Boehringer Ingelheim

A meta-analytic approach to the assessment of tolerability in early phase oncology trials

PSI 2026 annual conference, 15 June 2026

Lorenz Uhlmann

Life forward

Disclaimer

- The views and opinions expressed in the following presentation are those of the individual presenter and not necessarily those of Boehringer Ingelheim Pharma GmbH & Co. KG.
- The analyses are exploratory, non-validated, and shown solely to illustrate the presented statistical methodology.
- Cohorts of the Beamion LUNG-1 trial are presented. However, clinical interpretation of the results is neither intended nor valid.

Background

Trial: Phase I trial assessing efficacy and safety of zongertinib in patients with advanced or metastatic solid tumors with HER2 aberrations (Beamion LUNG-1)

Data: The analysis is based on four Cohorts from the study (patients on 120 mg QD only):

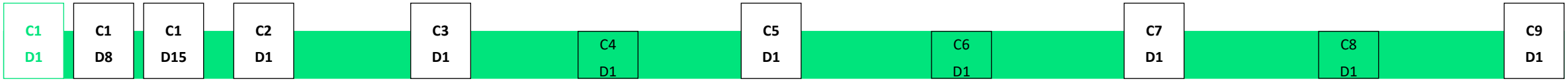
- **Cohort 1:** HER2 TKD mutation-positive non-squamous NSCLC previously treated (non-ADC treatment) (30 patients)
- **Cohort 2:** HER2 TKD mutation-positive non-squamous NSCLC first-line treatment (71 patients)
- **Cohort 4:** HER2 TKD mutation-positive NSCLC with active brain lesions, previously treated (30 patients)
- **Cohort 5:** HER2 TKD mutation-positive non-squamous NSCLC previously treated with HER2-directed antibody drug conjugates (31 patients)

Research question: Is there any substantial heterogeneity (i.e., difference) between cohorts in the patient-related outcome (PRO) EORTC IL46 (trouble with side-effects), overall and in the course over time, or can we pool data across all cohorts? Technical questions:

- How can we define heterogeneity in this setting?
- Which approaches can be used to detect such heterogeneity?

Boehringer Ingelheim

Patient-related outcome assessment schedule



PROs are scheduled to be administered to the patient prior to treatment.

- Cycle is 21 days

Tolerability PROs collected while on treatment at:

- Cycle 1: Days 1 (baseline)
- Cycle 1 day 8, cycle 1 day 15
- Cycle 2 day 1, cycle 3 day 1
- Cycle 5 day 1, cycle 7 day 1, cycle 9 day 1 (last cycle PROs collected)

- No PRO data collected post disease progression
- No PRO data collected post cycle 9

PRO of interest:

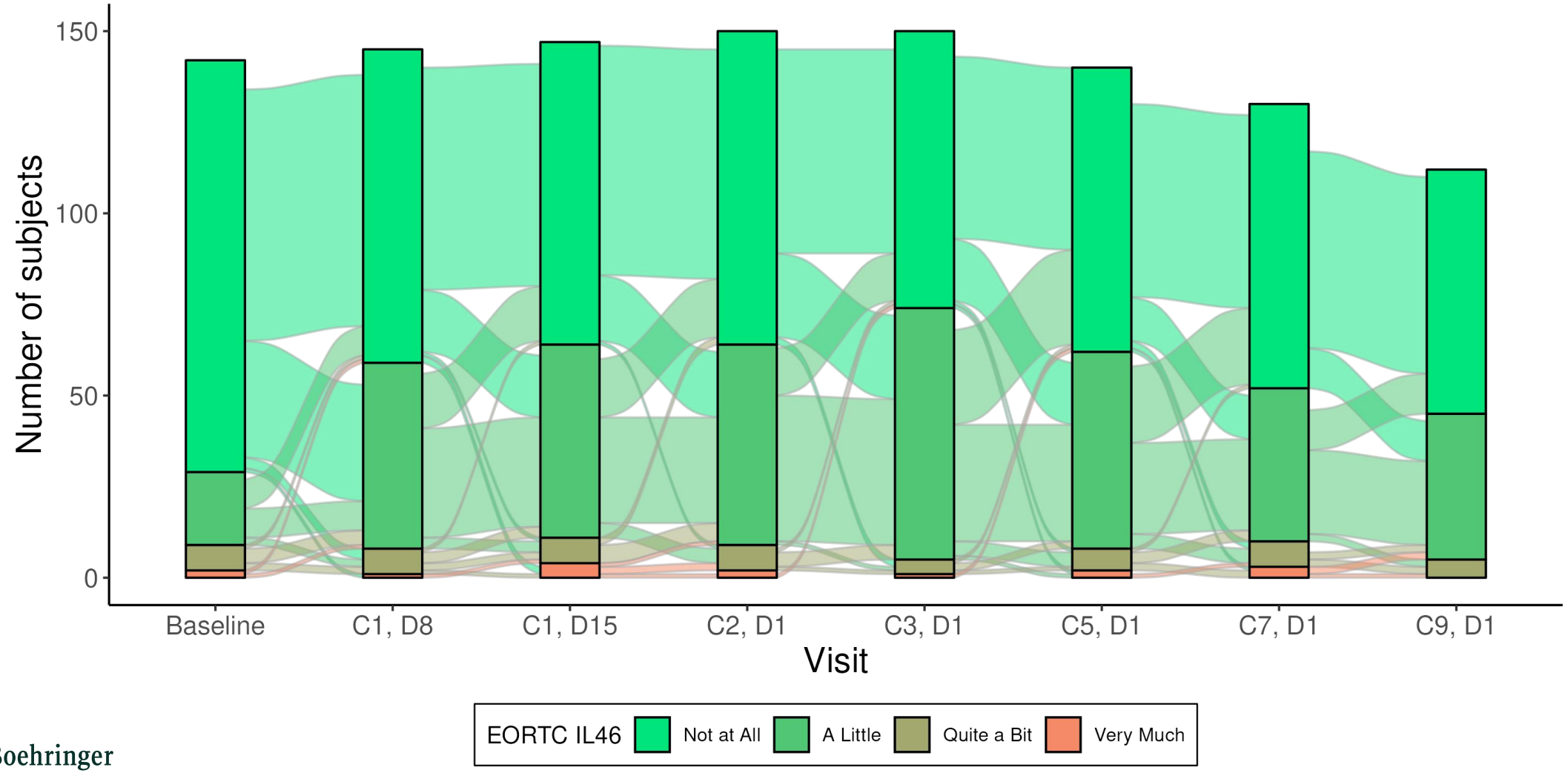
EORTC IL46: trouble with side-effects

During the past week:
To what extent have you been **troubled with side-effects** from your treatment?

