Generalizing the distribution and dynamics of health outcomes in clinical trials for novel treatments of rare disease to the real world

Poster No. MSR29

Devin Incerti[1], Darius N. Lakdawalla[1,2], Patricia Dorling[3], Partha Deb[4], Edward Vytlacil[5], Andrew Morris[1], Joseph C Cappelleri[3]

[1]EntityRisk, Inc. Princeton, NJ, USA; [2]University of Southern California, Los Angeles, CA, USA; [3]Pfizer Inc, New York, NY, USA; [4]Hunter College and NBER, New York, NY, USA; [5]Yale University, New Haven, CT, USA

Background

Rare disease studied: Duchenne muscular dystrophy ("DMD")
- Progressive, debilitating neuromuscular disorder
- Prevalence reported as 15.9 cases per 100,000 in the U.S.[1]

Novel therapy: PF-06939926 (developed by Pfizer)
- Designed to potentially slow or arrest DMD progression
- Phase-3 randomized control trial underway

Standard of care: Physiotherapy and glucocorticoids[2]

Health outcomes: Disease progression measured by the North-Star Ambulatory Assessment ("NSAA")[2]
- NSAA is comprised of scores of 0, 1, or 2 on 17 separate physical tasks. Higher scores indicate stronger performance on a given task, while 0 indicates inability to perform the task.

Real-world data source: The Cooperative International Neuromuscular Research Group ("CINRG") has established a large, multicenter natural history study of DMD patients and progression.[3]

Table 1. Summary of CINRG real-world data

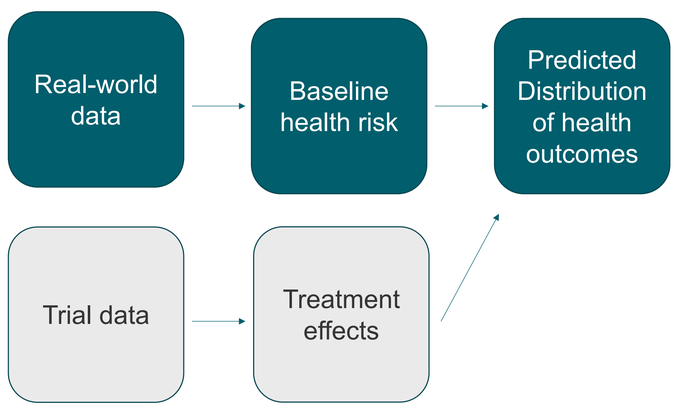
Data type	Longitudinal; inconsistent measurement timing
Subjects	135
N	441
Period	0 – 24 months after first NSAA measurement ("baseline")
Key health characteristics	Age, height, & weight at baseline; NSAA scores at baseline; receiving glucocorticoids

Note: Includes only subjects with more than one NSAA measurement

Objective

To predict the full distribution of health outcomes for patients receiving a novel treatment for rare disease, by leveraging real-world data to model baseline health risk in tandem with trial data to estimate treatment effects

Figure 1. Prediction framework



Methods

Prediction of real-world outcomes for DMD proceeds in **seven** steps, of which we have performed six thus far (See Figure 2).

Step 1: Model specification

- Use a Bayesian cumulative ordered logit model[4] to represent the ordinal nature of the response variable: **decline in NSAA skill count**
  - NSAA skill count is the number of tasks on which a subject scores 1 or 2.
- Model specifications reflect hypotheses about the functional form of the predictors and clinically significant changes in NSAA measurements.[5]
- Apply priors to the modeling specifications as follows:
  - Intercept: Normal(logit(p_i), 1) where p_i is the cumulative probability of each possible value of the ordinal outcome variable
  - Predictors: Regularized horseshoe with one degree of freedom for the global shrinkage parameter[6]

Table 2. Model specifications

Specification	Endpoint	Predictors	Description
Linear	NSAA skill decline of 0 / 1 / 2 / 3+ at time *t*	Age, height, weight, NSAA skills at index; days since index	All predictors linear
Interactions	See above	See above	Days since index interacted with all other predictors
Splines, 4 degrees freedom	See above	See above	Natural cubic splines (3 internal knots), fit on all predictors
Splines, 3 degrees freedom	See above	See above	Natural cubic splines (2 internal knots), fit on all predictors
Quadratic	See above	See above	Degree 2 polynomial, fit on all predictors

