Prioris.ai
When predictions matter

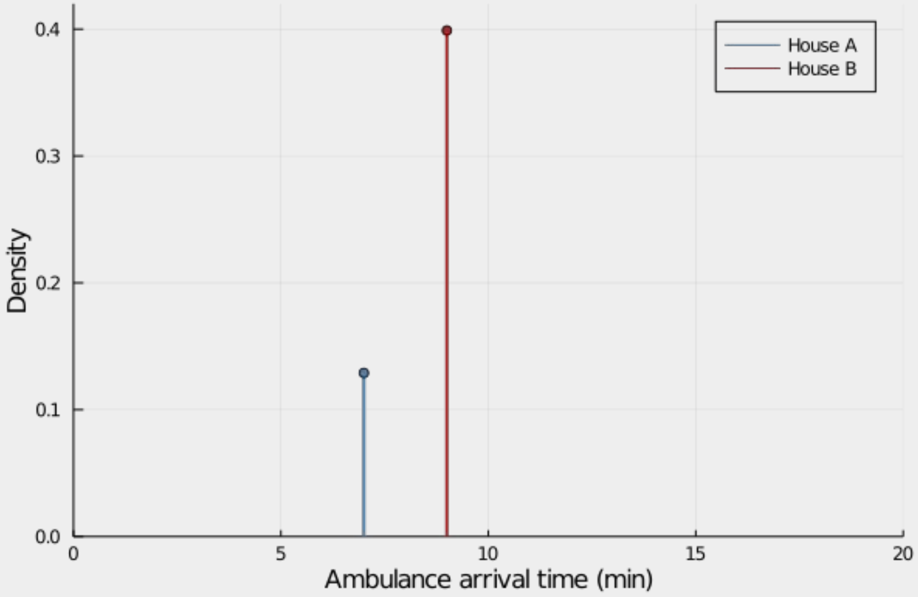
Quantifying and reporting prediction uncertainty with probabilistic models

16 Mar 2021

Stanley E. Lazic, PhD
stan.lazic@prioris.ai
@stanlazic

Does uncertainty matter?

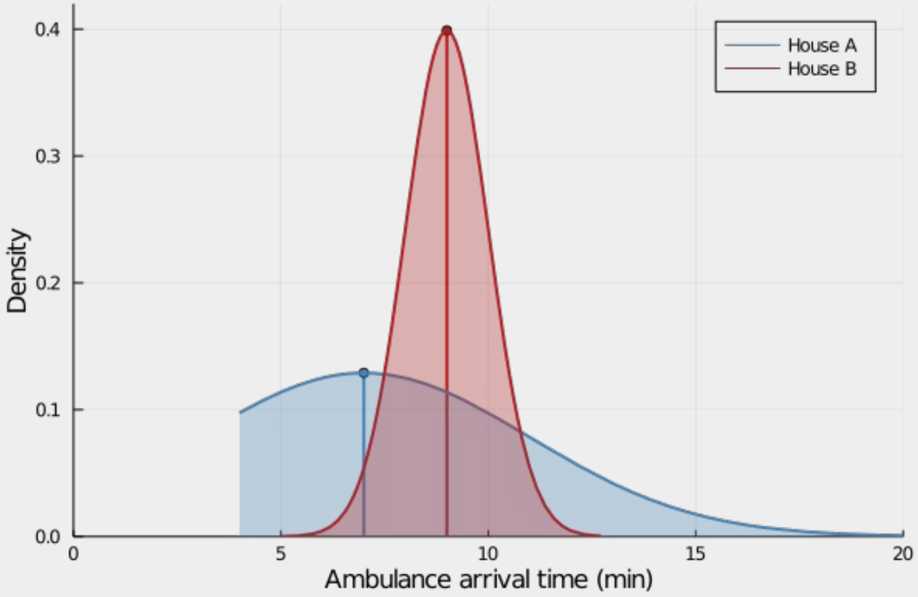
Assume you're moving to a new house but want to be close to a hospital due to a heart condition. Which house do you choose?



Assume you want an ambulance to arrive before 10 min.

Does uncertainty matter?

Assume you're moving to a new house but want to be close to a hospital due to a heart condition. Which house do you choose?



House A:

P(Arrival time > 10 min) = 0.29

House B:

P(Arrival time > 10 min) = 0.16

Uncertainty matters when:

1) The range of plausible values is as important as the best estimate.

2) You need to know that the model doesn’t know.