➔ **Not at All / A Little / Quite a Bit / Very Much**

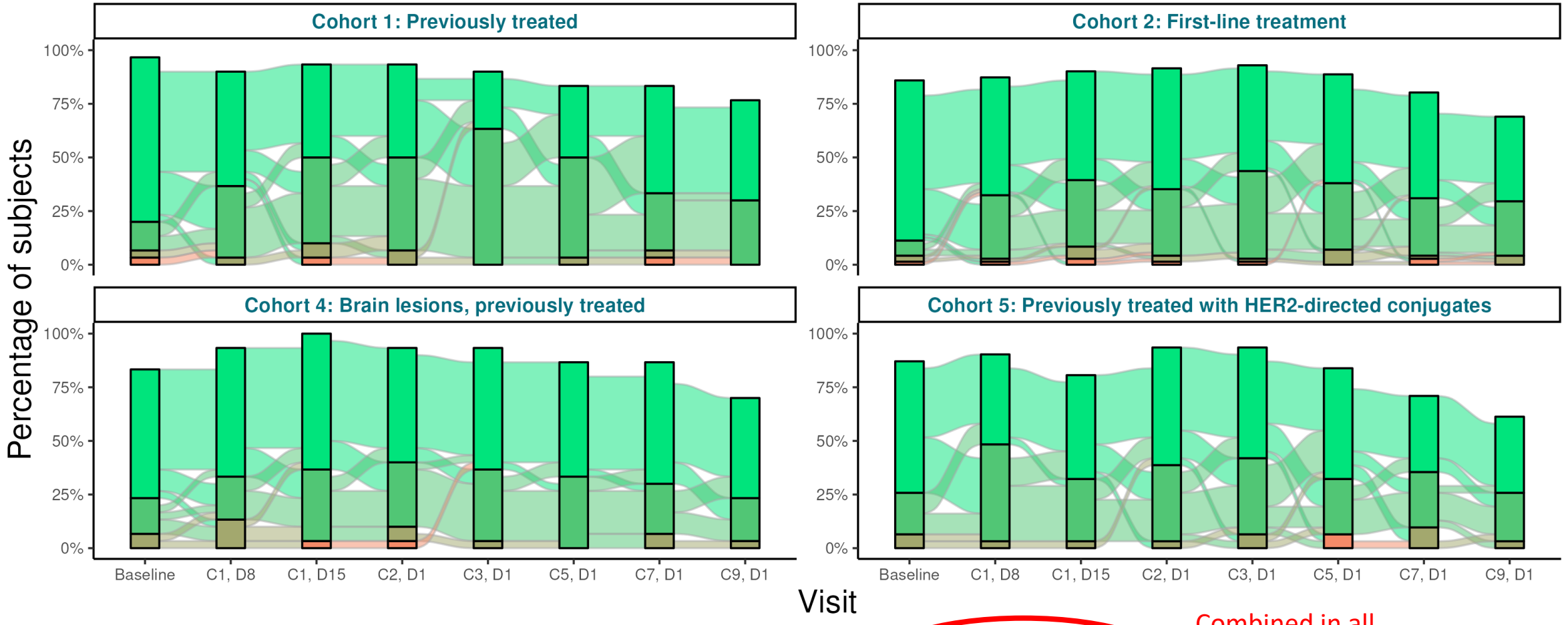
Boehringer Ingelheim

EORTC IL46 values over time



Boehringer Ingelheim

EORTC IL46 values over time – any difference between cohorts?

