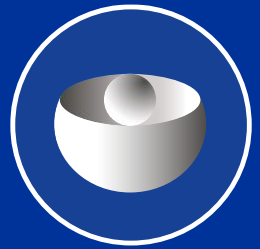


EUROPEAN MEDICINES AGENCY

Type-I-error rate inflation in mixed models for repeated measures caused by ambiguous or incomplete model specifications

PSI Journal Club, April 2024

Presented by Florian Lasch
Joined work with Sebastian Haeckl and Armin Koch, Hannover Medical School

Disclaimer

The views expressed in this presentation and in the following panel discussion are the personal opinion of the author and should not be understood as being made on behalf of or reflecting the position of the EMA or one of its committees or working parties.

Agenda

- Role of pre-specification in regulatory decision-making
- Pre-specification in practice
- Impact of model ambiguity on the type I error rate

The role of pre-specification in regulatory decision making

- context / decision dependent
  - Marketing authorisation
  - Demonstrating a significant benefit against satisfactory methods (treatments) for maintaining an orphan designation
- adds evidentiary weight to the findings
- less important in earlier phases
- critical for confirmatory claims
- Not only for classical statistical methodology, see [Reflection paper on the use of artificial intelligence in the lifecycle of medicines | European Medicines Agency (europa.eu)]

Pre-specification for confirmatory trials / analyses

"To **avoid concerns over data-driven selection of methods**, it is essential to pre-specify the selected methods in the statistical section of the study protocol or analysis plan (…)" (p. 6)

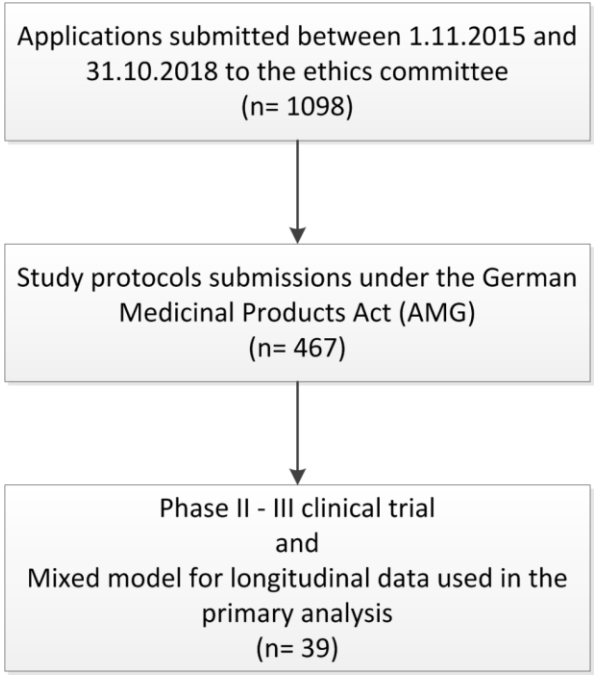
"Therefore, the precise option settings must be fully justified and predefined in advance in detail, so that the **results could be replicated**, if required, by an external data analyst and so that it can be established that **the choice has not been made post hoc**." (p. 10)

European Medicines Agency (EMA). Guideline on missing data in confirmatory clinical trials. 2011; 1–12.