3) You’re ranking items and want to know if the ranks can be reliability distinguished

Supervised machine learning (predictive modelling)

Classic ML:	Y | X

Supervised machine learning (probabilistic predictive modelling)

Classic ML:	Y | X
Probabilistic ML:	P(Y | X)

Supervised machine learning (probabilistic predictive modelling)

Classic ML:	Y | X, Model
Probabilistic ML:	P(Y | X, Model)

How do you get a distribution for a prediction?

1) Fully Bayesian (MCMC, Variational Inference)

2) Partially Bayesian (MC Dropout, Last-layer Laplace Approximation, Mean-Variance Estimation, ...)

3) Ensembles (Bootstrap, Deep Ensembles)

Sources of uncertainty

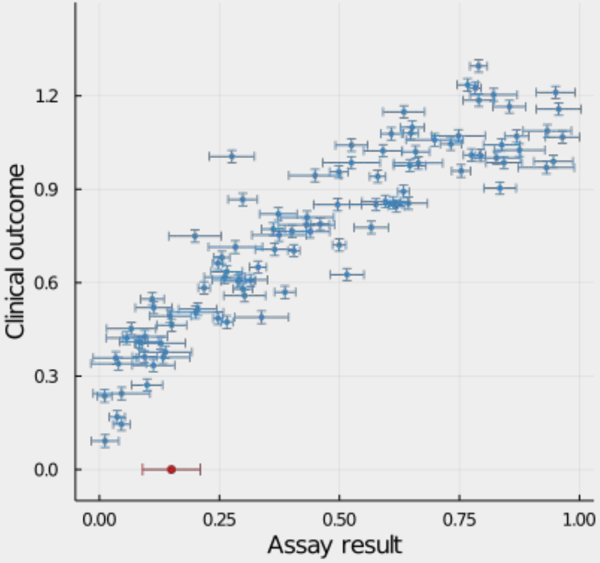
1) The data (X and Y)

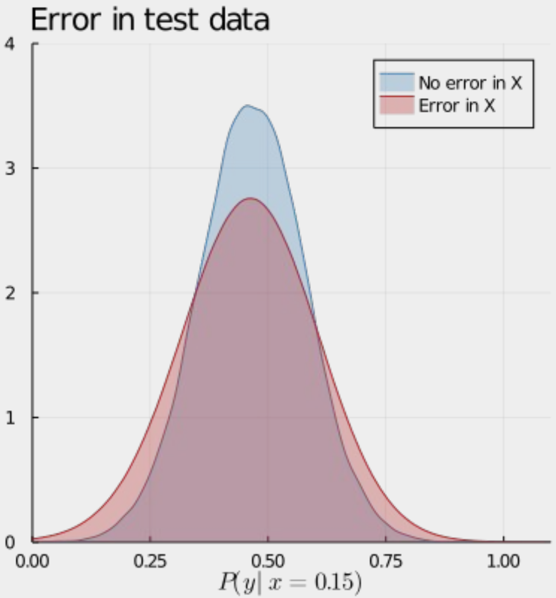
$$\left.\begin{array}{l}\text{2) Distribution function}\\ \text{3) Mean function}\\ \text{4) Variance function}\\ \text{5) Link function(s)}\\ \text{6) Parameters \& hyperparameters}\end{array}\right\}\ \text{The model}$$

For calibrated prediction intervals, key sources of uncertainty need to be incorporated.

Uncertainty in the data (X and Y)

1) Measurement error
2) Misclassification
3) Censoring/truncation
4) Binning
5) Missing data