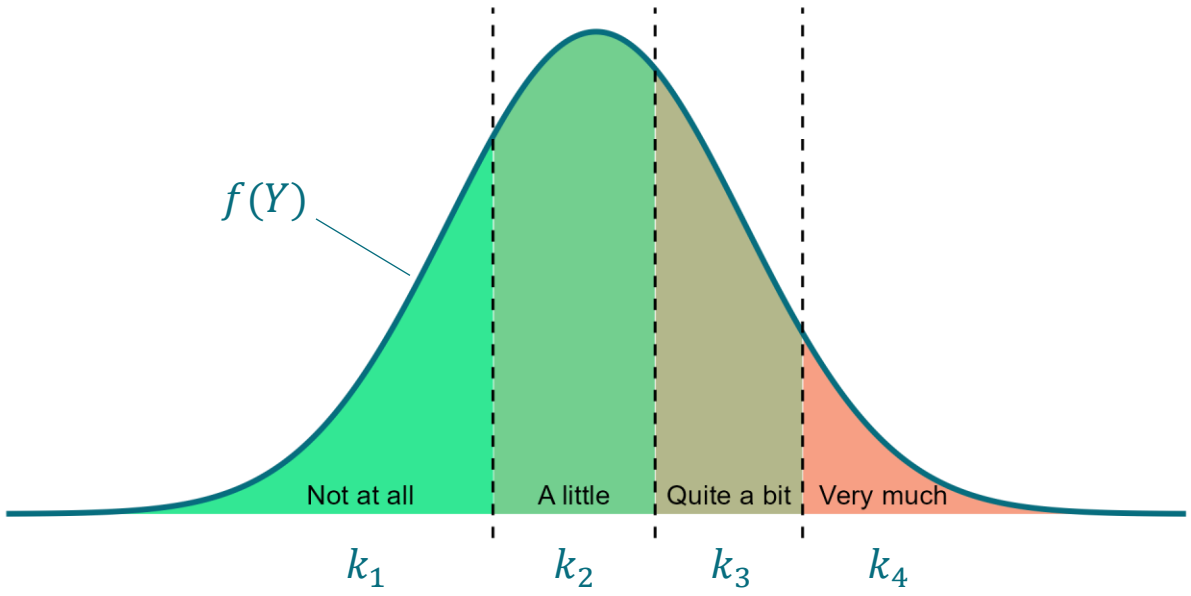


Combined in all models shown in this presentation

Boehringer Ingelheim

EORTC IL46 – modelling an ordinal outcome

Distribution of continuous variable Y spanning all categories of the ordinal variable



- Underlying latent variable Y follows a continuous distribution. Categories are modelled as

$$P(Y \leq k|x) = F(\alpha_k + \boldsymbol{x\beta}).$$

- Apply cumulative logit model:

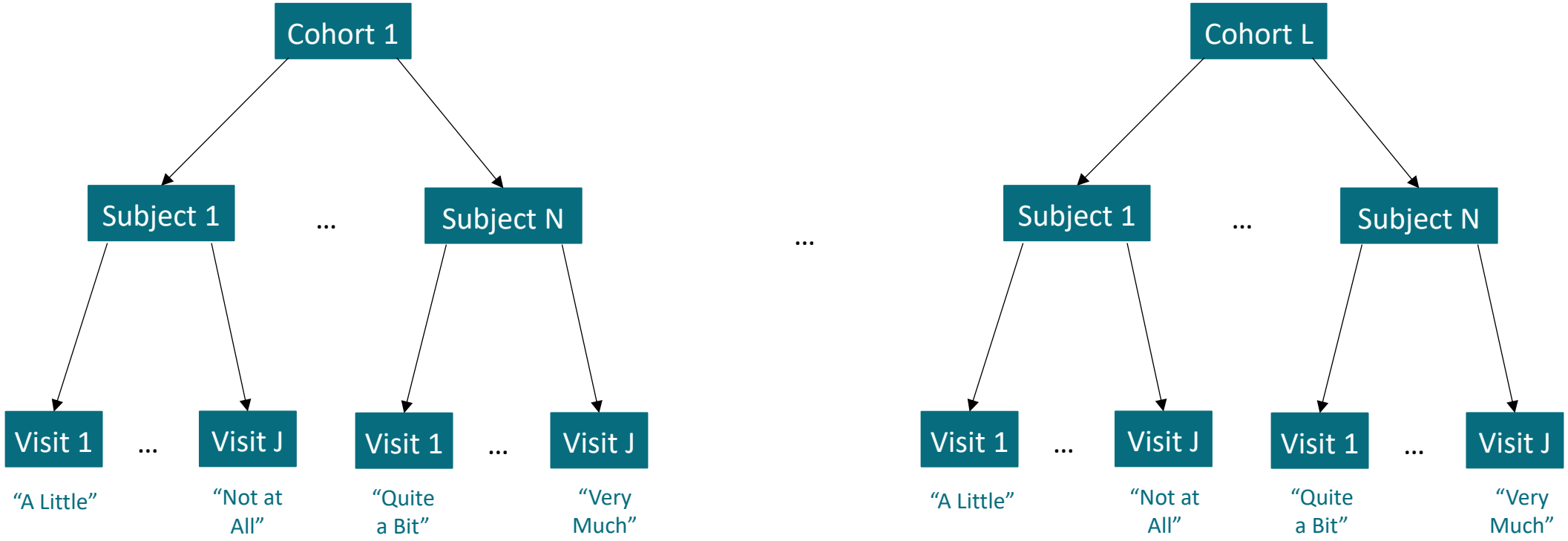
$$\log\left(\frac{P(Y \leq k|\boldsymbol{x})}{P(Y > k|\boldsymbol{x})}\right) = \alpha_k + \boldsymbol{x\beta}$$

- This implies proportional odds, i.e., an "equal shift" based on changes in predictors $\boldsymbol{x}$, characterised by the regression parameter $\boldsymbol{\beta}$ across categories.
  Each category comes with its own intercept α_k.

See, for example, Tutz (2012).