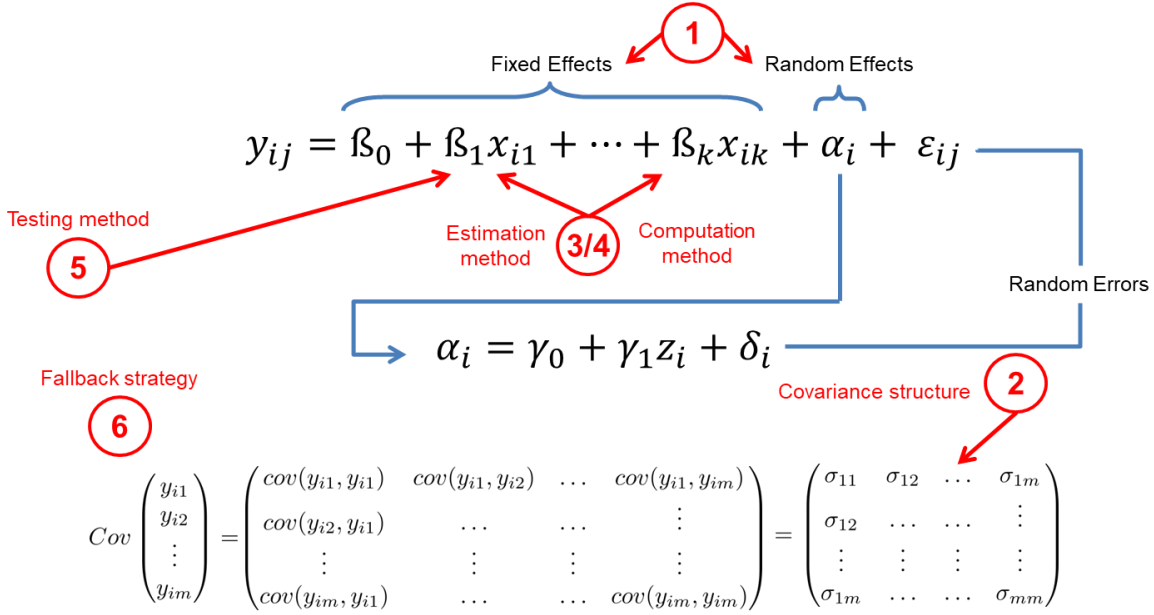
Pre-specification in practice

Empirical study on model specifications

- Focus on MMRMs
- Access to study protocols granted by MHH ethics committee



Mixed models for longitudinal data



Verbeke G, Molenberghs G. Linear Mixed Models for Longitudinal Data. Verlag New York, LLC: Springer; 2009.

Empirical study

TABLE 2 Reporting of primary mixed model analyses by clinical development phase

Evaluation Item	Development Phase		All Studies
	II (n = 15)	III (n = 24)	(n = 39)
Fixed and random effects	15/15 (100.0%)	22/24 (91.7%)	37/39 (94.9%)
Covariance structure	13/15 (86.7%)	17/24 (70.8%)	30/39 (76.9%)
Testing method	5/15 (33.3%)	9/24 (37.5%)	14/39 (35.9%)
Estimation method	4/15 (26.7%)	7/24 (29.2%)	11/39 (28.2%)
Computation method	0/15 (0.0%)	1/24 (4.2%)	1/39 (2.6%)
Fallback strategy	1/15 (6.7%)	6/24 (25.0%)	7/39 (17.9%)
SAP reference	3/15 (20.0%)	9/24 (37.5%)	12/39 (30.8%)
All items specified[a]	0/15 (0.0%)	0/24 (0.0%)	0/39 (0.0%)
Main items specified[b]	5/15 (33.3%)	7/24 (29.2%)	12/39 (30.8%)
Main items specified[b] or SAP reference	7/15 (46.7%)	14/24 (58.3%)	21/39 (53.8%)

