

The Decision-Equivalence Metric (DEM) Method Based Evaluation of Biomarker Surrogacy in MASH

Igor Vasyutin¹, Sabine Stübler¹, Angèle Fleury¹, Corinna Schoelch¹, Anaïs Glatard², Jan-Georg Wojtyniak¹

1. Boehringer Ingelheim Pharma GmbH & Co KG – Ingelheim, Binger Straße 173 55216 Ingelheim am Rhein; 2. Pharmetheus AB Kungsängstull 4 753 19 Uppsala, Sweden

Objectives

- Evaluate the surrogacy of blood biomarkers propeptide of type III collagen (PRO-C3) and enhanced liver fibrosis (ELF) score for histological endpoints in MASH F1/F3 using the Prentice criteria;
- Evaluate the surrogacy of PRO-C3 and ELF for histological endpoints in MASH using the decision-equivalence metric (DEM) approach;
- Compare surrogacy evaluation results obtained with the Prentice criteria and DEM approach;
- Assess DEM framework as a tool for surrogate endpoint evaluation.



Scan the QR code to view the e-poster

Introduction



Early and reliable go/no-go decisions are essential for accelerating drug development in metabolic dysfunction–associated steatohepatitis (MASH). Surrogate biomarkers such as the propeptide of type III collagen (PRO-C3) and the enhanced liver fibrosis (ELF) score have the potential to support mechanistic and quantitative decision-making¹; however, their ability to consistently reflect treatment effects with respect to histological endpoints remains insufficiently established.

Pharmacokinetic/pharmacodynamic (PK/PD) and exposure–response (ER) models enable quantitative characterization and prediction of biomarker dynamics, providing a framework for simulation-based evaluation of biomarker behavior under clinical trial conditions at both individual and population levels.

The Decision Equivalence Metric (DEM) is a recently proposed, simulation-based framework for assessing biomarker surrogacy by quantifying concordance between biomarker-based and endpoint-based decision making within pharmacometric models².

In this study, we applied the DEM framework to evaluate PRO-C3 and ELF as surrogate biomarkers for histological improvement in MASH and to investigate methodological limitations of the DEM approach.

Prentice criteria



Both PRO-C3 and ELF satisfied Prentice criteria 1 and 2, demonstrating significant treatment effects and associations with IHF, but did not meet criterion 3, as adjustment for the surrogate did not fully attenuate the treatment effect on IHF (Table 1). This finding is consistent with the well-recognized difficulty of achieving full compliance with all Prentice criteria in practice.

The proportion of treatment effect explained (Freedman PE) was 23.75% for PRO-C3 and 13.11% for ELF, indicating partial mediation of the treatment effect on IHF.

Table 1. Prentice criteria results for PRO-C3 and ELF.

#	Criterion	PRO-C3		ELF	
		Estimate	p-value	Estimate	p-value
0	The survodutide has a significant effect on the IHF.	0.016	2E-08	0.016	2E-08
1	The survodutide has a significant effect on the PRO-C3/ELF.	-0.178	1E-09	-0.037	7E-07
2	The PRO-C3/ELF is predictive of the IHF.	-0.039	5E-07	-0.096	4E-05
3	The PRO-C3/ELF fully captures (mediates) the survodutide effect on the IHF.	0.012	4E-05	0.014	2E-06

Methods



PK/PD models for PRO-C3 and ELF and an exposure–response (ER) model for improvement of histological findings (IHF) were developed using data from the survodutide Phase II trial 1404-0043 in patients with MASH and fibrosis stages F1–F3. IHF was defined as a ≥2-point decrease in NAFLD activity score (NAS), including ≥1-point improvement in lobular inflammation or ballooning, with no worsening of fibrosis.

Surrogacy of PRO-C3 and ELF for IHF was assessed using the Prentice criteria and DEM framework. The Prentice criteria define a valid surrogate endpoint as one that (i) is affected by treatment, (ii) is associated with the endpoint of interest, and (iii) accounts for the treatment effect on the endpoint of interest when included in the model³. IHF, PRO-C3, and ELF were analyzed using logistic or linear regression models, depending on nature. Treatment effect in regression models was represented by predicted average steady-state concentrations of survodutide; placebo subjects were assigned a concentration of zero.

Following DEM framework, PK/PD and ER models were integrated to simulate 1,500 parallel virtual trials incorporating between-subject variability and parameter uncertainty. Independent hypothesis testing was performed for PRO-C3/ELF and IHF. DEM was defined as the proportion of simulated trials with concordant decisions based on PRO-C3/ELF and IHF and was evaluated across multiple sample sizes² (Figure 1).

Modeling was performed in NONMEM® v7.5; simulations and analyses were conducted in R v4.4.1.

The DEM



DEM analyses demonstrated a monotonic increase in decision agreement with increasing sample size. At smaller sample sizes (8–20 participants per arm), DEM values were reduced, reflecting greater uncertainty and variability in both biomarker and histological endpoints.

From approximately 30 participants per arm onward, DEM values exceeded 90% for both biomarkers, indicating strong concordance between surrogate-based and histology-based trial conclusions.

Notably, the power trajectories for PRO-C3, ELF, and histological improvement closely mirrored the observed DEM patterns, suggesting that increasing statistical power translated directly into improved alignment between surrogate and clinical decision-making.