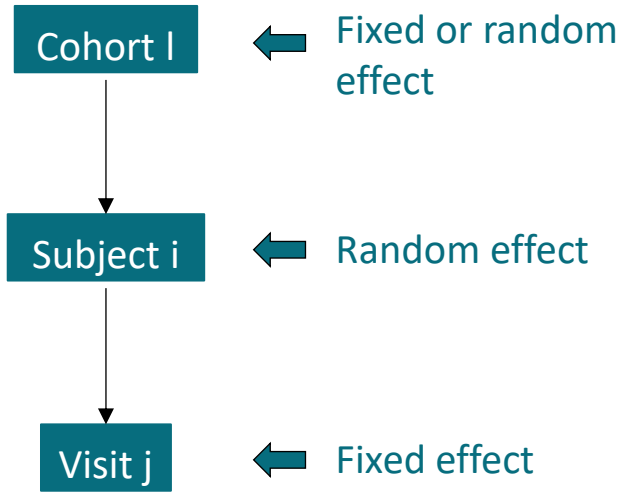
Boehringer Ingelheim

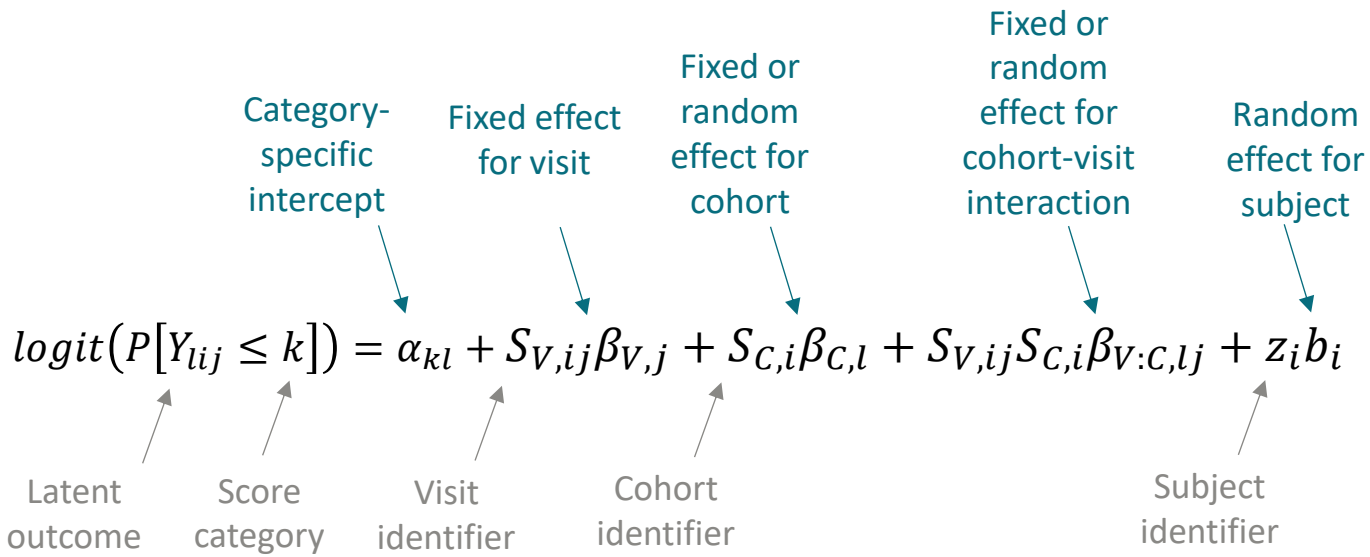
Hierarchical data structure



Model definition (Generalized Mixed Model)

- Hierarchical data structure needs to be accounted for.
- Effects for cohort, subject, and visit needs to be included.
- Interaction term for cohort and visit may also be of interest.





The ideas of model implementations shown on the next slides are heavily based on Whitehead et al. (2001).

Model overview

Heterogeneity I AL1

Generalized mixed model with fixed or random cohort effect

→ assumes equal shift between response categories over time

Heterogeneity II

Generalized mixed model with cohort-specific intercept

→ accounts for different response distributions between cohorts

Heterogeneity III

Generalized mixed model with interaction term of cohort and visit

→ checks for difference in shift over time across cohorts

Heterogeneity IV

Bayesian meta-analytic approach

→ ranks cohorts with respect to transition odds

Model overview

Heterogeneity I

Generalized mixed model with fixed or random cohort effect

→ assumes equal shift between response categories over time

Heterogeneity II

Generalized mixed model with cohort-specific intercept

→ accounts for different response distributions between cohorts

Heterogeneity III

Generalized mixed model with interaction term of cohort and visit

→ checks for difference in shift over time across cohorts

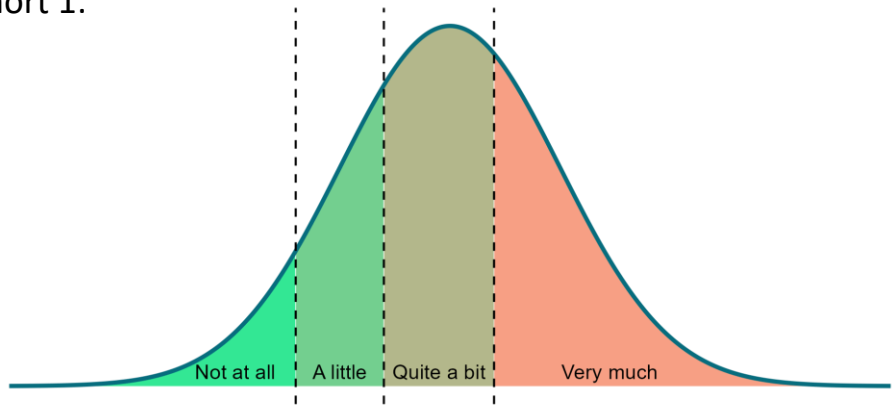
Heterogeneity IV

Bayesian meta-analytic approach

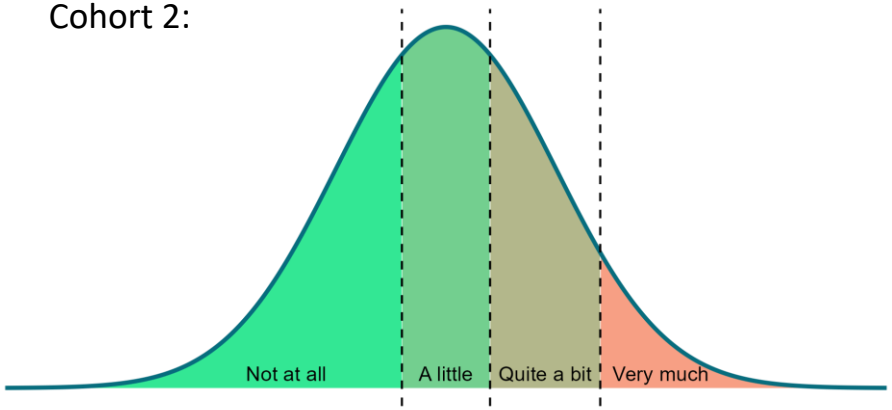
→ ranks cohorts with respect to transition odds

Heterogeneity I – fixed cohort effect

Cohort 1:



Cohort 2:



Example: general shift from Cohort 1 to Cohort 2

➔ Subjects in Cohort 2 are more likely to be less burdened.