Step 2: Model Fitting

- Pre-process raw data according to model specifications[7]
- Fit Bayesian models to real-world data using Stan

Step 3: Model Checking

- Check for convergence[8]
- Check alignment of estimated predictor effects with clinical literature (See Figure 3)[9]

Figure 2. Modeling workflow

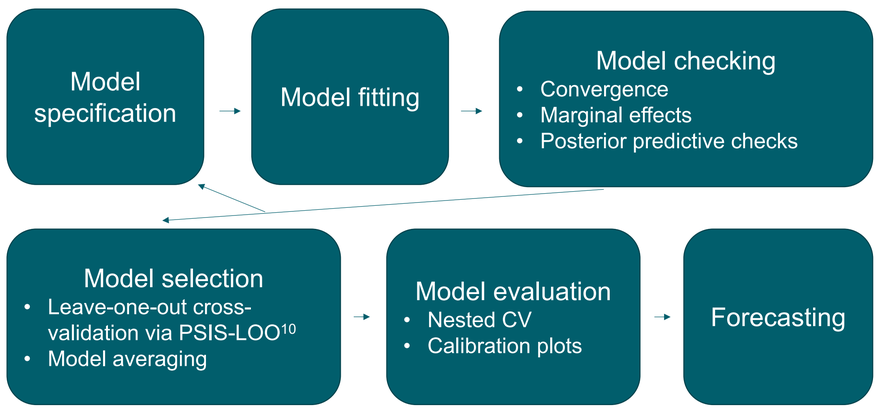
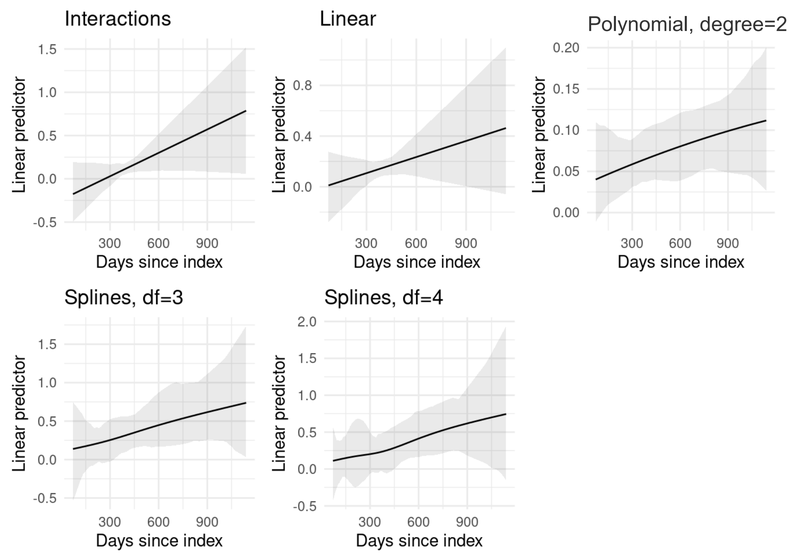


Figure 3. Marginal effects of days since index on probability of NSAA decline

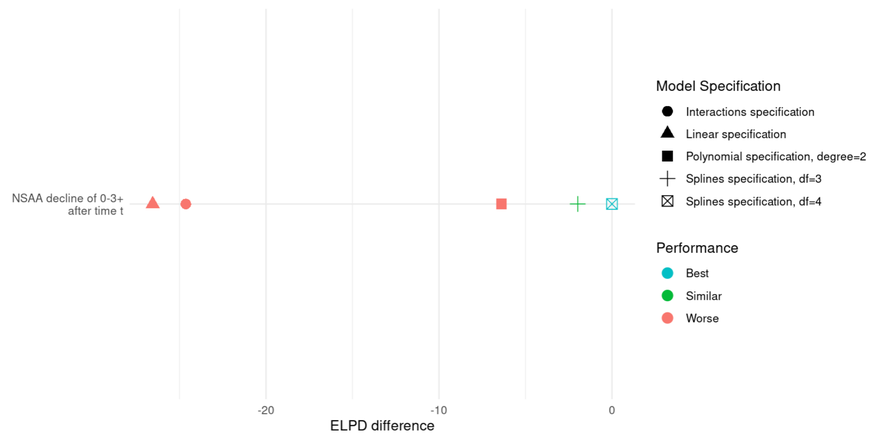


Solid lines reflect estimated linear predictor values for the days since index predictor. The gray areas reflect 95% credible intervals.

