

Evaluation of treatment effects in genetic ataxias using SARA score modelling: comparison of multiple trial designs by a large trial simulation framework

Niels Hendrickx¹ Alzahra Hamdan² Andrew C. Hooker² Xiaomei Chen² Rebecca Schüle^{3,4,5}
 ARCA study group Andreas Träschütz^{3,4} Matthis Synofzik^{3,4} Mats O. Karlsson² France Mentré¹



Emmanuelle Comets^{1,6} the EVIDENCE-RND consortium
¹Université Paris Cité and Université Sorbonne Paris Nord, Inserm, IAME, F-75018 Paris, France ²Pharmacometrics Research Group, Department of Pharmacy, Uppsala University, Uppsala, Sweden ³Division Translational Genomics of Neurodegenerative Diseases, Hertie Institute for Clinical Brain Research (HIH), University of Tübingen, Tübingen, Germany; ⁴German Center for Neurodegenerative Diseases (DZNE), Tübingen, Germany. ⁵Division of Neurodegenerative Diseases, Department of Neurology, Heidelberg University Hospital, Germany ⁶Univ Rennes, Inserm, EHESP, Irset - UMR_S 1085, 35000, Rennes, France.

CONTEXT

- Genetic cerebellar ataxias are progressive rare neurological diseases, often with multi-systemic damage to other neurological systems, causing debilitating impairment of gait, balance, speech, and fine motor skills
- Patient's disease severity is evaluated with the Scale for the Assessment and Rating of Ataxia (SARA) score, a composite clinical score (0-40), which can be modelled using total score [1], or through Item Response Theory (IRT) models [2]
- There are currently no disease modifying drugs for most ARCA, and promising treatment trials with robust designs are needed [3]
- This work aimed to study the influence of the inclusion criterion for two modelling strategies (IRT-based versus total score-based) and the influence of design on the power and type 1 error of simulated clinical treatment trials for a hypothetical disease modifying effect.

INFLUENCE OF SIMULATION MODEL AND INCLUSION CRITERION

Modelling SARA score

- Modelling of the SARA score with an IRT model [2] on the ARCA registry [4]
 - Item Characteristic Curve parameters fitted on the ARCA population (990 patients, 1932 measurements), then fixed.
 - linear model (with time since onset as regressor) fitted on the latent variable scale, on the Autosomal-Recessive Spastic Ataxia Charlevoix Saguenay (ARSACS) (173 patients, 349 measurements) from the ARCA registry [4]
 - normal random effects on base and slope, with a correlation, estimation with NONMEM 7.5.1 [5] LAPLACE
- Modelling the total SARA score with a 4 parameter logistic model versus the time since onset (TSO), on the patients with ARSACS from the ARCA registry
 - lognormal random effects on all parameters except γ , estimation with NONMEM 7.5.1 [5] SAEM

$$f(TSO) = \delta + \frac{\gamma}{1 + e^{\beta - \alpha \times TSO}}$$

Parameter	Value	RSE (%)
δ	6.16	17
γ	28.75	7
β	3.94	13
α (yr^{-1})	0.11	13
ω_δ	0.31	22
ω_γ	0	-
ω_β	0.20	20
ω_α	0.09	79
σ	2.02	5

