Safeguarding public health

CHMP's view on multiplicity; through assessment, advice and guidelines

Rob Hemmings
Statistics Unit Manager, MHRA
CHMP member
Chair, CHMP Scientific Advice Working Party
Biostatistics Working Party Member

Disclaimer

- If the CHMP chair disagrees he will say so…

Contents

- Multiplicity issues in clinical trials
- Relevant guidelines
- General comments and literature reference
- Special topics
- Assessment
- Scientific Advice
- Conclusions

What are ‘multiplicity issues in clinical trials’?

- “More than one chance **to win**” - to win what?
  - A marketing authorisation (MA)?
  - A trial that is formally a success?
  - Statement on an individual endpoint?

- “Multiplicity of inferences is present in virtually all clinical trials.” – understatement? – all benefit-risk decisions are holistic. Multiple testing is necessary; some is ‘confirmatory’, some is ‘exploratory’.

- “The usual concern with multiplicity is that, if it is not properly handled, unsubstantiated **claims** for the effectiveness of a drug may be made as a consequence of an inflated rate of false positive conclusions.”

- **Proper interpretation of MA, trial, endpoint all requires awareness of this multiplicity concept so that strength of evidence is properly understood.**

Guidelines – E9

- “Multiplicity may arise from multiple primary variables, multiple comparisons of treatments, repeated evaluation over time and/or interim analyses.” **= Problem statement**

- “Methods to avoid or reduce multiplicity are sometimes preferable when available, such as the identification of the key primary variable (multiple variables), the choice of a critical treatment contrast (multiple comparisons) … In confirmatory analyses, any aspects of multiplicity which remain after steps of this kind have been taken should be identified in the protocol; adjustment should always be considered” **= Solutions, pre-specification and ‘adjustment’**

Guidelines – Multiplicity issues in clinical trials

- "A clinical study that requires no adjustment of the type I error is one that consists of the two treatment groups, that uses a single primary variable, and has a confirmatory statistical strategy that pre-specifies just one single null-hypothesis relating to the primary variable and no interim analysis… all other situations require attention to the potential effects of multiplicity."

- The guideline describes matters for which multiplicity should be considered: endpoints, doses, timepoints, hypotheses, subgroups. Focus is at the level of the **trial**.
- Other topics: secondary endpoints, subgroups, 'responder analyses', composite endpoints

Guidelines – SmPC

- Clinical Efficacy and Safety
  - "It may be appropriate to provide limited information, relevant to the prescriber, such as the main results (**statistically compelling** and clinically relevant) regarding pre-specified end points or clinical outcomes in the major trials … Such information on clinical trials should be concise, clear, relevant and balanced …

- Paediatric population
  - "… the main results regarding pre-specified endpoints should be provided, **whether positive or negative**. If data are considered inconclusive, this should be stated."

General comments

- Regulatory philosophy
  - Ensure evidence of therapeutic efficacy
  - Ensure positive risk-benefit
  - Product information that is instructive and informative
  - Licensing decisions that are proportionate and scientific without compromising standards

- Since licensing (risk-benefit) decisions are holistic, statistical methods to handle of multiplicity are, formally, neither necessary nor sufficient but:
  - makes for a more simple life
  - facilitates planning for the sponsor and a more predictable regulatory review
  - enhances (statistical) strength of evidence

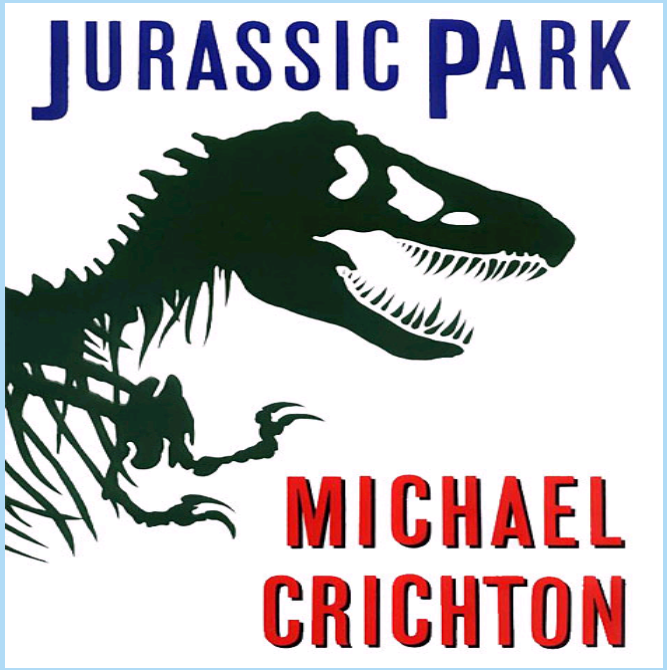
General comments

- Multiplicity is generally well-handled through pre-specification, hierarchy and adjustment

- Until things go wrong…and the potential for “Texas sharp-shooting” explains the emphasis on pre-specification and Type I error control

- This is not a controversial guideline, why think about changing it?
  - New strategies
  - Multi-regional drug development
  - Subgroups for confirmatory conclusions
  - Evidence at interim (or anytime nominal level adjusted?)
  - Simultaneous confidence intervals