Cohort effect, equal shift across all categories

Applied model: $$logit\left(P\left[Y_{lij} \leq k\right]\right) = \alpha_k + S_{C,i}\beta_{C,l} + S_{V,ij}\beta_{V,j} + z_i b_i$$

⇨ Likelihood ratio test

Boehringer Ingelheim

Heterogeneity I – applied to EORTC IL46 data

Result of the likelihood ratio test: **p=0.4518**

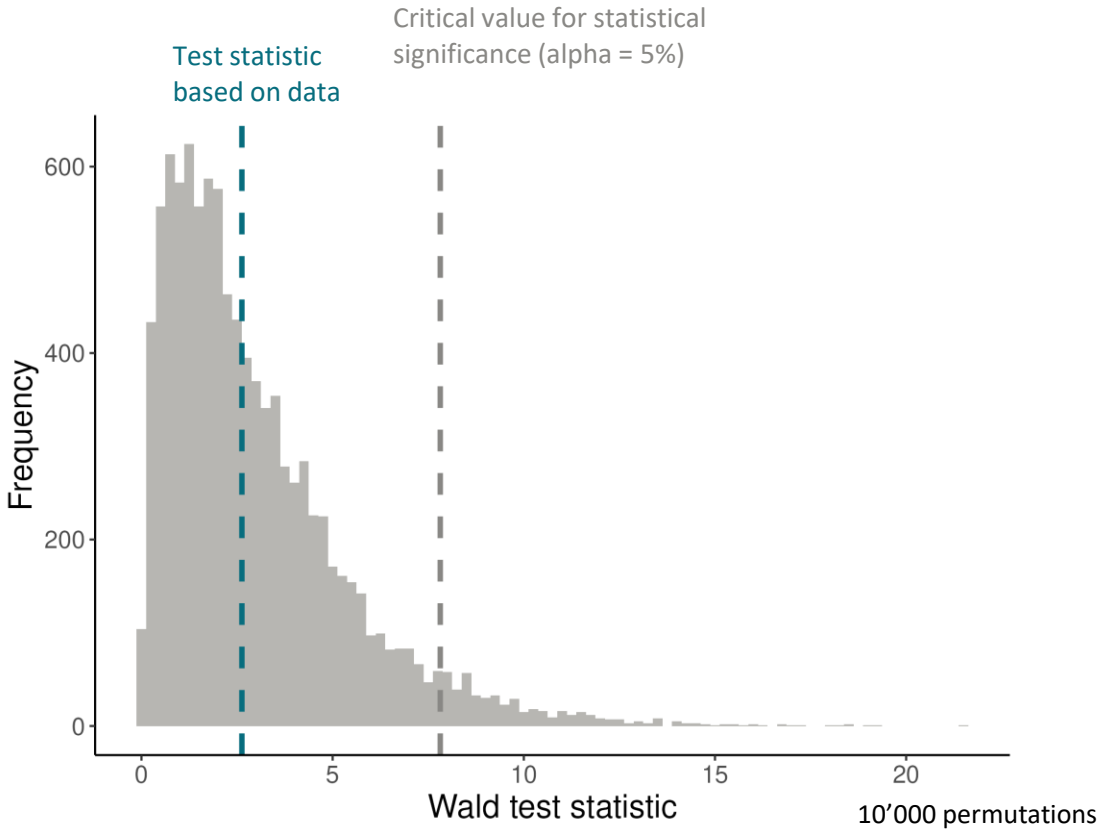
Fixed effects of cohorts:

	OR [95% CI]	p-value
Cohort 2	0.61 [0.30; 1.26]	0.1844
Cohort 4	0.54 [0.23; 1.29]	0.1655
Cohort 5	0.81 [0.35; 1.89]	0.6232

OR: odds ratios; CI: confidence interval;
Cohort 1 is used as reference category.

⇨ No strong statistical evidence of a systematic difference between cohorts.

Visualization of cohort effect:
Compare distribution of expected Wald test statistics under the assumption of no cohort effect with test statistic based on data:



Boehringer Ingelheim

Heterogeneity I – random cohort effect

Random cohort effect

Applied model: $$logit\left(P\left[Y_{lij} \leq k\right]\right) = \alpha_k + S_{C,l}b_{C,l} + S_{V,ij}\beta_{V,j} + z_i b_{S,i}$$

$$b_{C,l} \sim N(0, \sigma_c^2)$$

$$b_{S,i} \sim N(0, \sigma_S^2)$$

Idea:
Calculate the ratio of the variance across cohorts and the total variance (ICC).

Challenges:

- Correlation between $b_{C,l}$ and $b_{S,i}$ needs to be accounted for (see Turner et al., 2000).
- Approximation of the remaining variance in a logistic model (see, for example, Goldstein, 2002).

⇨

Based on our data:

No meaningful estimation of σ_c^2 due to low number of cohorts.