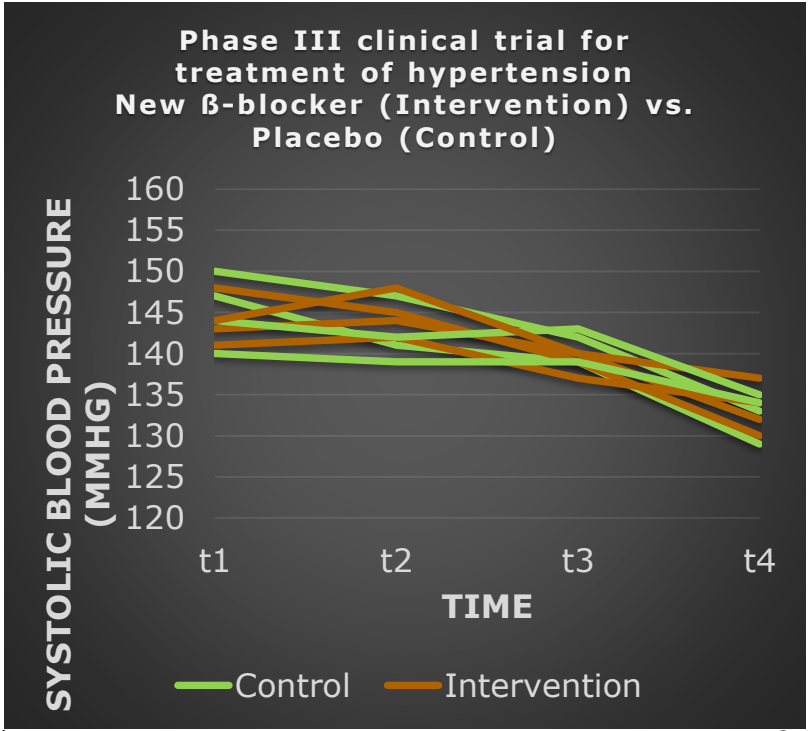
Häckl S, Koch A, Lasch F. Empirical evaluation of the implementation of the EMA guideline on missing data in confirmatory clinical trials: Specification of mixed models for longitudinal data in study protocols. *Pharmaceutical Statistics*. https://doi.org/10.1002/pst.1964

Impact of model ambiguity on Type I error

Häckl S, Koch A, Lasch F. Type-I-error rate inflation in mixed models for repeated measures caused by ambiguous or incomplete model specifications. Pharm Stat. 2023 Nov-Dec;22(6):1046-1061. doi: 10.1002/pst.2328

Motivational example

- Phase III clinical trial
- Estimand of interest:
- Effect of a new beta-blocker against a control in patients with elevated blood pressure on the difference in means in the change in systolic blood pressure at t4 as compared to baseline. (no IEs considered).



Simulation study – data generation

Simple data generating mechanism

General model $Y_i = X_i\beta + \varepsilon_i \qquad \varepsilon_i \sim N_i(0, R_i)$

Outcome measured 4 times

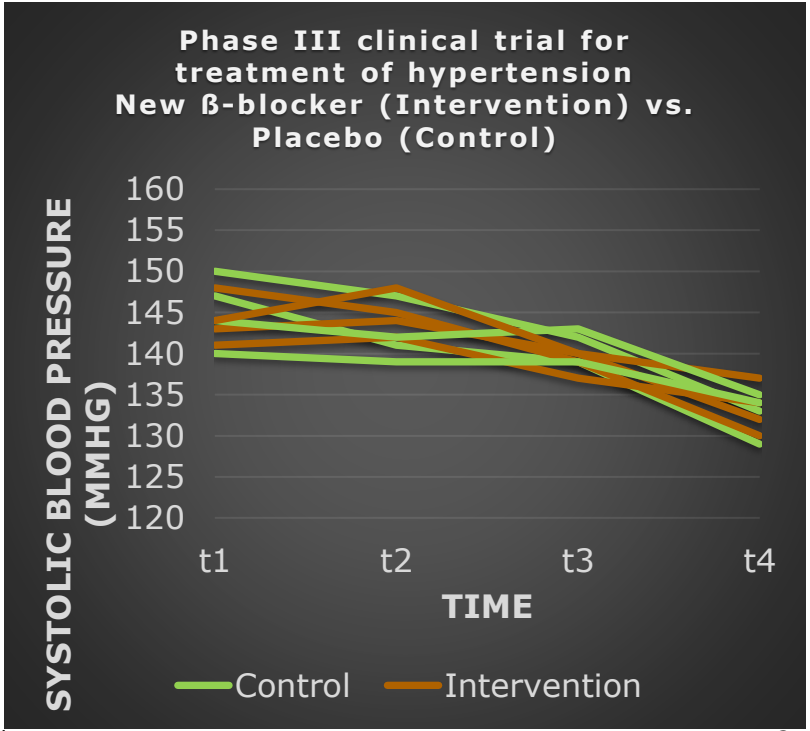
No treatment effect at any time point

Homogeneous compound symmetric variance-covariance structure to model within-patient errors (R_i)

Trial characteristics

- Sample size (50, 100, 500)
- Allocation ratio (1:1, 2:1)

