ICH-M15 for Bayesian Modelling

A systematic model assessment framework to support design submission and discussion

EFSPI RegStat, 12th September 2025

Tobias Mielke, Fei Chen (Johnson & Johnson)
Mouna Akacha, Simon Newsome (Novartis)
Cornelia-Ursula Kunz, Christian Stock (Boehringer-Ingelheim)

EFSPI
EUROPEAN FEDERATION OF STATISTICIANS IN THE PHARMACEUTICAL INDUSTRY
Representing Statistical Associations in Europe

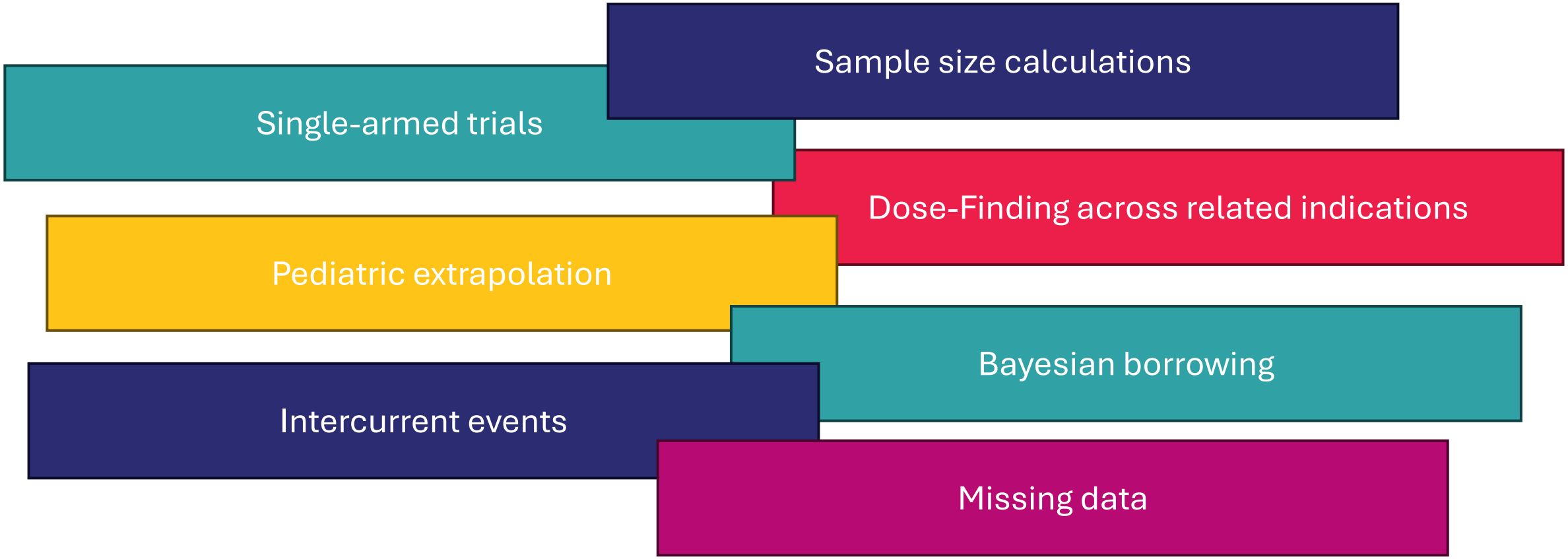
Modelling and Assumptions in DD

Absence of randomized data does not mean that we know nothing.

Discussion of assumptions and uncertainty in those can be highly informative.

Transparency on assumptions is required for understanding of risks and their likelihood

Assumptions in Drug Development



How to deal with assumptions?
How to deal with uncertainty?

Modelling and Assumptions in Drug Development

1. Drug development starts with assumptions, i.e.: “actionable science”
2. Many conventional statistical analyses integrate assumptions
   - *… which are not always fulfilled, yet still accepted*
3. Increasing transparency on assumptions through Estimands, Bayesian and Causal Inference frameworks
4. Increasing uptake of decision frameworks integrating quantification of uncertainty across Pharmaceutical Industry (i.e. elicitation & assurance)
5. ICH M15 to leverage learnings from assumption-rich disciplines (e.g. PMX)

Common situation in Drug Development?

1. Historical Placebo and/or active control data available
2. New trial to establish PoC, Dose-Response or confirm efficacy
3. How much concurrent control data do we need to accrue?

What are the relevant risks? What are consequences of wrong decisions?

What information is required to limit those risks?

Given current knowledge, how likely would learnings alter our decision?

