(12.9%)
Pyrexia	(8.7%)	(13.6%)
Respiratory, thoracic and mediastinal disorders		
Total number of patients with at least one adverse event	(16.0%)	(20.6%)
Dyspnoea	(9.9%)	(14.0%)
Cough	(9.5%)	(11.9%)
Metabolism and nutrition disorders		
Total number of patients with at least one adverse event	(19.0%)	(15.4%)
Decreased appetite	(19.0%)	(15.4%)
Investigations		
Total number of patients with at least one adverse event	(5.7%)	(10.5%)
Alanine aminotransferase increased	(5.7%)	(10.5%)

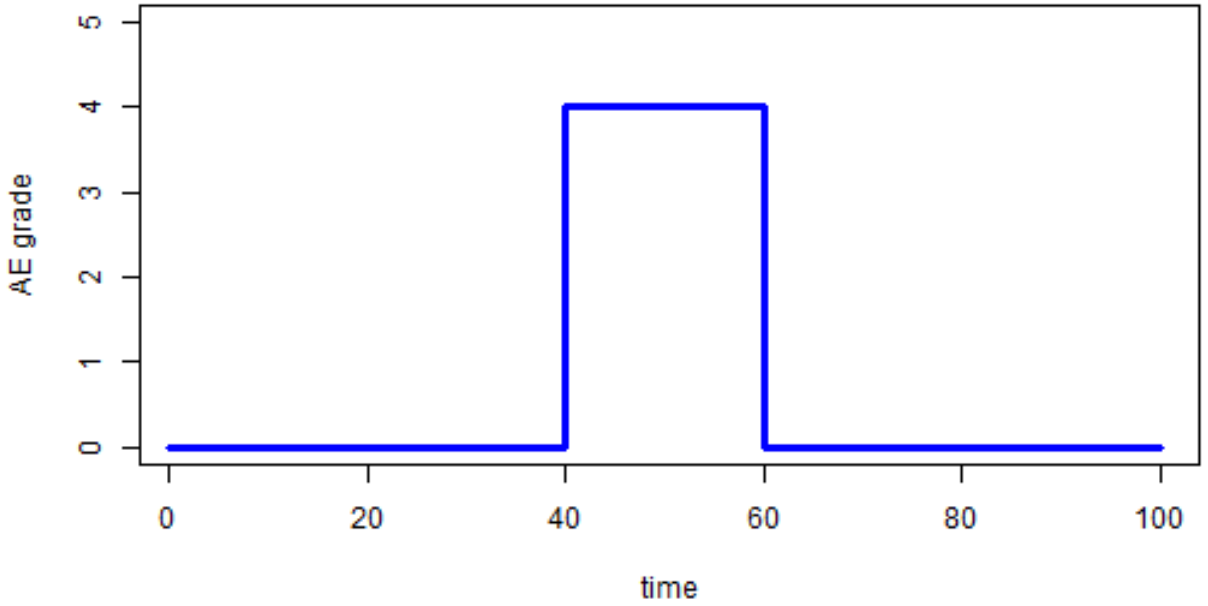
* Edited to remove product names and absolute frequencies

Some Issues with Standard Presentation

- Plethora of different tables and analyses but not well connected
  - Connection only visible in patient narratives
  - No distinction between one AE or multiple AEs per patient, per SOC, …
- Hard to assess