→ 6 scenarios/trials, each with 10.000 simulations



Simulation study – MMRM specifications

Model parameter	Number of options	Precise specifications
Estimation method	4	ML, REML, Empirical sandwich estimation, MIVQUE(0)
Computational method	2	Newton-Raphson algorithm, Fisher scoring
Variance-covariance structure	7	Homogeneous CS, AR(1), TOEP, Heterogeneous CS, AR(1), TOEP Unstructured
Hypothesis testing	2	Wald test, LRT
DDFM estimation method	4	Residual, Between-Within, Satterthwaite, Kenward-Rogers

$$\alpha_{emp}(method_i) = \frac{\sum_{k=1}^{n} 1_{p_{i,k}<0.05}}{n}$$

indicator that the null hypothesis is rejected for specification **i** in the **k**-th simulation run

Clusters of analysis methods

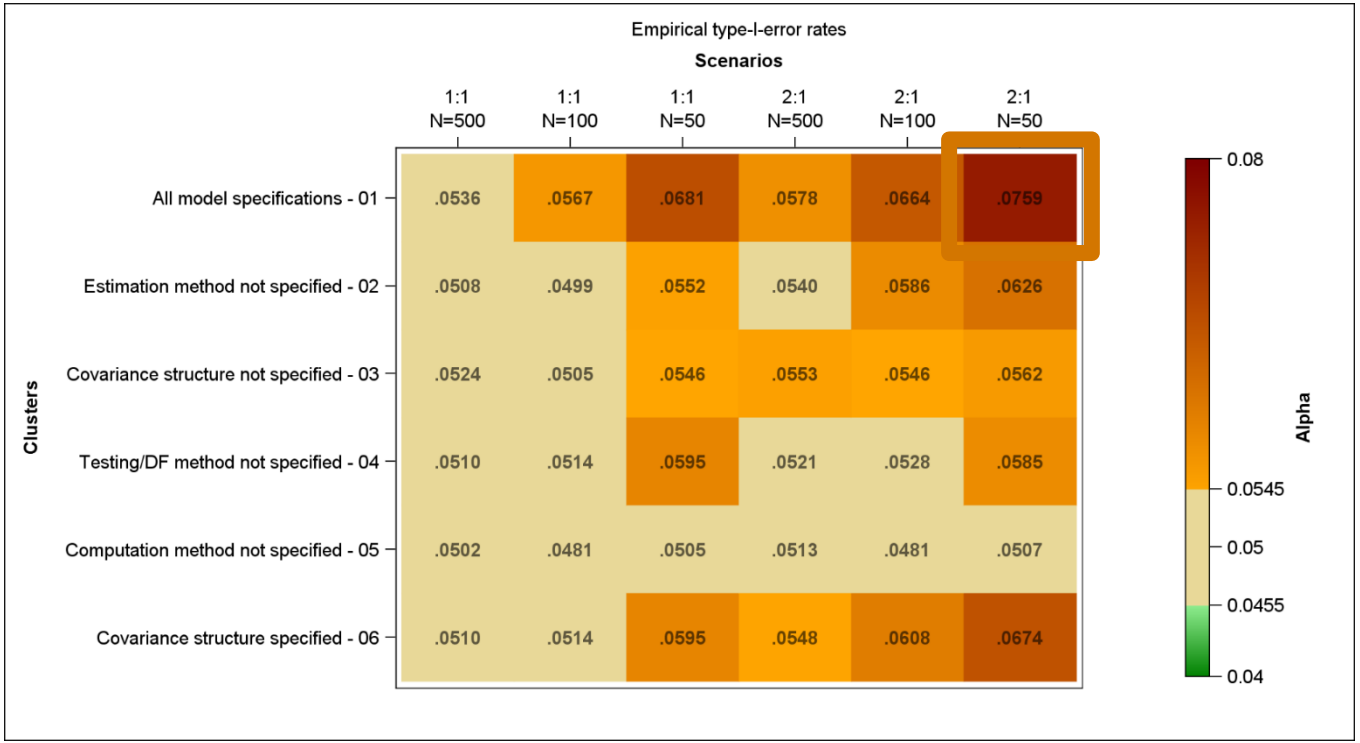
Model item	Model specifications (analysis methods) N=560					
	cluster 1 n=560	cluster 2 n=4	cluster 3 n=14	cluster 4 n=8	cluster 5 n=2	cluster 6 n=40
Estimation method	ns	ns	**REML**	**ML**	**REML**	ns
Covariance structure**	ns	**UN**	ns	**UN**	**UN**	**UN**
Testing method	ns	**WALD**	**WALD**	ns	**WALD**	ns
DDFM	ns	**BW**	**BW**	ns	**BW**	ns
Computation method	ns	**NR**	**NR**	**NR**	ns	ns