Disclaimer: OS Borrowing & ICH-M15

OS Borrowing Discussion at EMA Bayes Workshop (June 2025):

1. OS Borrowing considerations were submitted in the design stage of two parallel contemporaneous development programs.

2. A structured model and assumption assessment framework, such as ICH-M15, was not utilized during the submission process.

3. Both development programs subsequently changed and the option of OS control data borrowing across them was no longer of relevance.

Would ICH M15 have supported the submission? **Yes.**

Would it have changed the outcome? **Likely not.**

Motivation: Borrowing OS data in Oncology

Two parallel contemporaneous development programs:

- Same indication
- Largely overlapping populations
- Partially same comparator (single comparator “A” vs investigator’s choice of “A or B”)

Independently designed trials with PFS as primary endpoint:

- OS key secondary endpoint, but last in testing sequence
  - Median OS in the targeted patient population ~40m
  - Studies sized primarily for PFS with OS HR assumption slightly more conservative

  → Long duration to reach enough events for adequate power on OS

Parallel positive information on the Roche CID-pilot on Lymphoma:

“When to implement borrowing, if not for this case?”

What approach was taken?

1. Systematic evaluation of Bayesian dynamic borrowing methods for the given case.
2. Internal strategic discussions on how to move ahead:
   - Why OS only? Why not also PFS?
   - Go bold, proposing full pooling of matched control data?
   - How much details to provide on statistical properties?
3. **Submitted briefing book for SA targeting OS control data borrowing only:**
   - **Justification:** Rationale, designs, eligbility criteria, assumptions
   - **Methods:** Pocock criteria, propensity score weighting & Bayesian methodology
   - **Properties:** Type-1 error vs. power (simulation results)

Regulatory Feedback Received

FDA and EMA did not accept the proposed Bayesian borrowing as primary analysis approach to OS after PFS success:

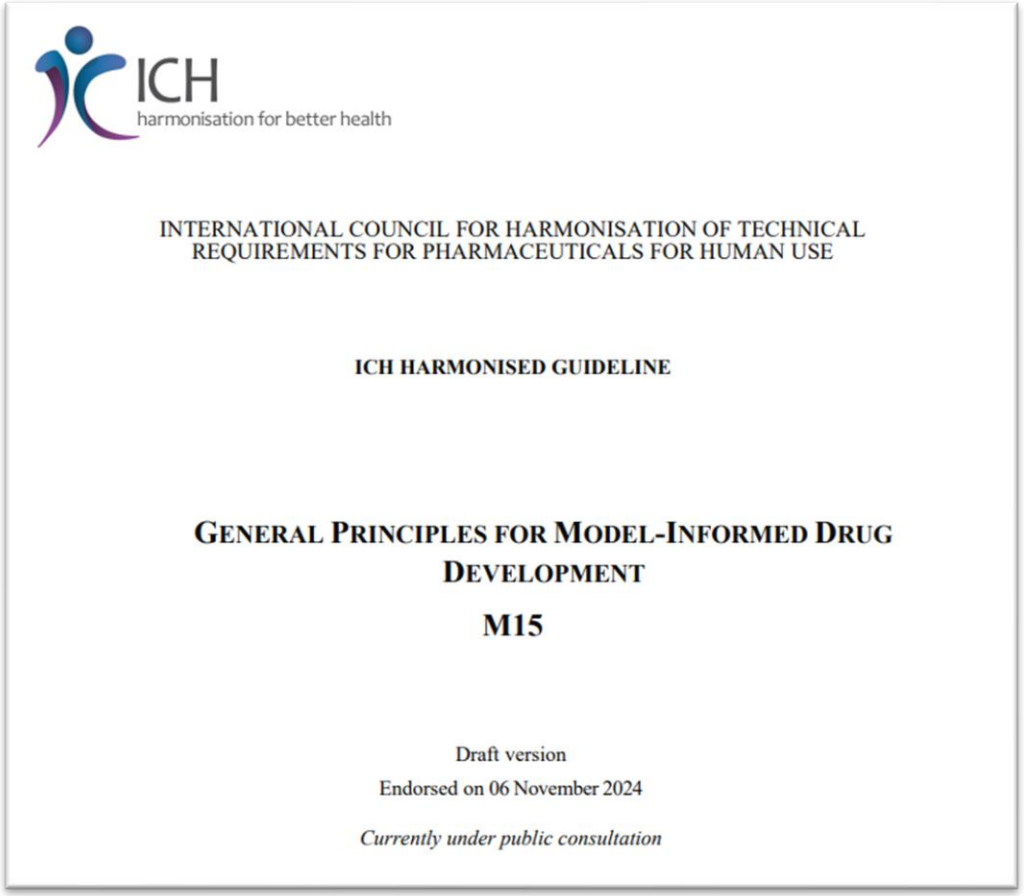
1. OS may be impacted by subsequent treatments, which could potentially differ.
2. Potential differences in patient populations and response, which may not be controlled for.
3. Increased risk of multiple false positives by relying on same control data across two trials.