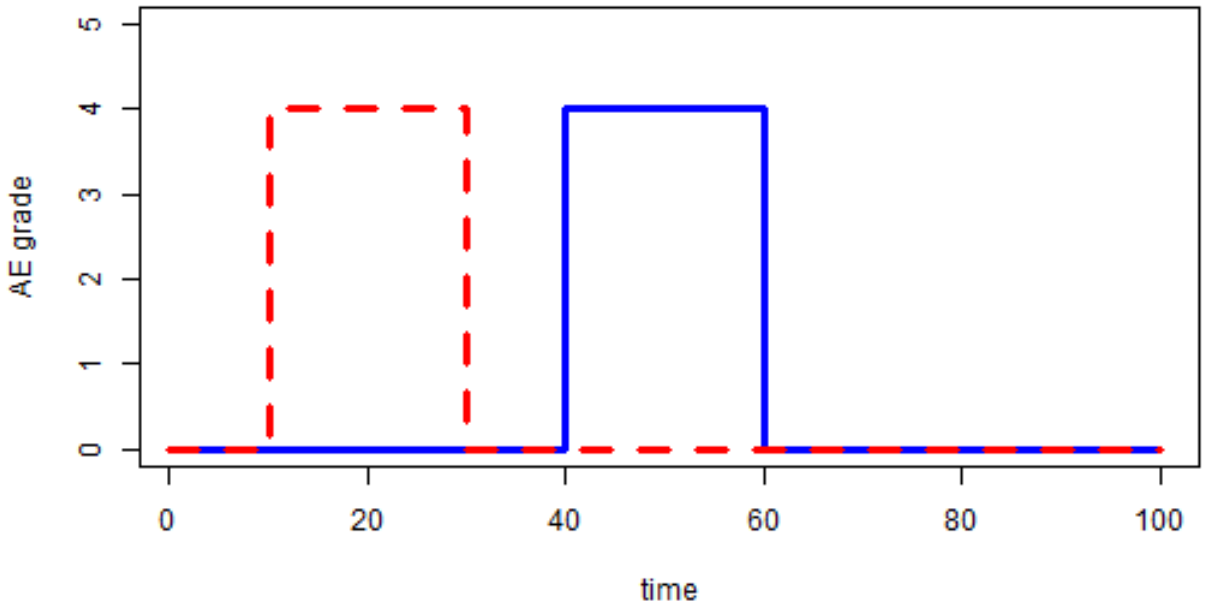
- Further issues see next slides

Example



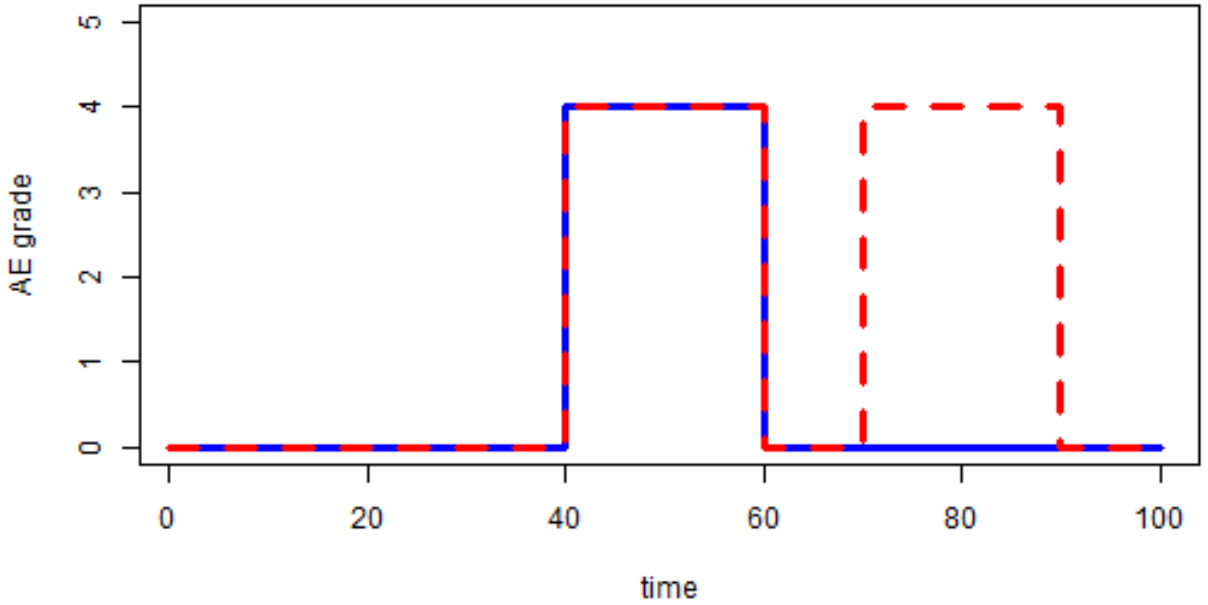
Example

Earlier onset, same duration, same grade



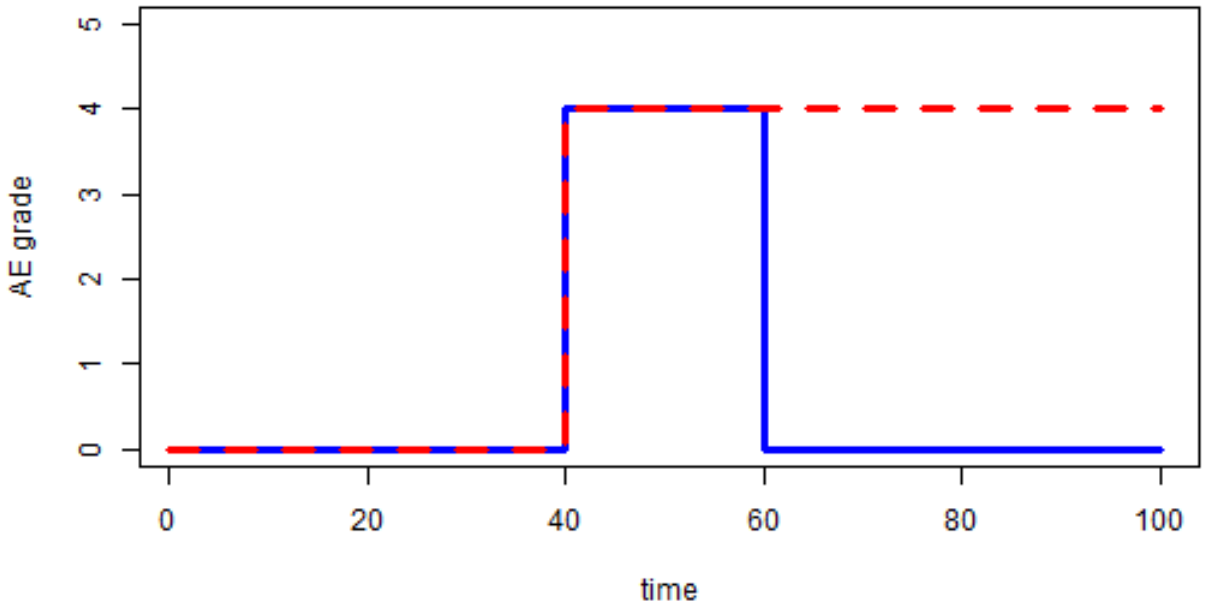
Example

Same onset, same duration, same grade, but recurring



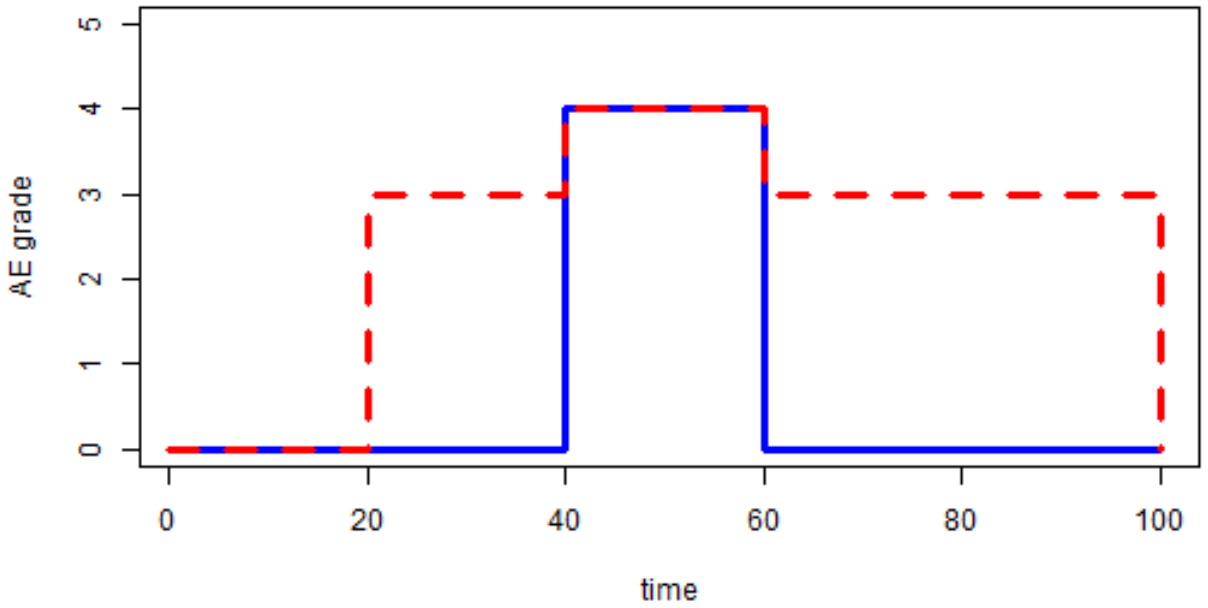
Example

Same onset, same grade, longer duration



Example

Same onset of maximum severity, same maximum grade, complex path



Estimands

Goal

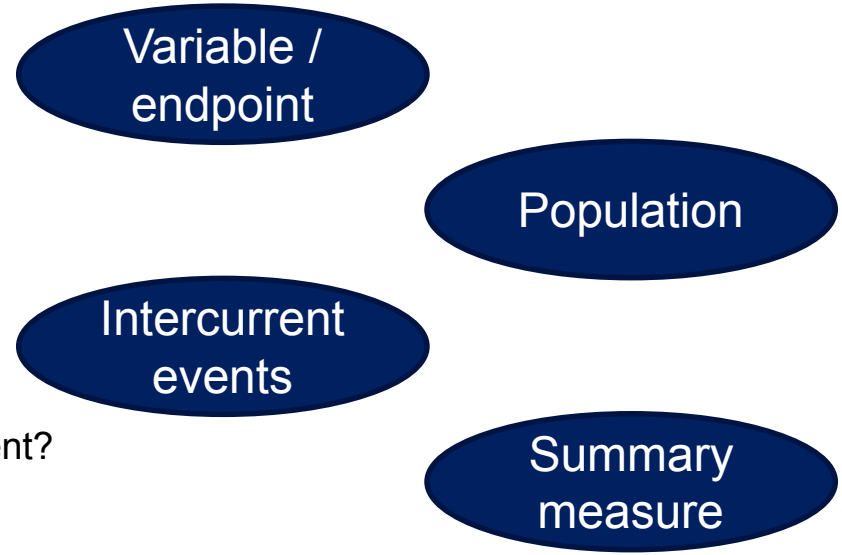
Define clinically relevant estimand

- Estimand should be defined upfront **for important safety** endpoints
  - Variable and summary measure should be defined in a sensible and clinically meaningful way
  - Patient population needs some thoughts
  - Role of different intercurrent events needs to be carefully considered
- Of note:
  - Estimands not only relevant for primary / key secondary efficacy endpoints
  - Estimands do not rely on statistical testing, estimation is a suitable goal

- Currently, we often implicitly use a while-on-treatment estimand (+ margin after last dose) for safety
  - Population, statistical measure and treatment of ICE often not well defined

Questions

- What about…
  - … timing of AEs?
  - … duration of AEs?
  - … severity of AEs?
  - … reversibility of AEs?
  - … recurrence of AEs?
  - … the impact of dose (reduction)?
  - … timing and frequency of treatment?
  - … intercurrent events such as rescue treatment?
  - … (unequal / short) duration of follow up?
  - … competing risks (e.g. death)?