* For (pre-)specified covariance structures the same variance-covariance structure is assumed for each treatment arm.
DDFM = denominator degrees of freedom; ns = not (pre-)specified; REML = restricted maximum likelihood; ML = maximum likelihood; UN = unstructured pattern; WALD = Wald test; BW = Between-Within; NR = Newton-Raphson

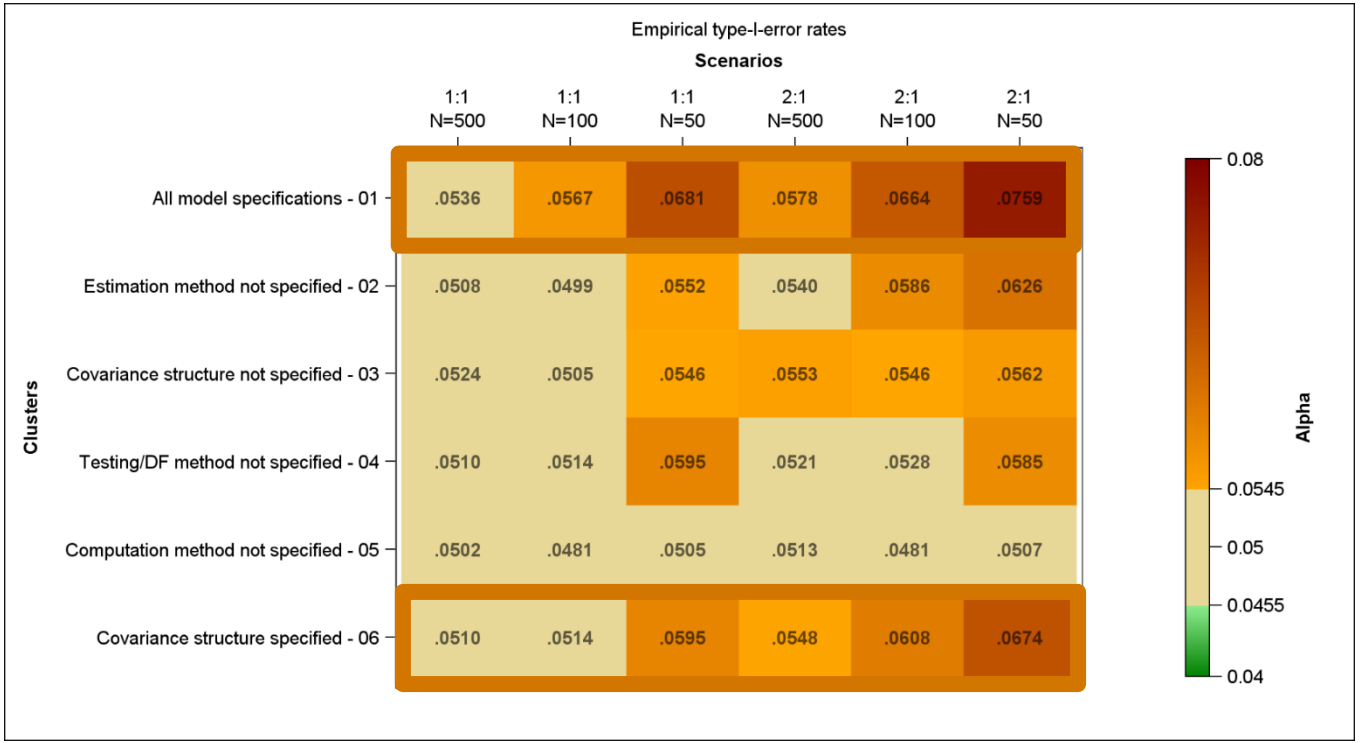
$$\alpha_{total}(cluster_j) = \frac{\sum_{k=1}^{n} 1_{p_{j,k}<0.05}}{n}$$