Similar considerations apply likely to most borrowing cases.

How could those concerns be addressed?

Structured Modelling Assessment Framework

Let‘s learn from MIDD…



ICH
harmonisation for better health

INTERNATIONAL COUNCIL FOR HARMONISATION OF TECHNICAL REQUIREMENTS FOR PHARMACEUTICALS FOR HUMAN USE

ICH HARMONISED GUIDELINE

GENERAL PRINCIPLES FOR MODEL-INFORMED DRUG DEVELOPMENT

M15

Draft version

Endorsed on 06 November 2024

Currently under public consultation

Generic & structured framework for

- planning regulatory interactions on model-informed drug development
- implementation, reporting and submission of modelling results

What specifically?

- Clear structure for defining modelling approaches
- Defining problem, situation and assumptions
- Spelling out modelling impact & risk
- Verification & validation steps

Enabling transparent & constructive discussions?

Assessment elements

- Question of interest: **What specific problem do we want to address?**
- Context of use: **In exactly which situation / circumstances?**
- Model influence: **Weight of model in decision making vs. other relevant information**
- Consequence of wrong decision: **With respect to patient safety/efficacy**
- Model risk: **Contribution of model outcomes to wrong decision.**
- Model impact: **Contribution of model outcome vs. current regulatory practice**

No surprises:
Obvious specification steps when deciding on implementation of innovation

Model risk & Model Impact

„Potential differences in patient populations and response, which may not be controlled for.

Increased risk of multiple false positives by relying on same control data across two trials.“

What would be consequences of a wrong decision *due to the model*?

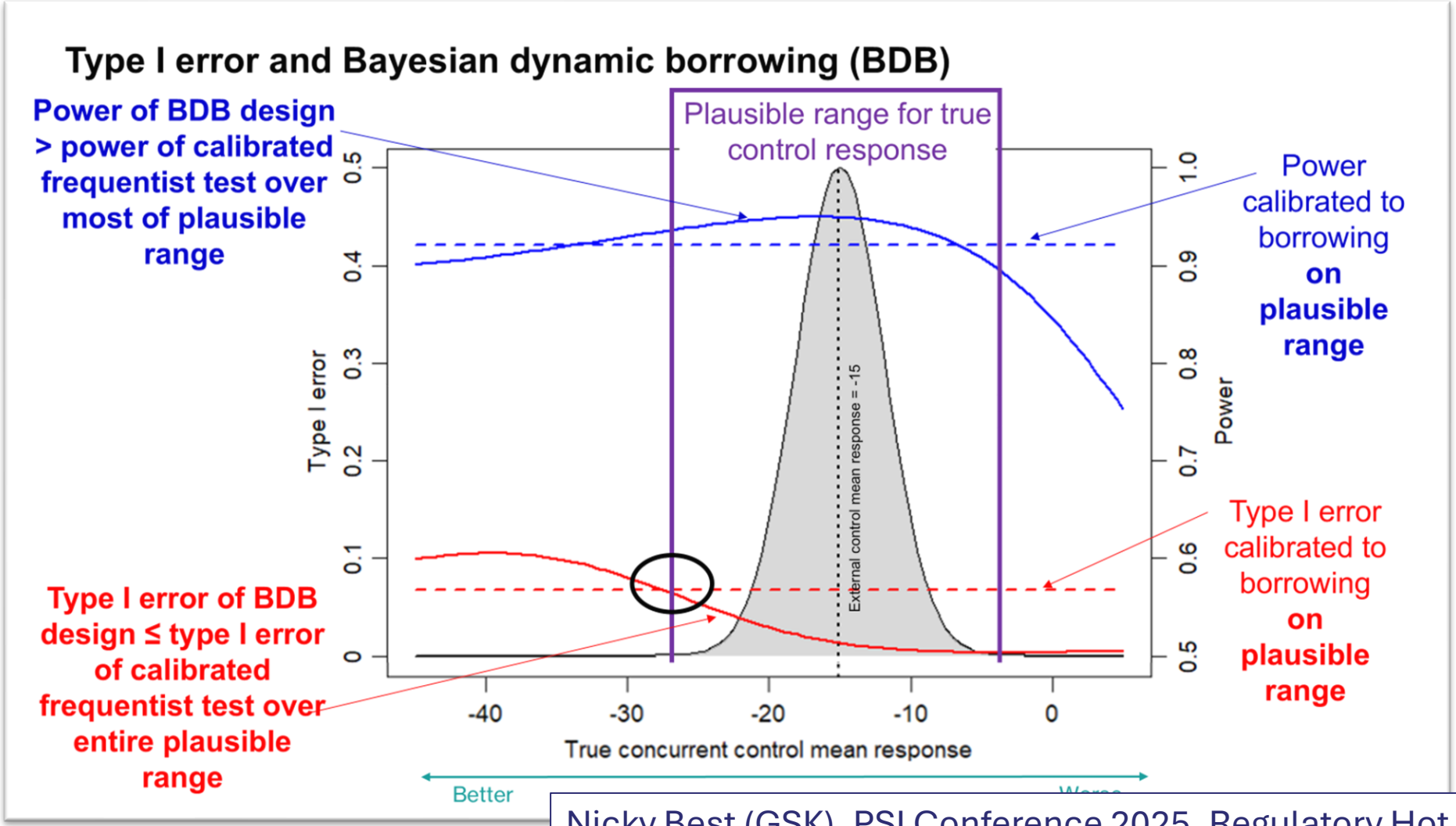
- Limited consequence to approval, as this would be granted based on PFS.
- False superiority claim on OS captured in the label