Boehringer Ingelheim

Model overview

Heterogeneity I

AL1

Heterogeneity II

Heterogeneity III

Heterogeneity IV

Generalized mixed model with fixed or random cohort effect

→ assumes equal shift between response categories over time

Generalized mixed model with cohort-specific intercept

→ accounts for different response distributions between cohorts

Generalized mixed model with interaction term of cohort and visit

→ checks for difference in shift over time across cohorts

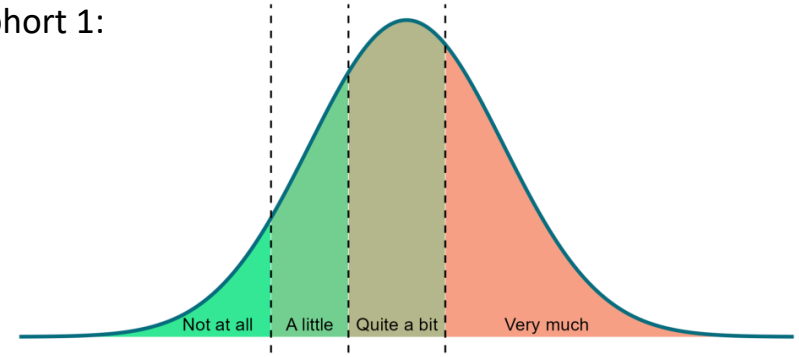
Bayesian meta-analytic approach

→ ranks cohorts with respect to transition odds

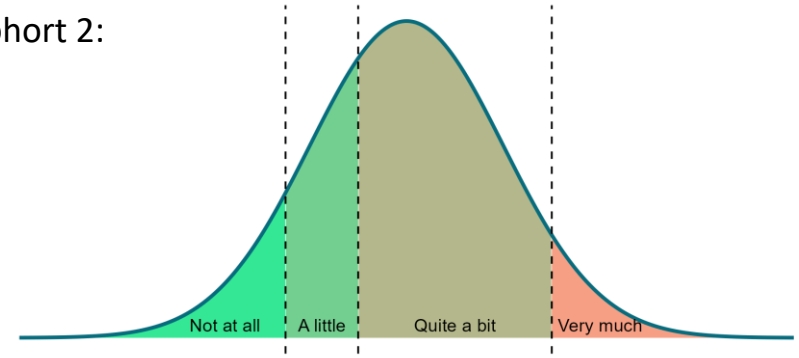
Heterogeneity II – generally different likelihood of categories

Example: Distribution is different between Cohort 1 and Cohrt 2 at baseline

Cohort 1:



Cohort 2:



Cohort- and category-specific intercept

Applied model:

$$logit\left(P\left[Y_{lij} \leq k\right]\right) = \alpha_{kl} + S_{V,ij}\beta_{V,j} + z_i b_{S,i}$$

Compare to model with only category-specific intercept:

$$logit\left(P\left[Y_{lij} \leq k\right]\right) = \alpha_k + S_{V,ij}\beta_{V,j} + z_i b_{S,i}$$

Implementation (in standard software) and interpretation not straightforward (possible?)

➔ Include only baseline data

➔ Likelihood-ratio test between models: p=0.5891

➔ Looping through all visits: p≥0.1110

Boehringer Ingelheim

Model overview

Heterogeneity I

AL1

Heterogeneity II

Heterogeneity III

Heterogeneity IV

Generalized mixed model with fixed or random cohort effect

→ assumes equal shift between response categories over time

Generalized mixed model with cohort-specific intercept

→ accounts for different response distributions between cohorts

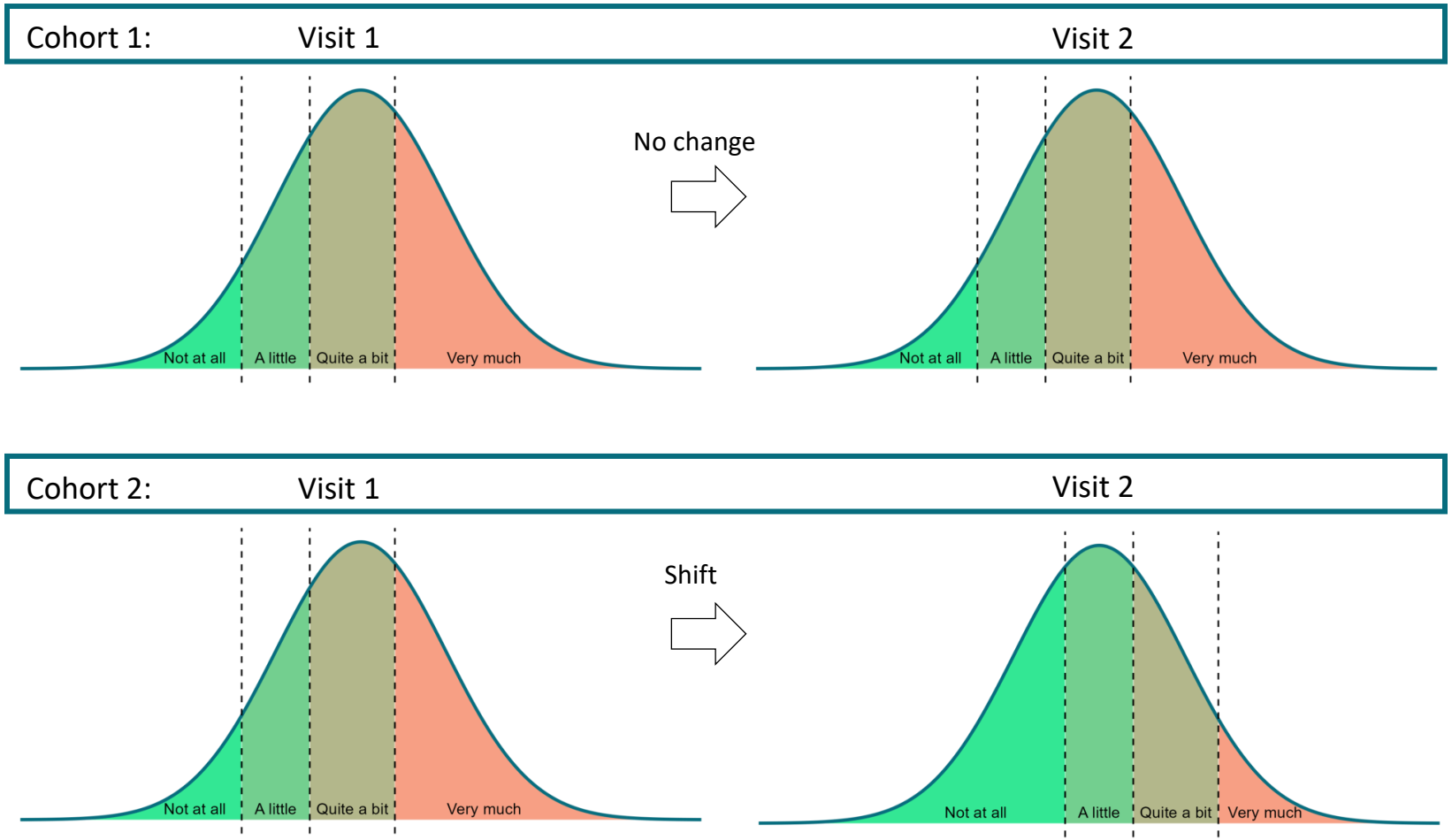
Generalized mixed model with interaction term of cohort and visit