at least one of the contained models rejects the null hypothesis

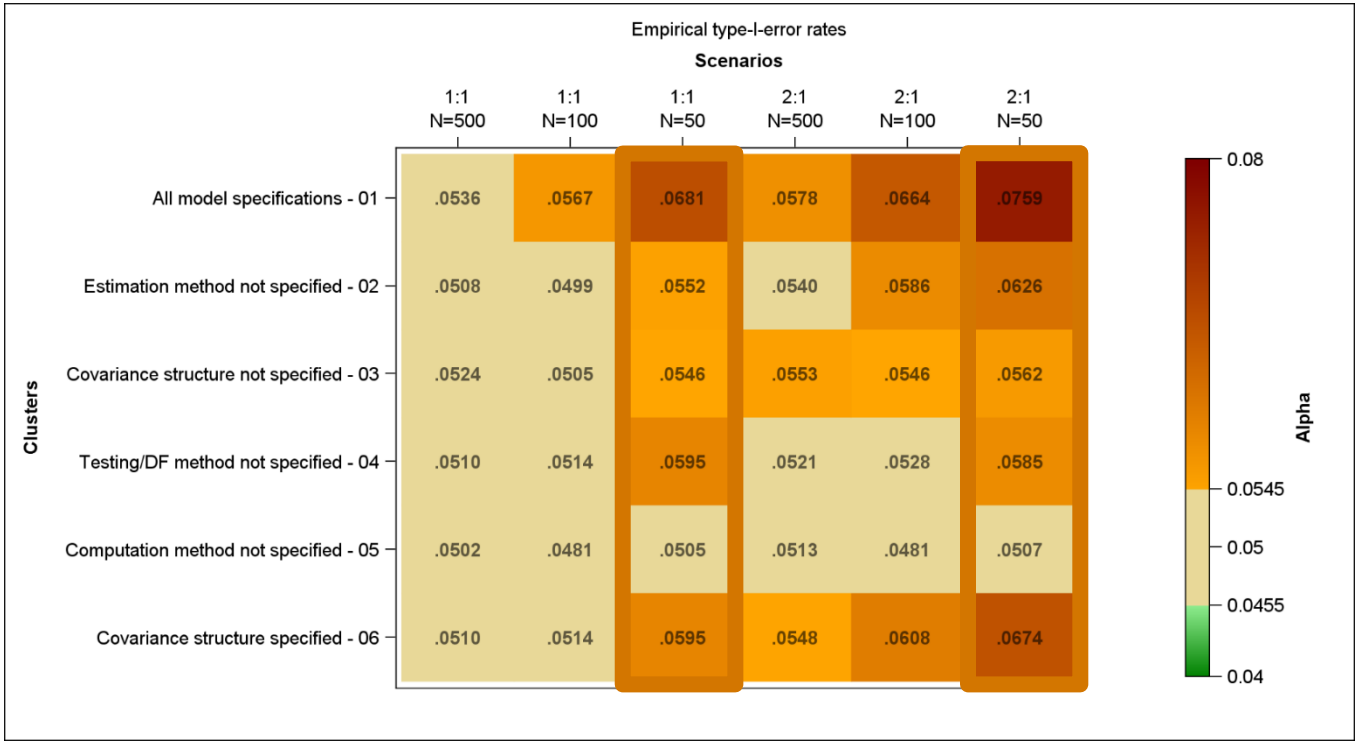
Simulation study – Results



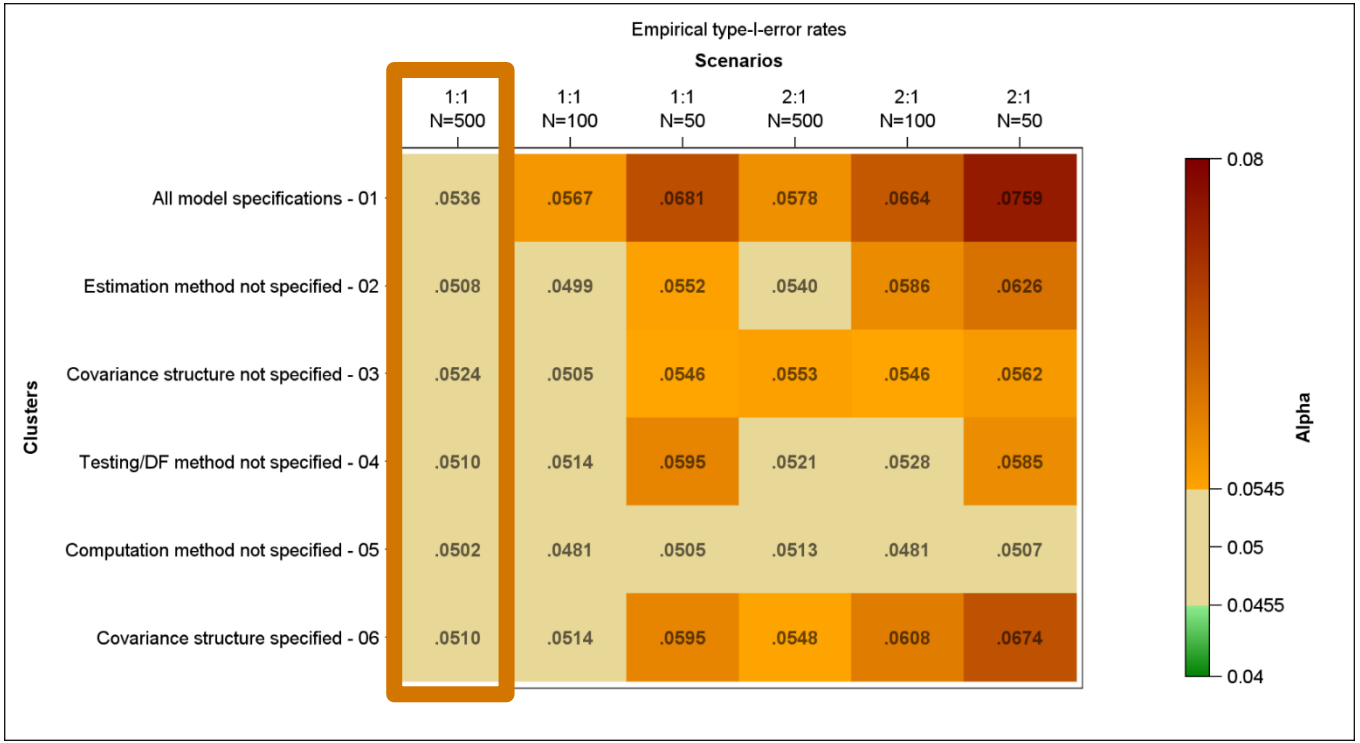
Simulation study – Results



Simulation study – Results



Simulation study – Results



Simulation study – discussion

- Ambiguous model specifications lead to Type I error rate inflation

- Our simulation study underestimates magnitude of the inflation
  - More options for model specification are possible than included in the simulation study

- The magnitude of the type I error rate inflation depends on:
  - The number of not specified model item
  - The type of not specified model item
  - The clinical setting (sample size, randomisation ratio, etc.)

- No post-hoc solution.

Conclusion

- Complex models are… complex
- A small differences in the specification can have a notable impact on the results
- Ambiguous specification leads to type I error rate inflation.

- Empirical study + simulation study:
  - The practice did not conform with the required standard.
  - It matters.

Any questions?