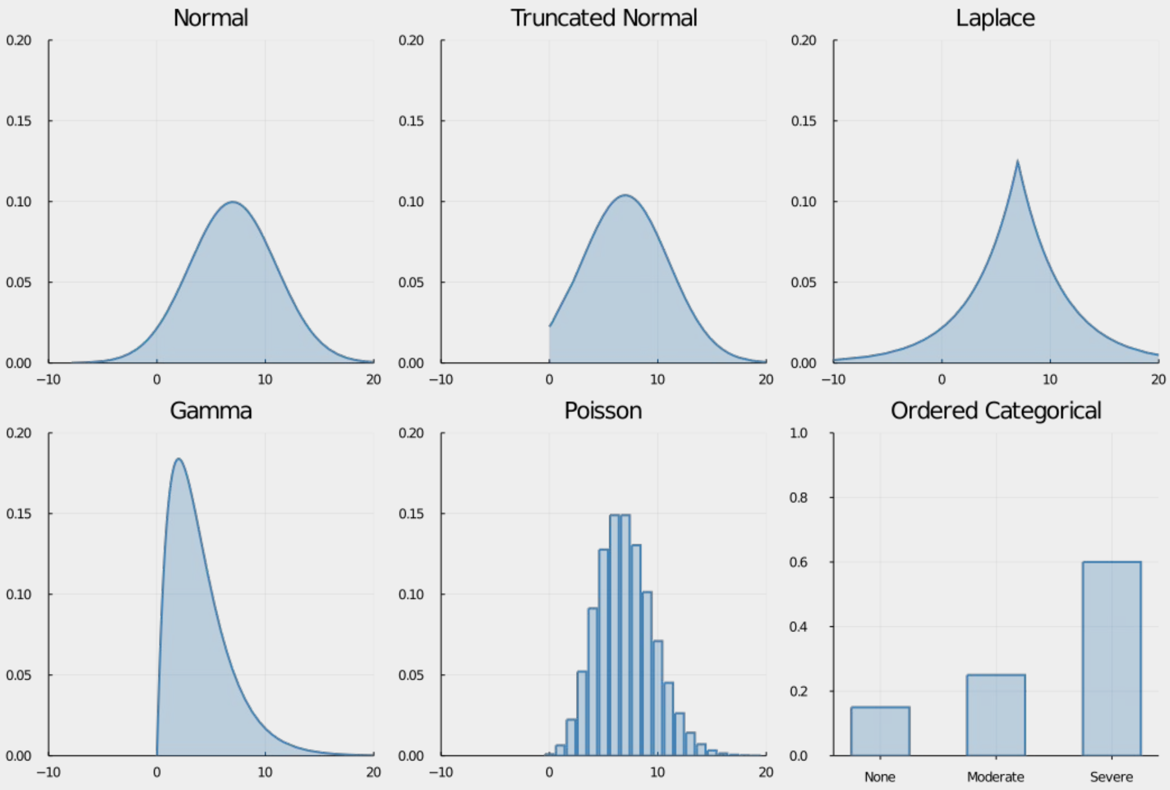


Uncertainty can be in the training data, test data, or both.

Distribution function

The likelihood or data generating distribution represents our uncertainty in Y.

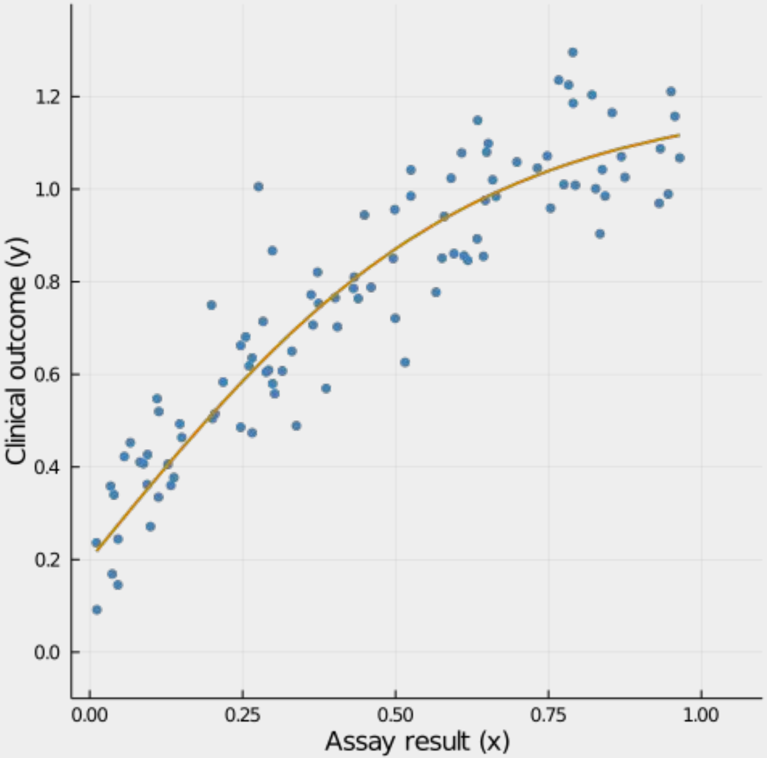
Background knowledge can narrow the options.