→ checks for difference in shift over time across cohorts

Bayesian meta-analytic approach

→ ranks cohorts with respect to transition odds

Heterogeneity III – differing trend over time



Model components:

- Category-specific intercept
- Fixed effect for visit
- Fixed or random effect for cohort
- Fixed or random effect for interaction of cohort and visit

Boehringer Ingelheim

Heterogeneity III – applied to EORTC IL46 data

All visits included

Fixed effect for cohort:visit interaction

Model: $$logit\left(P\left[Y_{lij} \leq k\right]\right) = \alpha_k + S_{V,ij}\beta_{V,j} + S_{C,i}\beta_{C,l} + S_{V,ij}S_{C,i}\beta_{V:C,lj} + z_i b_i$$

- α_k: Category-specific intercept
- $S_{V,ij}\beta_{V,j}$: Fixed effect for visit
- $S_{C,i}\beta_{C,l}$: Fixed effect for cohort
- $z_i b_i$: Random effect for subject

⇨ Likelihood ratio test (cohort:visit interaction): p=0.7185

⇨ No strong statistical evidence of differences between cohorts

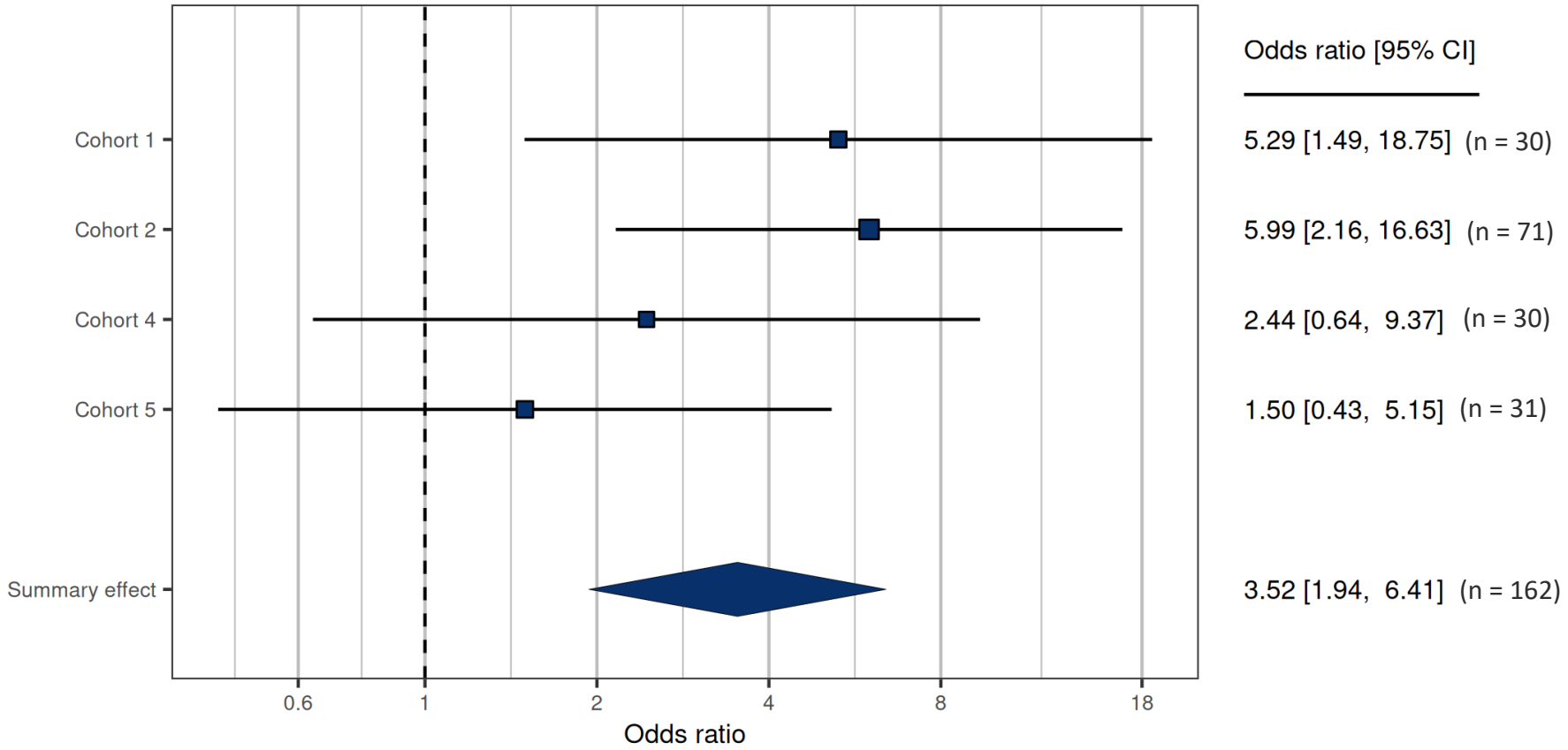
Boehringer Ingelheim

LU2

Heterogeneity III – differing trend – focusing on the end of Cycle 1

Fit the following model per cohort: $logit\left(P\left[Y_{lij} \leq k\right]\right) = \alpha_k + S_{V,ij}\beta_{V,j} + z_i b_i$

➔ Compare the estimated odds ratio for visit at C02, D01:



Odds ratios <1 indicate a higher probability for better outcomes, i.e., an odds ratio <1 is favourable.