Step 4: Model Selection

- Evaluate each individual specification's out-of-sample ("OOS") performance
- Measure OOS performance using estimated log-pointwise density ("ELPD")
- Estimate ELPD via Pareto-smoothed importance sampling leave-one-out ("PSIS-LOO") cross validation[10]
- Combine the best performing models – the two **Splines specifications** – to form a "weighted average" posterior predictive distribution (See Figure 4)
- Weight each model by its relative OOS performance[11]

Figure 4. Relative model performance measured by ELPD via PSIS-LOO

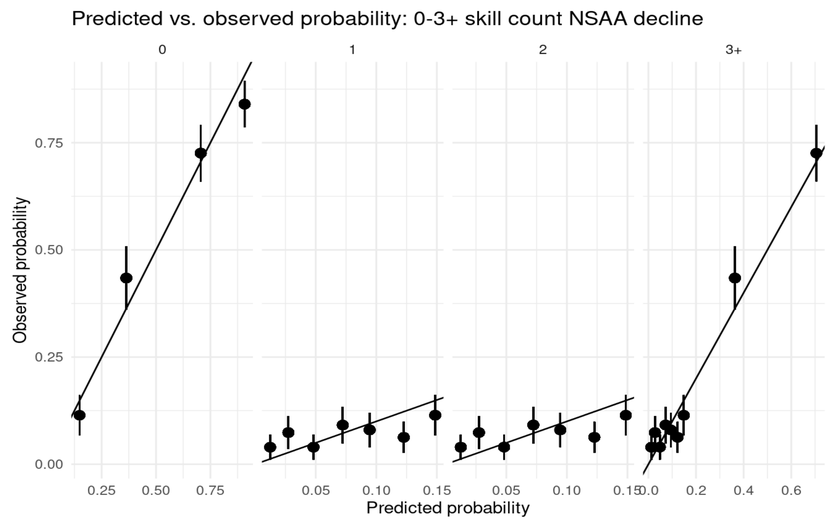


Specifications in the "Best" and "Similar" groups are included in the average posterior predictive distribution

Step 5: Model Evaluation

- Implement a nested cross-validation workflow[12] to estimate the "structural" uncertainty for the **entire process**
- Calibration plots show the accuracy of the posterior predictive distribution for sub-population baseline health & risk profiles (See Figure 5).

Figure 5. Calibration plots from nested CV



Solid lines represent a perfectly calibrated model. Points represent predicted probabilities of NSAA skill decline resulting from the modeling workflow, with error bars showing 95% credible intervals.

Step 6: Forecasting

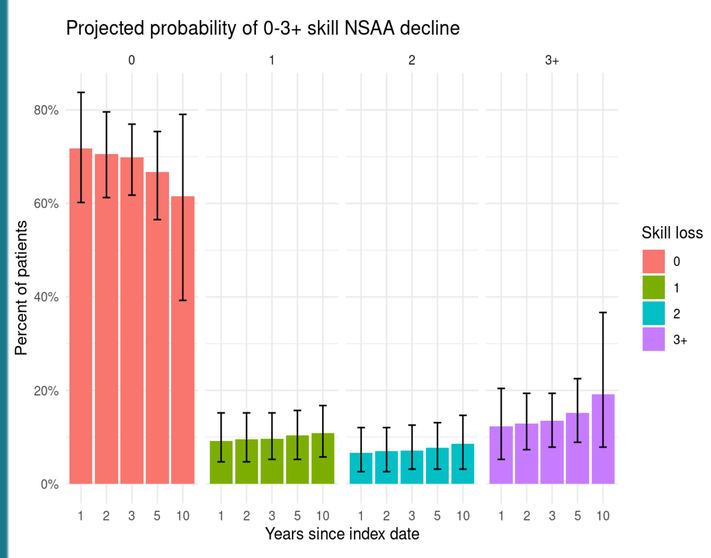
- Simulate the posterior predictive distribution for the target population at timepoints of interest using the "weighted average" approach described in step 4 (See Figure 6)

Step 7: Incorporate Treatment Effects

- Estimate a treatment effect by combining the trial and real-world data
- Apply the effect to the predictive model
- Simulate the distribution of health outcomes for the treated population
- Not yet implemented as the phase 3 RCT is ongoing

Results

Figure 6. Projected distribution of NSAA skill declines



Error bars reflect 95% credible intervals.