Mean function

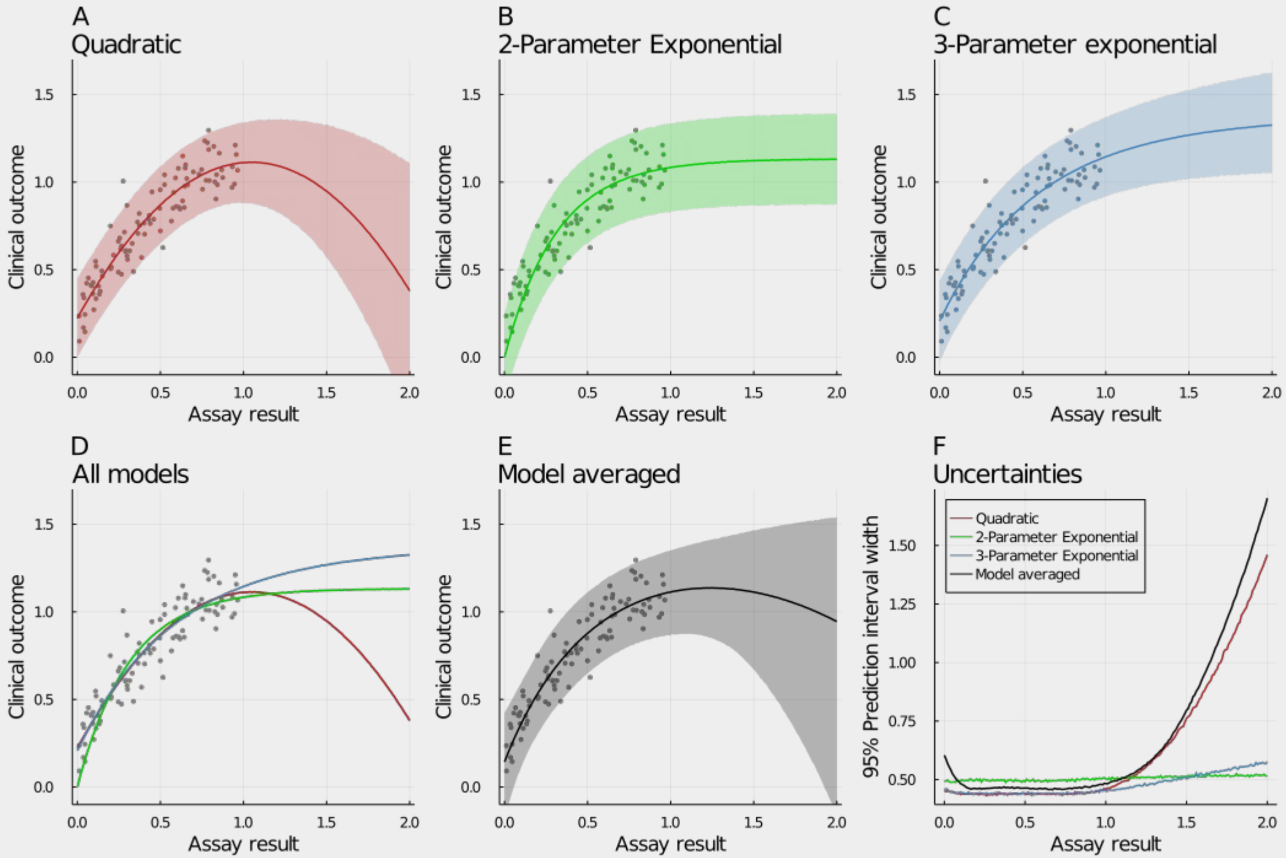
The functional or structural form of the model describing how Y changes as X changes:

- NN architecture
- Tree(s) model
- (Non)linear regression model
- Differential equation