Literature reference

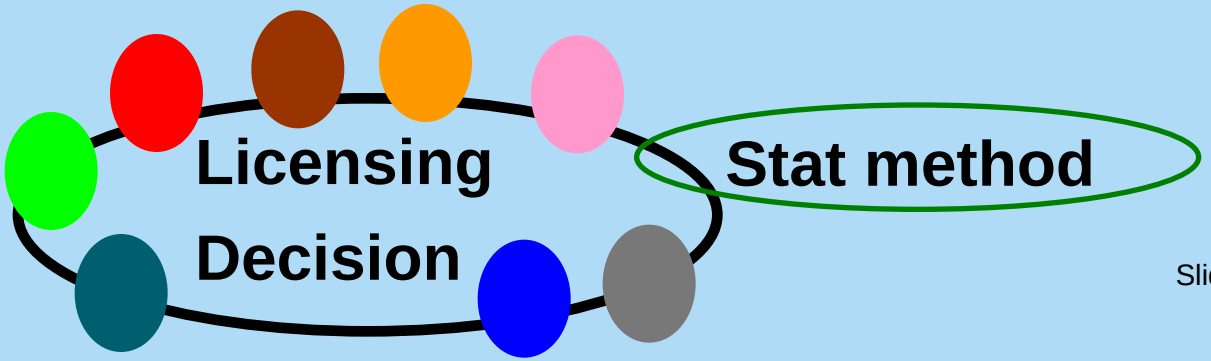


Dr Ian Malcolm (Jeff Goldblum)

> *"your scientists were so preoccupied with whether they **could** that they didn't stop to think if they **should**."*

<u>Important questions:</u>

1. **Must we address multiplicity? <u>OR</u>**
2. **Is it preferable to address multiplicity? <u>AND ONLY THEN</u>**
3. **How to address multiplicity?**



Special topics – new strategies

- 1. ‘Using’ alpha
  - Brilliant
  - All advances welcome – where ‘must’ or ‘preferable’
- 2. Demonstrating control of alpha
  - Simulation

Special topics – secondary endpoints

- Secondary endpoints
  - Usually: ***variables expressing supportive evidence*** – no **claims** are intended – tests are exploratory
  - Sometimes: ***variables which may become the basis for additional claims***

- For discussion (though not here!)
  - Pulmonary Hypertension and Erectile Dysfunction
  - Treatment and Maintenance of effect
  - Disability progression in RRMS or Structural Damage in RA
  - SGRQ / TDI in asthma
  - ORR in last line oncological indication

Special topics – secondary endpoints

- SmPC section 4.1 vs 5.1 – helps

- Decision-making is not only statistical
  - Totality of data
  - Pharmacology / “biological plausibility” / External evidence
- Scenario 1 - New pharmacological class, uncertain mechanism, single pivotal trial
- Scenario 2 - Known pharmacological class, clear mechanism, multiple pivotal trials

- Increases (statistical) strength of evidence

- Should consideration of **need to adjust** and **how to adjust** go beyond the data from the trial? Communication between disciplines to be improved? **Are hierarchies non-sensical?**

Special topics – interim decisions

- Trial stops for ‘efficacy’ at p=0.001 based on primary endpoint
  - Should consider more but that’s another talk…

- How to quantify effects
  - 95% CIs
  - 99.9% CIs

- What about secondary variables?

- More complex for stopping at 2nd, 3rd interim analysis?

Special topics – multi-regional development

• 1. Investigating regional differences
  - Discussed elsewhere under ‘subgroups’

• 2. Addressing different requirements / preferences between regions
  - Conduct separate development programmes, or plan a programme with several pivotal trials which individually prioritise the standards for different regions.
  - Increase the size of the trial so that it can simultaneously achieve all regulatory requirements within one framework without loss of power
  - Choose one set of regulatory standards over others for the purpose of planning and reporting; possibly, the highest hurdle
  - Develop **multiple SAPs** for a single trial, so that the same trial data separately analysed will address the different standards of each region.

Special topics – multi-regional development

- Multiple SAPs = multiple chances to win.
- Not in EU – still one chance to win
- Not without problems…

Assessment

- Strict on confirmatory, ‘scientific’ on exploratory
- Secondary endpoints treated as supportive

- Lower standards?
  - No, balance of data, science, pharmacology, communication
  - Nominal significance levels are important for planning and analysing RCTs
  - Matters important for prescribers are not only those with statistical significance proven to the highest standards
  - Need to be better at describing uncertainty?

- Marketing authorisation, meta-analyses, safety data

- Multiple SAPs – have been accepted

Scientific Advice

- Multiplicity is generally well-handled through pre-specification, hierarchy and adjustment

- Multiple SAPs
  - Accepted if the best solution

- Complex adjustments
  - “Is our complex adjustment acceptable?” – Yes.
  - “Is our adjustment necessary” – good question!

Concluding remarks

- Proper interpretation of MA, trial, endpoint all requires awareness of this multiplicity concept – but this is not a statistical problem – it is instead a problem for inference, and hence design, that can be addressed through statistical methods. Solutions must respect context.

- Plenty of room for further discussion, e.g.
  - Type I error control ‘validated’ by simulation
  - What constitutes an ‘additional claim’?

- New methods to be understood and integrated, but don’t be so preoccupied with whether you could that you don’t stop to think about whether you should.