Anticancer guideline – Rev. 5
(Example from pre-estimand times)

- Revised / reconsidered treatment of safety data, especially in the light of immune modulators in comparison to conventional cytotoxic drugs (chemotherapy)

"*The aim of this revision is to find ways on how to report AEs in order to improve the understanding of the toxicity and tolerability profiles of medicinal products. This could include:* ***incidence*** *and* ***prevalence*** *per period of time,* ***time to event****,* ***time-adjusted analyses*** *for AEs (e.g. by different cut-off dates or event rates per 100 patient-years) if justified based on the event rate profiles over time. It is not anticipated that all AEs would need to be reported in such detail, however. Selection criteria could for example include events leading to dose reduction or interruption, SAEs, events that are likely to affect tolerability and events of special interest, e.g. based on pre-clinical data.*"

(Concept paper, EMA/130525/2015)

Anticancer guideline – Rev. 5
(Example from pre-estimand times)

- Timing and duration of AEs
  - *"[AEs] (…) 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 or interruption."*
  - *"[AEs] (…) 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 treatment benefit."*
  - ➢ Timing and persistence of AEs needs to be taken into account
  - ➢ Impact of intercurrent events such as dose reduction (and consequence for safety profile)
- Differences in time on treatment
  - *"A common problem (…) when the experimental drug shows substantially improved efficacy and patients therefore stay longer on the experimental arm than on the comparator arm. This introduces a bias by observation time if the collection of AEs is stopped at the time of study drug discontinuation or shortly thereafter."*
  - ➢ Important to take into account if differences are expected

Based on slides from Yolanda Barbachano - MHRA

Anticancer guideline – Rev. 5
(Example from pre-estimand times)

- Treatment discontinuation / treatment switching
  - *"Extended safety data collection, including off-therapy and on-new therapy, may therefore be included in the study design"*
  - ➢ Treatment policy estimand?
- Different / multiple estimands needed for different questions
  - *"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"*
  - ➢ Combination of duration and severity might be of relevance
  - ➢ Depending on safety endpoint
- Landmark analyses
  - *"All [MAAs] (…) 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 SAT

Based on slides from Yolanda Barbachano - MHRA

PubMed Search - „estimand AND safety“

(URL: https://pubmed.ncbi.nlm.nih.gov/?term=estimand+safety; 14.07.2020)

- Only 17 hits
- Mostly on efficacy estimands (and standard safety analysis)
- Some observational trials
- Only **3 papers** really discuss estimands for safety in RCTs
  - Akacha M, Bretz F, Ruberg S. *Estimands in clinical trials - broadening the perspective*. Stat Med. 2017;36(1):5-19. doi:10.1002/sim.7033
  - Unkel S, Amiri M, Benda N, et al. *On estimands and the analysis of adverse events in the presence of varying follow-up times within the benefit assessment of therapies*. Pharm Stat. 2019;18(2):166-183. doi:10.1002/pst.1915
  - Ratitch B, Bell J, Mallinckrodt C, et al. *Choosing Estimands in Clinical Trials: Putting the ICH E9(R1) Into Practice*. Ther Innov Regul Sci. 2020;54(2):324-341. doi:10.1007/s43441-019-00061-x
- In comparison: 213 hits for „estimand“ (without restriction to safety)
- Of note: Only one occurrence of „safety“ in ICH E9 (R1) addendum

➢ A lot to be done.

Statistical Testing

Statistical Testing

- Statistical testing for differences often not suitable
  - Standard H_0 (no difference between arms) is not applicable
  - No primary endpoint with clear hypothesis
- Actually, often we want to have reassurance that there is no clinically relevant difference in safety
  - „Equivalence Testing“
- Sometimes significant improvement in safety endpoints of relevance (e.g. for mortality or other severe endpoints)
  - Standard approach to multiplicity as for efficacy endpoints
  - see Multiplicity GL (EMA/CHMP/44762/2017)

Statistical Testing
to aid flagging safety signals

- Flagging **only** significant differences not enough
  - Multiplicity correction even anti-conservative
  - Important but rare AEs might not be flagged but could be highly relevant
  - Yet, „multiplicity“ is of course an issue and needs to be taken into account when interpreting results.