Mean function uncertainty

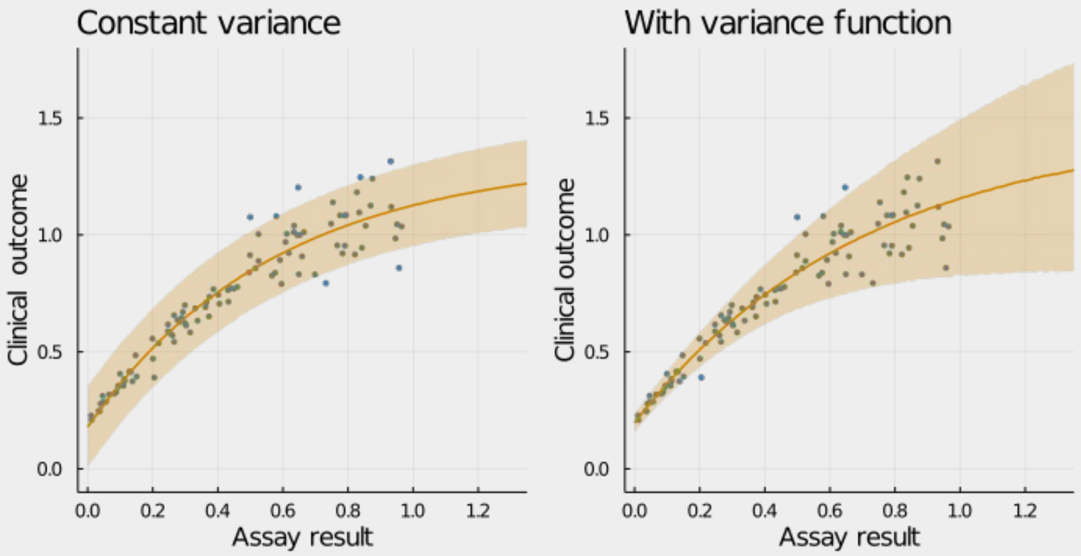
- 3 models fit to the data.
- Model-averaged prediction is high where models make different predictions.



Variance function

Describes how the uncertainty in Y varies with X.

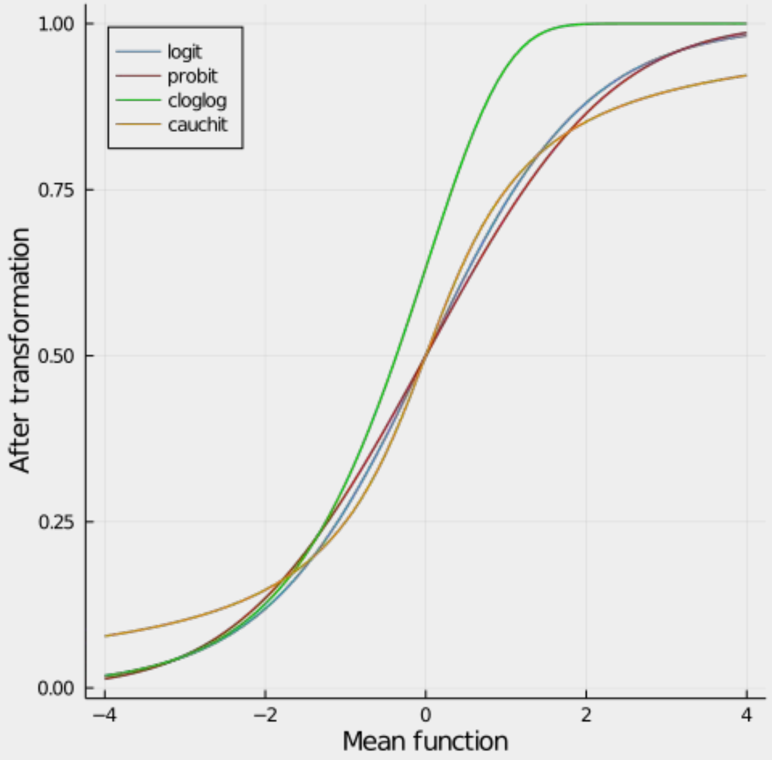
- Constant variance often assumed
- Variance function can be a NN, or
- A simple function of the mean function



Link function(s)

Nonlinear transformation of the mean and/or variance functions to keep the values within an acceptable range.