What would be consequences of a correct decision due to the model?

- More patients treated with drug improving survival
- Inform on detriment or absence of survival benefit
- Inform decision in presence of intercurrent events (e.g. treatment switching)

Model risk & Model Impact - Continued

… and what is the likelihood of wrong decisions?



Additional considerations

Appropriateness of proposed MIDD: Why is the model suitable?

- Effect on (related) primary endpoint established based on stand-alone data.
- Majority of Pocock criteria fulfilled.
- Populations made comparable using propensity score weighting.
- Bayesian dynamic borrowing robustifies analyses against potential differences in response.

Technical criteria: Key criteria for evaluation of acceptability

- Maximum acceptable weight of borrowed data vs. stand-alone evidence?
- Acceptable level of type-1 error for specific *(unlikely)* difference in control mOS?
- Required level of evidence from stand-alone data?
- Similarity of concurrent control data on multiple endpoints, i.e. ORR, PFS and OS?

Model evaluation

Model evaluation: verification, validation & applicability assessment:

Model is going to be thoroughly assessed upon availability of data using pre-specified sensitivity analyses:

Model not applicable or not valid? → Model not fit for decision making

What are the criteria for verification, validation and applicability?

- Study design could be optimized to increase probability of “evaluation success”, e.g.:
  - Definition of inclusion & exclusion criteria, assessment schedule
  - Determination of sample size & allocation ratio
- Upfront discussion & agreement to maximize chance of modelling success.

ICH-M15 to support statistical modelling/assumptions

- Increased transparency on assumptions and approaches to validate those
- Structured approach for discussion and transparency on concerns.
- No free pass for all cases – but context of use specific.
- Broadly applicable to Bayes and novel statistical methods in general

Larger emphasis on risk management vs. strict type-1 error?

Why to trust only the limited in-trial data rather than totality of data?

Thank you for the opportunity to present and discuss ICH-M15 in the context of Bayesian Statistics

Assessment Framework Overview

Table 1: Guideline Overview: Sequence of MIDD in Relation to the Relevant Guideline Sections

Stages	Planning and Regulatory Interaction		Implementation, Reporting, and Submission		
Sequence of Activities	**Key Assessment Elements**	**Additional Considerations for Interaction with Regulator and to Inform Decision-Making**	**Model Evaluation**	**Model Analysis Reporting**	**Documentation for Regulatory Interactions and Submissions**
	• Question of Interest • Context of Use • Model Influence • Consequence of Wrong Decision • Model Risk • Model Impact	• Appropriateness of Proposed MIDD • Technical Criteria for model evaluation and model outcomes[1] These should be documented (e.g., in a Model Analysis Plan [MAP]).	• Verification • Validation • Applicability assessment	• Model Analysis Report(s) (MAR)	• Regulatory documents, including + Outcome of MIDD Evidence Assessment + References to all relevant MAPs and MARs
Relevant Guideline Section	Section 2.1 and Appendix 1	Sections 2.2 and 4.1 and Appendix 1	Section 3	Section 4.2 and Appendix 2	Sections 2 and 4.3 and Appendix 1

Note: Terms used in this table are defined in relevant guideline sections.

[1] Results derived from M&S (i.e., via model-based predictions or simulations) and associated conclusions that are typically aligned to a Question of Interest.

Inform Decision-Making

Design stage

Analysis stage