➢ See also Multiplicity GL (EMA/CHMP/44762/2017) which briefly addresses safety

Subgroup Analyses

Subgroup Analyses

- Should be used for safety as well to be able to assess subgroup differences in efficacy in conjunction with safety
- Further complicates assessment as it increases the chance for spurious signals or spurious lack of signal for rare AEs
- ➢ Briefly mentioned in Subgroup GL (EMA/CHMP/539146/2013)

- **Pre-specify** important safety endpoints and **relevant subgroups** and discuss expected outcome *a priori*
  - Clinical and biological justification very important

Benefit / Risk

Benefit / Risk

- For patients weighing of risks and benefits usually a very personal decision not to be confused with regulatory B/R

- PROs for safety events (e.g. PRO-CTCAE)
  - take impact of AEs on patients into account
  - resolve some of the issues I previously discussed
  - might also be of relevance in regulatory assessment

Regulatory Benefit / Risk Assessment

- Numbers are usually based on different populations
  - „PP“ vs „ITT“
- Intercurrent events such as rescue medication might be treated differently for efficacy and safety
  - Comparison problematic from statistical point of view
- Difficult to appropriately *weight* benefits and risks
  - Comparisons are usually „informal“, e.g. based on the Effects Table
  - Further (semi-)quantitative methods are hardly ever used

Example of Effects Table

Effects Uncertainties

Effect	Short Description	Unit	Pembrolizumab 200 mg QW3	chemotherapy	Uncertainties / Strength of evidence	Ref
Favourable Effects						
OS	duration of survival from randomization to death regardless of cause	months (95% CI)	16.4 (14.0, 19.7)	12.1 (11.3, 13.3)	Efficacy not demonstrated for the TPS 1-49% subgroup (target population of the current extension of indication) OS in TPS 1-49%: median OS 13.4 (10.7, 16.9) vs 12.1 (11.0, 14.0) months; HR 0.90 (0.76, 1.06)	CSR
PFS	survival without progression from randomization to PD or death whichever occurred first BIRC per RECIST 1.1	median months (95% CI)	5.4 (4.3, 6.2)	6.6 (6.3, 7.3)	PFS not reaching statistical significance PFS in TPS 1-49%: median PFS 4.2 (4.1, 5.2) vs 6.8 (6.3, 8.1) months; HR 1.27 (95%CI 1.08, 1.50)	CSR
Unfavourable Effects						
Tolerability	Drug-related AEs	%	63.7	89.9	The rate of overall AEs in the pembrolizumab group was comparable to the reference datasets	CSR
	Grade 3-5 AEs	%	51.3	56.9		
	Drug-related G 3-5 AEs	%	18.4	41.1		
	SAEs	%	40.4	30.4		
	Death due to AEs	%	10.7	7.6		
	Discontinuation due to AEs	%	20.4	14.8		
	Discontinuation due to SAEs	%	16.4	9.3		
Selected AEOSIs	Pneumonitis	%	8.2	0.5	An increased incidence of pneumonitis was reported compared with the reference datasets	CSR
	Hypothyroidism	%	11.9	1.5		

Summary

- Safety assessment often based on
  - plethora of „unconnected“ (and potentially statistically inadequate) frequency tables

- Statistical considerations not well developed
  - partly due to imprecise / inappropriate questions („equivalence of safety“)
  - complexity of situation
  - partly due to rare events

- Multiple analyses (from different angles) required to provide a good overview and understanding of all safety aspects

Outlook

- Graphical display instead of extensive tables
  - Use graphics such as forest plots to show safety profiles
- Appropriate statistical methods / summary measures are urgently needed
  - Training of clinicians and statisticians dealing with safety needed
- Estimand framework need to be further developed and discussed in the light of safety analyses
  - Clinical and statistical input needed
- Do we need a paradigm shift in safety reporting?