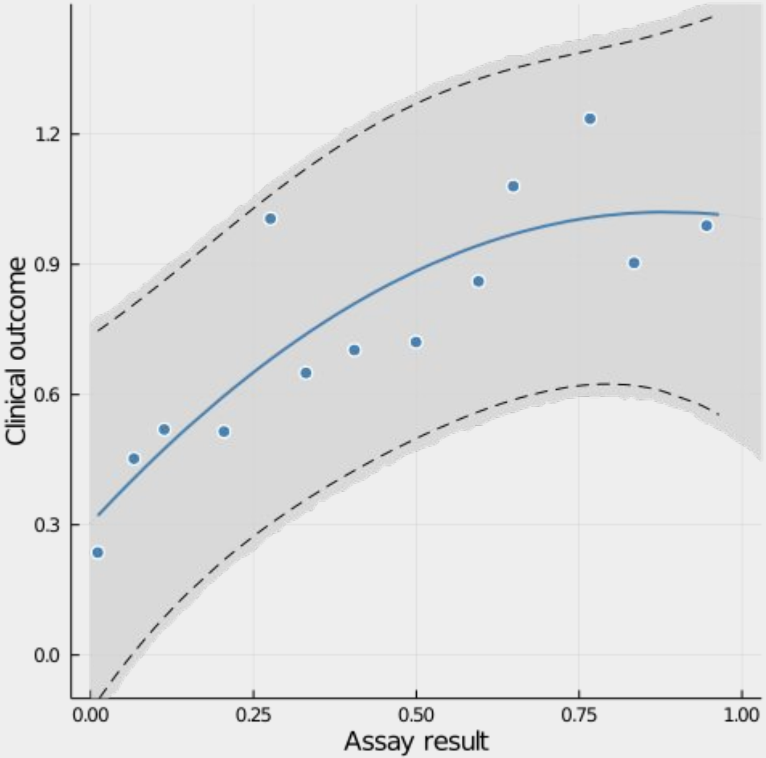
- Similar to activation functions for NNs.



Parameters & Hyperparameters

Unknown coefficients or weights for the mean and variance functions.

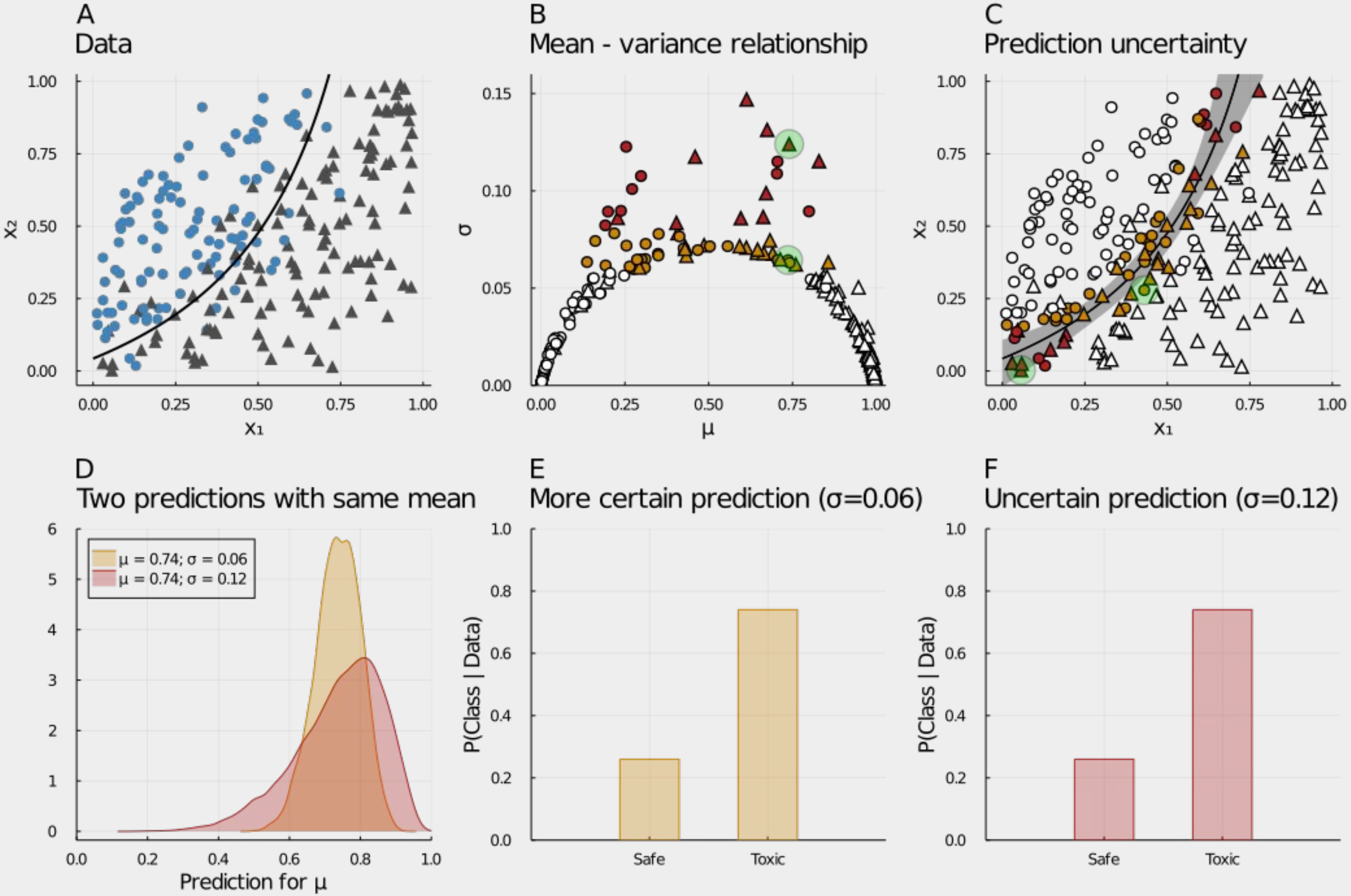
- Estimated from the data (weights + biases)
- Fixed (weight decay)