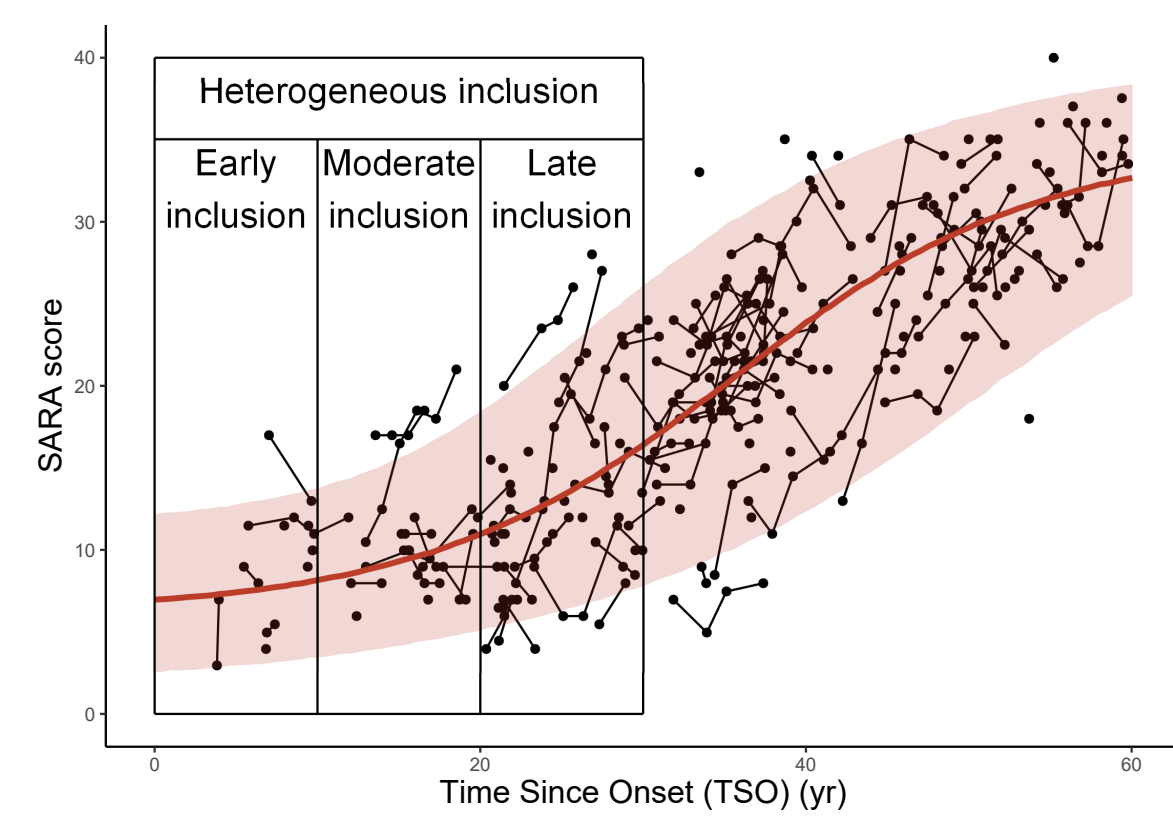


Table 1. Parameter estimates of the logistic model

Figure 1. Median prediction and 90% CI over 500 simulated random effects for the logistic model

Simulation study

- Simulations with IRT model: drug effect on the slope parameter
- Simulations with logistic model on total score: drug effect on α , with γ fixed to its population value in ARSACS: 28.75
- Settings**
 - 500 replicates of 100 patients, parallel design, 1:1 allocation ratio
 - drug effects of 0% and 50%
 - 5 year trials with one visit every 6 months
- Inclusion criteria**
 - 4 cases studies based on TSO at inclusion (early, moderate, late and heterogeneous) (Fig.1)
- Estimation**
 - IRT model simulations: analysed with the IRT model, a linear model and the logistic model
 - logistic model simulations: analysed with the logistic model and a linear model versus TSO
 - parameter estimation: NONMEM 7.5.1 [5], with the Laplace method for the IRT model, and SAEM for the logistic and the linear model
- Evaluation:** tests using the Likelihood Ratio Test
 - empirical type 1 error: fraction of replicates simulated with a drug effect of 0% where the LRT is significant
 - power corrected for the type 1 error (significance threshold chosen such that the type 1 error is 5% for that threshold)