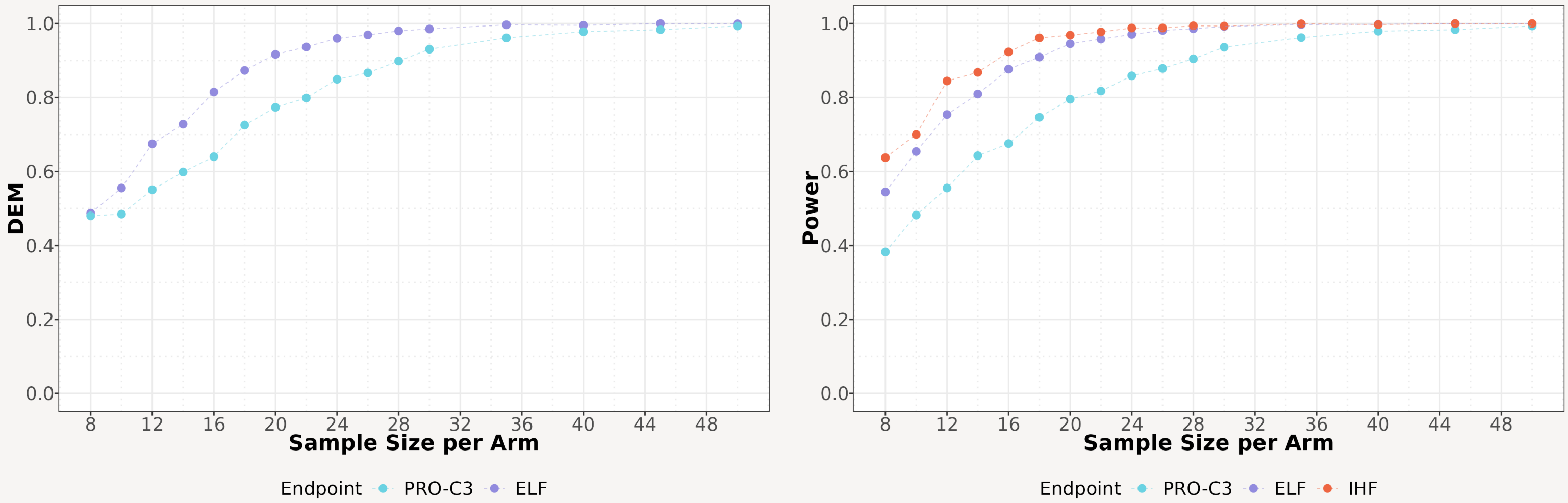


Figure 2. (A) Decision agreement of PRO-C3/ELF-based endpoints with IHF-based endpoint for different sample size; (B) trial power trajectories for PRO-C3/ELF/IHF-based endpoints. All endpoints were simulated for week 48 of survodutide/placebo treatment.

Step 1: Model development

$$S = f(Z)$$
$$EI = g(Z)$$

S – surrogate
EI – endpoint of interest
Z – treatment

Step 2: Clinical trial simulations

Simulate 1500 trials for specific trial size N
Re-sampling performed from distributions (uncertainty and variability)

Step 3: DEM score calculation

For each simulated trial, perform hypothesis testing using surrogate and EI

Calculate DEM as

$$\frac{\# \text{ trials where decisions match}}{\text{total \# of simulated trials}}$$

Figure 1. DEM framework.

Conclusions

Neither PRO-C3 nor ELF fulfilled all three Prentice criteria, consistent with the stringency of classical surrogacy definitions. Using DEM, both biomarkers demonstrated high concordance with histology-based decisions, exceeding 90% agreement at sample sizes of ~30 participants per arm.

DEM dependency on endpoint power is an important limitation, as high agreement may partly reflect study precision rather than true surrogate performance. Joint modeling and power-adjusted DEM approaches may improve assessment of surrogate reliability and support model-informed drug development in MASH.

Abbreviations

DEM, Decision-Equivalence Metric; ELF, enhanced liver fibrosis score; ER, exposure–response; IHF, improvement of histological findings; MASH, metabolic dysfunction–associated steatohepatitis; NAFLD, Non-Alcoholic Fatty Liver Disease; NAS, NAFLD activity score; PK/PD, Pharmacokinetic/pharmacodynamic; PRO-C3, propeptide of type III collagen.

References

- 1 – Sanyal A, Shankar S, Yates K, et al. The Nimble Stage 1 Study Validates Diagnostic Circulating Biomarkers for Nonalcoholic Steatohepatitis. Res Sq [Preprint]. 2023 Jan 19;rs.3.rs-2492725. doi: 10.21203/rs.3.rs-2492725/v1. Update in: Nat Med. 2023 Oct;29(10):2656-2664. doi: 10.1038/s41591-023-02539-6
- 2 – <https://sxpisg.github.io/events/pastevents/posts/webinar-12Sep2023.html>
- 3 – Prentice RL. Surrogate endpoints in clinical trials: definition and operational criteria. Stat Med. 1989 Apr;8(4):431-40. doi: 10.1002/sim.4780080407

Acknowledgments

This study is supported and funded by Boehringer Ingelheim. The authors met the criteria for authorship as recommended by the International Committee of Medical Journal Editors (ICMJE) as supported by Good Publication Practice (GPP) guidelines. The authors did not receive payment related to the development of this poster. Boehringer Ingelheim was given the opportunity to review the poster for medical and scientific accuracy, as well as intellectual property considerations.



Boehringer
Ingelheim