Model overview

Heterogeneity I

AL1

Generalized mixed model with fixed or random cohort effect

→ assumes equal shift between response categories over time

Heterogeneity II

Generalized mixed model with cohort-specific intercept

→ accounts for different response distributions between cohorts

Heterogeneity III

Generalized mixed model with interaction term of cohort and visit

→ checks for difference in shift over time across cohorts

Heterogeneity IV

Bayesian meta-analytic approach

→ ranks cohorts with respect to transition odds

Ranking of cohorts (borrowing ideas from network meta-analysis)

- Implement a Bayesian model (without interaction of cohort and visit).*
- In a Bayesian setting, prior distributions need to be defined. Flat priors may be preferable.

$$logit\left(P\left[Y_{lij} \leq k\right]\right) = \alpha_k + S_{V,ij}\beta_{V,j} + S_{C,i}\beta_{C,l} + z_i b_i$$

$$\alpha_k \sim N(0, 1'000\ ) \text{ (ordered)}$$

$$\beta_{V,1} = 0$$

$$\beta_{V,j} \sim N(0, 1'000),\ j = 2, \ldots, 8$$

$$\beta_{C,1} = 0$$

$$\beta_{C,l} \sim N(0, 1'000),\ l = 2, 4, 5$$

$$b_i \sim N(0, \tau^2)$$

$$\tau^2 \sim dunif\,(0, 1'000) \text{ (or use } invGamma(0.001, 0.001))$$

*For a discussion and the implementation of Bayesian network meta-analysis, see Dias et al., 2011.

Bayesian meta-analytic approach

In (network) meta-analytic models, ranking of arms (here: cohorts) is established (by comparing β parameters during MCMC estimation). If ranking reveals big differences between cohorts, this might imply heterogeneity.

Ranking of cohorts:

Cohort	Probability of being the "best" cohort
1	2.8%
2	30.8%
4	56.4%
5	10.0%

⇨ Cohort 4 comes with the highest probability of being the best cohort.

Odds ratio for C4 vs. C1 (best vs. worst cohort):
estimate [95% credible interval]: 0.53 [0.21, 1.32]*

*Odds ratios <1 indicate a higher probability for better outcomes, i.e., an odds ratio <1 is favourable.

⇨ No strong statistical evidence of differences between cohorts

Boehringer Ingelheim

Summary and discussion

- Several ways to check for heterogeneity between cohorts
- Definition of heterogeneity important as it implies specific assumptions on the structure
- Frequentist and Bayesian approaches available based on
  - generalized linear mixed models (i.e., proportional odds model with mixed effects)
  - meta-analytic approaches
- Heterogeneity assessable based on model comparisons (via likelihood ratios tests), permutation approaches (with corresponding visualizations), and ranking of cohorts (via MCMC approaches)
- In our example: no strong statistical evidence for heterogeneity detected

⇨ **Pooling of data from all cohorts is deemed acceptable.**

Obvious caveat: No evidence for a difference is (generally) not any evidence for no difference!

References

- Dias, S., Welton, NJ., Sutton, AJ., & Ades, AE. (2011). NICE DSU Technical Support Document 2: A Generalised Linear Modelling Framework for Pairwise and Network Meta-Analysis of Randomised Controlled Trials. (Technical Support Document in Evidence Synthesis; No. TSD2). National Institute for Health and Clinical Excellence. http://www.nicedsu.org.uk
- Goldstein, H., Browne, W., & Rasbash, J. (2002). Partitioning Variation in Multilevel Models. *Understanding Statistics*, *1*(4), 223–231.
- Turner, RM., Omar, RZ., Yang, M., Goldstein, H., Thompson, SG. (2000). A multilevel framework for meta-analysis of clinical trials with binary outcomes. *Statistics in Medicine,* 19:3417–3432.
- Tutz, G. (2012). Regression for Categorical Data. Cambridge University Press.
- Whitehead, A., Omar, R.Z., Higgins, J.P.T., Savaluny, E., Turner, R.M. and Thompson, S.G. (2001), Meta-analysis of ordinal outcomes using individual patient data. Statist. Med., 20: 2243-2260.

Software

- JAGS: A Program for Analysis of Bayesian Graphical Models Using Gibbs Sampling, Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003), March 20–22, Vienna, Austria
- R Core Team (2025). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.
- Wickham H, François R, Henry L, Müller K, Vaughan D (2023). dplyr: A Grammar of Data Manipulation. R package version 1.1.4, https://dplyr.tidyverse.org.
- H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.
- Wickham H, Miller E, Smith D (2023). haven: Import and Export 'SPSS', 'Stata' and 'SAS' Files. R package version 2.5.4, https://haven.tidyverse.org.
- Jason Cory Brunson and Quentin D. Read (2023). ggalluvial: Alluvial Plots in 'ggplot2'. R package version 0.12.5. http://corybrunson.github.io/ggalluvial/.
- van den Brand T (2024). ggh4x: Hacks for 'ggplot2'. R package version 0.3.0, https://github.com/teunbrand/ggh4x.
- Christensen R (2023). ordinal-Regression Models for Ordinal Data. R package version 2023.12-4.1, https://CRAN.R-project.org/package=ordinal.
- Kossmeier M, Tran U, Voracek M (2020). metaviz: Forest Plots, Funnel Plots, and Visual Funnel Plot Inference for Meta-Analysis. R package version 0.3.1, https://github.com/Mkossmeier/metaviz.
- Su Y, Yajima M (2024). R2jags: Using R to Run 'JAGS'. R package version 0.8-9.