Further information

florian.lasch@ema.europa.eu

Official address Domenico Scarlattilaan 6 ● 1083 HS Amsterdam ● The Netherlands
Telephone +31 (0)88 781 6000
Send us a question Go to www.ema.europa.eu/contact

Follow us on **@EMA_News**

References

Häckl S, Koch A, Lasch F. Type-I-error rate inflation in mixed models for repeated measures caused by ambiguous or incomplete model specifications. Pharm Stat. 2023 Nov-Dec;22(6):1046-1061. doi: 10.1002/pst.2328. Epub 2023 Jul 30. PMID: 37519010.

Häckl S, Koch A, Lasch F. Empirical evaluation of the implementation of the EMA guideline on missing data in confirmatory clinical trials: Specification of mixed models for longitudinal data in study protocols. *Pharm Stat*. 2019;18(6):636-644. doi:10.1002/pst.1964

European Medicines Agency (EMA). POINTS TO CONSIDER ON MULTIPLICITY ISSUES IN CLINICAL TRIALS DISCUSSION IN THE EFFICACY WORKING PARTY. 2002.

European Medicines Agency (EMA). Guideline on Missing Data in Confirmatory Clinical Trials. 2011:1-12.

Empirical study – Baseline Characteristics

Trial characteristic	phase II (n=15)	phase III (n= 24)	all trials (n=39)
Sponsor			
Pharmaceutical company[#]	12 (80%)	23 (96%)	35 (90%)
Top 21 pharmaceutical company	9 (60%)	13 (54%)	22 (56%)
Planned sample size			
mean (±standard deviation)	141 (±125)	651 (±992)	455 (±815)
median (minimum, maximum)	99 (30, 500)	232 (15, 4126)	180 (15, 4126)
Therapeutic area*			
Blood or blood-forming organs	0 (0.0%)	3 (12.5%)	3 (7.7%)
Endocrine, nutritional or metabolic diseases	0 (0.0%)	2 (8.3%)	2 (5.1%)
Mental, behavioural or neurodevelopmental disorders	0 (0.0%)	2 (8.3%)	2 (5.1%)
Nervous system			
Visual system	0 (0.0%)	3 (12.5%)	3 (7.7%)
Ear or mastoid process	1 (6.7%)	2 (8.3%)	3 (7.7%)
Circulatory / cardiovascular system	1 (6.7%)	0 (0.0%)	1 (2.6%)
Respiratory system	0 (0.0%)	2 (8.3%)	2 (5.1%)
Digestive System	6 (40.0%)	6 (25.0%)	12 (30.8%)
Skin	3 (20.0%)	1 (4.2%)	4 (10.3%)
Immune System	1 (6.7%)	1 (4.2%)	2 (5.1%)
Genitourinary system	2 (13.3%)	1 (4.2%)	3 (7.7%)
Developmental abnomalies	1 (6.7%)	0 (0.0%)	1 (2.6%)
	0 (0.0%)	1 (4.2%)	1 (2.6%)

* ,Only ICD-11 [20] superior categories 01-20 are considered since categories 21-26 do not represent therapeutic areas
[#]: Including the top 21 pharmaceutical companies

Empirical study – results by sponsor type

Evaluation item	Sponsor type		
	Minor (n=22)	Major (n=17)	
Fixed and random effects	20/22 (90.9%)	17/17 (100.0%)	
Covariance structure	15/22 (68.2%)	15/17 (88.2%)	
Testing method	5/22 (22.7%)	9/17 (52.9%)	
Estimation method	3/22 (13.6%)	8/17 (47.1%)	
Computation method	0/22 (0.0%)	1/17 (5.9%)	
Fallback strategy	2/22 (9.1%)	5/17 (29.4%)	
SAP reference	9/22 (40.9%)	3/17 (17.6%)	
All items specified[+]	0/22 (0.0%)	0/17 (0.0%)	
Main items specified*	3/22 (13.6%)	9/17 (52.9%)	
Main items specified* or SAP reference	11/22 (50.0%)	10/17 (58.8%)	

[+] excluding reference to SAP

* Main items are fixed/random effects, covariance structure and testing method

[#]: p-value derived from Chi^2-test comparing proportions between sponsor types

[##]: p-value derived from Fisher's exact test comparing proportions between sponsor types