Results

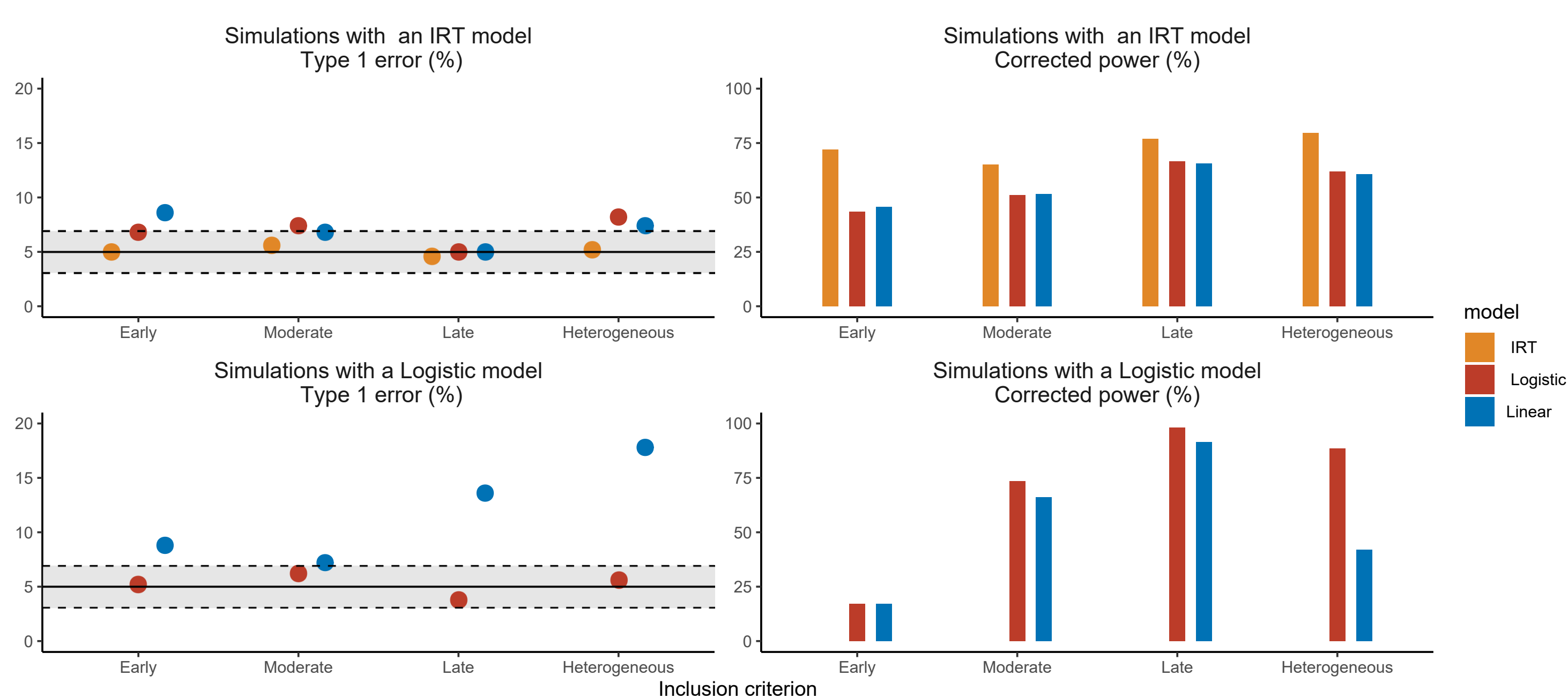


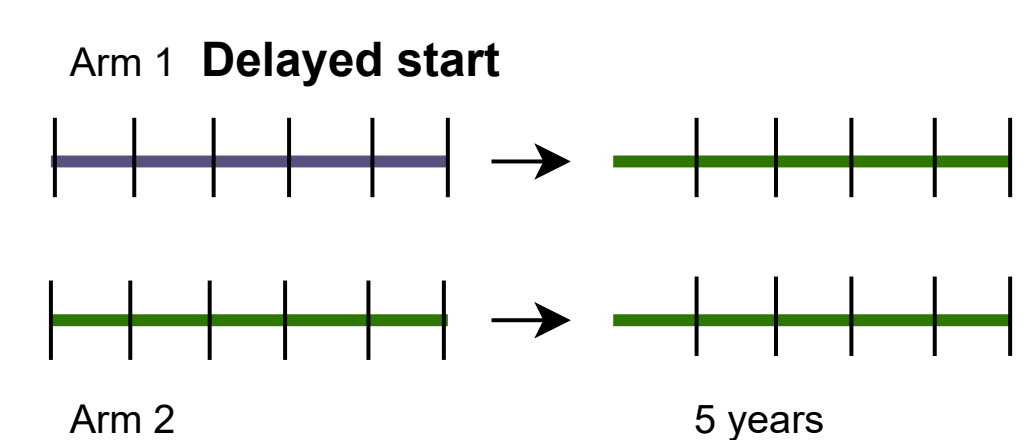
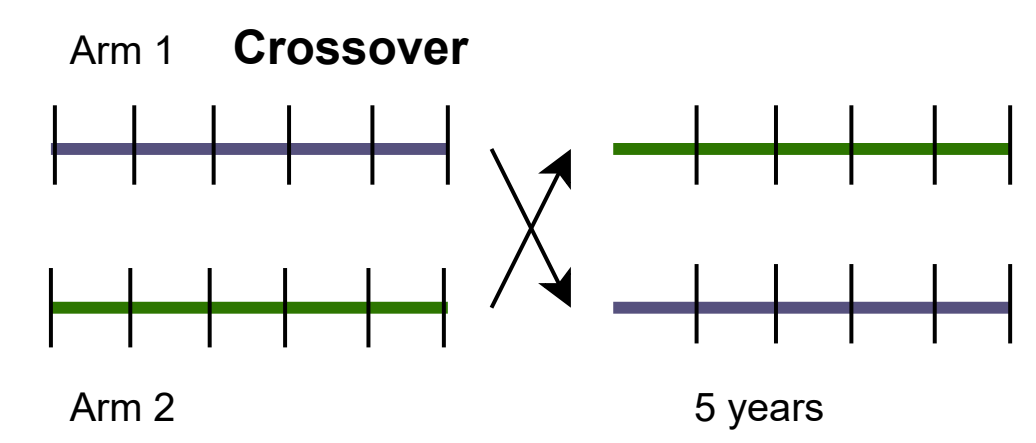
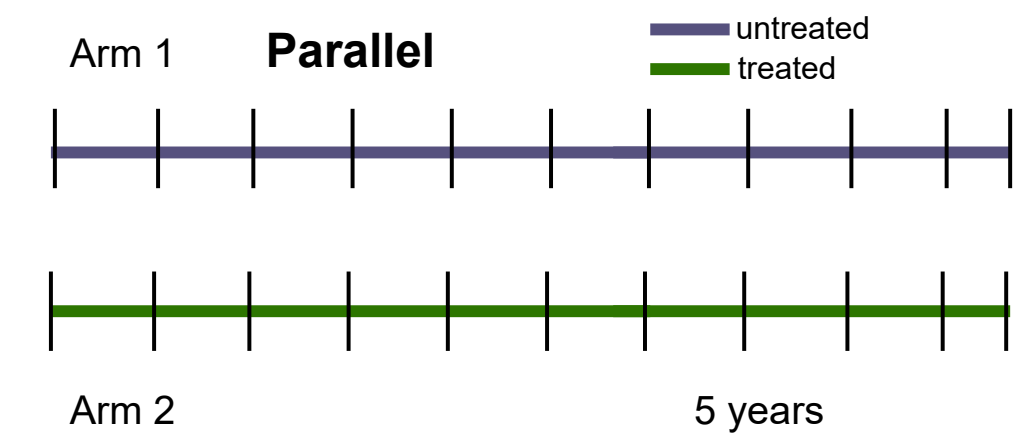
Figure 2. Simulations under an IRT model/a logistic model. Type 1 error and power corrected for the type 1 error for 4 inclusion criteria