Conclusions

The prediction framework detailed here:
- Captures structural uncertainty via model weighting
- Generalizes to the real-world population
- Generates the full distribution of predicted outcomes, rather than just means
- Enables extrapolation of baseline health risk and treatment effects into the future
- Provides evidence to inform outcomes-based contracting and parameterize disease models for cost-effectiveness analysis
- Can be applied to a wide variety of other rare disease estimation problems

References

1. Ryder S, Leadley RM, Armstrong N, et al. The burden, epidemiology, costs and treatment for Duchenne muscular dystrophy: an evidence review. *Orphanet J Rare Dis.* 2017;12:79
2. Birnkrant DJ et al., DMD Care Considerations Working Group. Diagnosis and management of Duchenne muscular dystrophy, part 1: diagnosis, and neuromuscular, rehabilitation, endocrine, and gastrointestinal and nutritional management. *Lancet Neurol.* 2018 Mar; 17(3):251-267
3. See https://cinrgresearch.org/duchenne-natural-history.
4. See Bürkner, P.-C. (2017). brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software*, 80(1), 1–28. https://doi.org/10.18637/jss.v080.i01
5. Here we show results using NSAA skill losses as an endpoint, but we have also executed this predictive workflow using the NSAA total score. NSAA total score is the sum of a subject's scores on the NSAA 17 items.
6. The horseshoe prior is implemented via the brms package in R with default settings. Preliminary analysis suggests the modeling results are not highly sensitive to adjusting the horseshoe parameters. See Piironen & Vehtari (2016). On the Hyperprior Choice for the Global Shrinkage Parameter in the Horseshoe Prior. https://arxiv.org/pdf/1610.05559v1.pdf
7. Includes calculating interaction term values, fitting splines / polynomials, and then scaling each predictor such that they have mean 0 and standard deviation 1.
8. Convergence measured by Rhat. *See* https://mc-stan.org/misc/warnings.html#r-hat
9. For example, we confirm that subjects are more likely to experience NSAA declines over time, and heavier patients less likely than lighter patients. *See, e.g.,* Goemans N, Wong B, Van den Hauwe M, Signorovitch J, Sajeev G, et al. (2020) Prognostic factors for changes in the timed 4-stair climb in patients with Duchenne muscular dystrophy, and implications for measuring drug efficacy: A multi-institutional collaboration. *PLOS ONE* 15(6): e0232870.
10. Out of sample performance is measured by ELPD, estimated via PSIS-LOO. Estimated log-pointwise density can be summarized as a Bayesian version of log-likelihood. Intuitively, the ELPD selects the model in which predictions are probabilistically most like the test data. Pareto-smoothed importance sampling leave-one-out cross validation is a computationally efficient approximation of leave-one-out cross validation. *See* Vehtari, A., Gelman, A. & Gabry, J. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Stat Comput* 27, 1413–1432 (2017); https://avehtari.github.io/modelselection/CV-FAQ.html#11_What_is_the_interpretation_of_ELPD__elpd_loo__elpd_diff.
11. Yuling Yao, Aki Vehtari, Daniel Simpson, Andrew Gelman "Using Stacking to Average Bayesian Predictive Distributions (with Discussion)," Bayesian Analysis, Bayesian Anal. 13(3), 917-1007, (September 2018)
12. We implement nested CV with 5 outer folds. Nested CV is a solution for inefficiencies and high-variance resulting from evaluating models trained on small datasets. *See* Piironen, J., Vehtari, A. Comparison of Bayesian predictive methods for model selection. *Stat Comput* 27, 711–735 (2017).

Disclosures

Study sponsored by Pfizer, Inc.

Abstract No. 118023. Presented at ISPOR Europe 2022, 7 November, 2022

entityrisk know uncertainty