- Simulations with IRT model:** type 1 error controlled for the IRT model, inflated for linear and logistic models in some scenarios. Power highest for IRT model in late and heterogeneous inclusion
- Simulations with logistic model:** type 1 error controlled for the logistic but not for the linear model. Lower power for linear model versus logistic model in the heterogeneous inclusion scenario due to non-linear changes in SARA score over a 30 year period

INFLUENCE OF THE DESIGN

Designs

- Inclusion criterion: heterogeneous
- 100 patients
- Duration: 5 years, visit every 6 months
- Parallel design
- Cross-over design
 - patients switch treatment group mid-trial (at 2.5 years)
- Delayed start design
 - patients in arm 1 are treated at the start of the study
 - patients in arm 2 are treated after 2.5 years in the study



Settings

- Simulation**
 - simulation model: logistic model on total score
 - 500 replicates
 - drug effect of 0% and 50% (1:1 allocation ratio)
- Estimation**
 - estimation models: as previous simulation
 - estimation methods: NONMEM 7.5.1 [5] with the FOCEI method for both models
 - evaluation: type I error and corrected power

Results

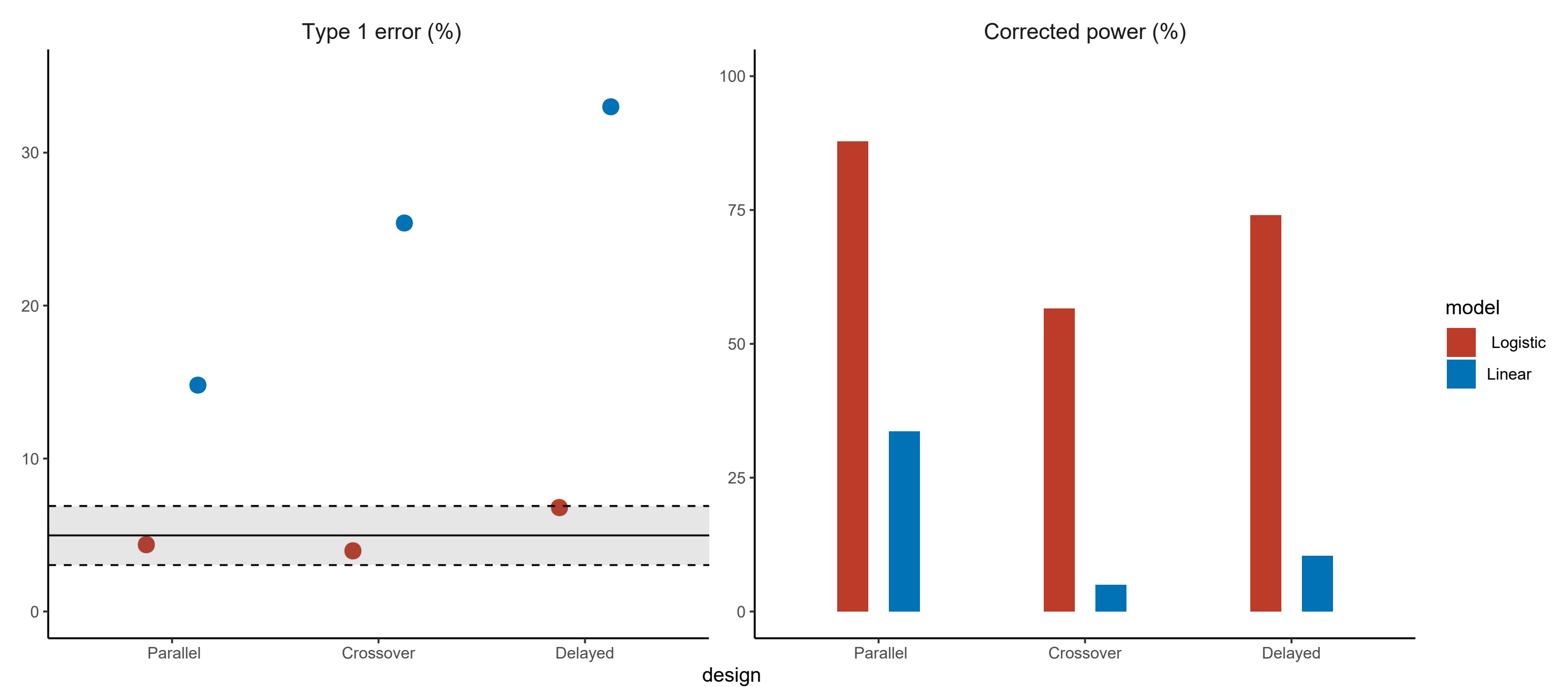


Figure 3. Type 1 error and corrected power for the three designs, simulated with the logistic model with heterogeneous design

- Logistic model:** controlled type 1 error, with the highest power for the parallel design, followed by the delayed start and crossover designs
 - slow disease progression, with high residual error: need to observe progression for a long period to observe an effect → can explain why crossover design has less power than delayed start and parallel designs
- Linear model,** inflated type 1 error in every scenario (from 15 to 33%), resulting in very low corrected power.

CONCLUSION

- Inclusion criterion was found to highly influence power for parallel designs, with the late and heterogeneous inclusion the most powerful
- The choice of simulation model (IRT versus logistic) also had an impact: simulating under the IRT model preserves the discreteness of the data, resulting in residual error lower for lower SARA scores. Simulating under the logistic model treated the data as continuous, resulting in a different error structure, influencing the distribution of power across inclusion criteria
 - in the simulations with the IRT model: power of the IRT model homogeneous across inclusion criteria
 - in the simulations with the logistic model: power of the logistic model much higher with late and heterogeneous inclusion (fast progression) and very low power for early inclusion
- Among the three tested designs, parallel had the highest power
- The linear model performed poorly in simulations with heterogeneous inclusion criterion
 - perspective: use a linear model accounting for time since inclusion in the study to correct for model misspecification

ACKNOWLEDGEMENTS

This work was supported by the European Joint Programme on Rare Diseases (EJPRD) (Grant Agreement n°825575) WP20 Innovation Statistics project "EVIDENCE-RND" (to F.M. R.S., and M.S), the EJPRD PROSPAX consortium, further supported by the Clinician Scientist program "PRECISE.net" (to A.T, M.S. and R.S.). We also thank Lionel de la Tribouille for the use of the CATIBioMed calculus facility.

REFERENCES

- Hendrickx et al. *AAPS J.* 2024.
- Hamdan et al. *CPT Pharmacometrics Syst Pharmacol.* 2024.
- Salem et al. *P CNS Neurol Disord Drug Targets.* 2023.
- Träschütz et al. *Front Neurol.* 2021.
- Beal et al. *Ellicot City: Icon Development Solutions* 2009.