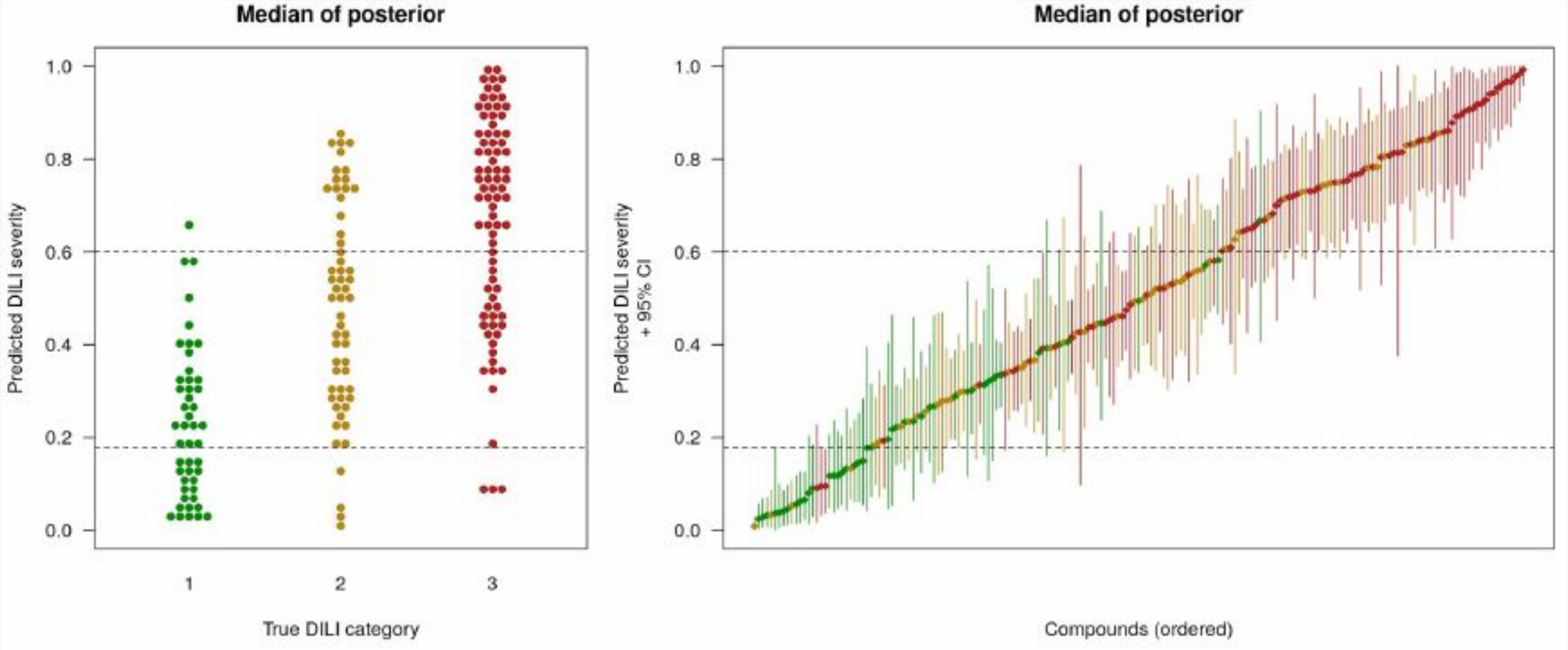
Dashed lines = 95% PI without parameter uncertainty
Grey shaded area = 95% PI with parameter uncertainty

Uncertainty for binary outcomes

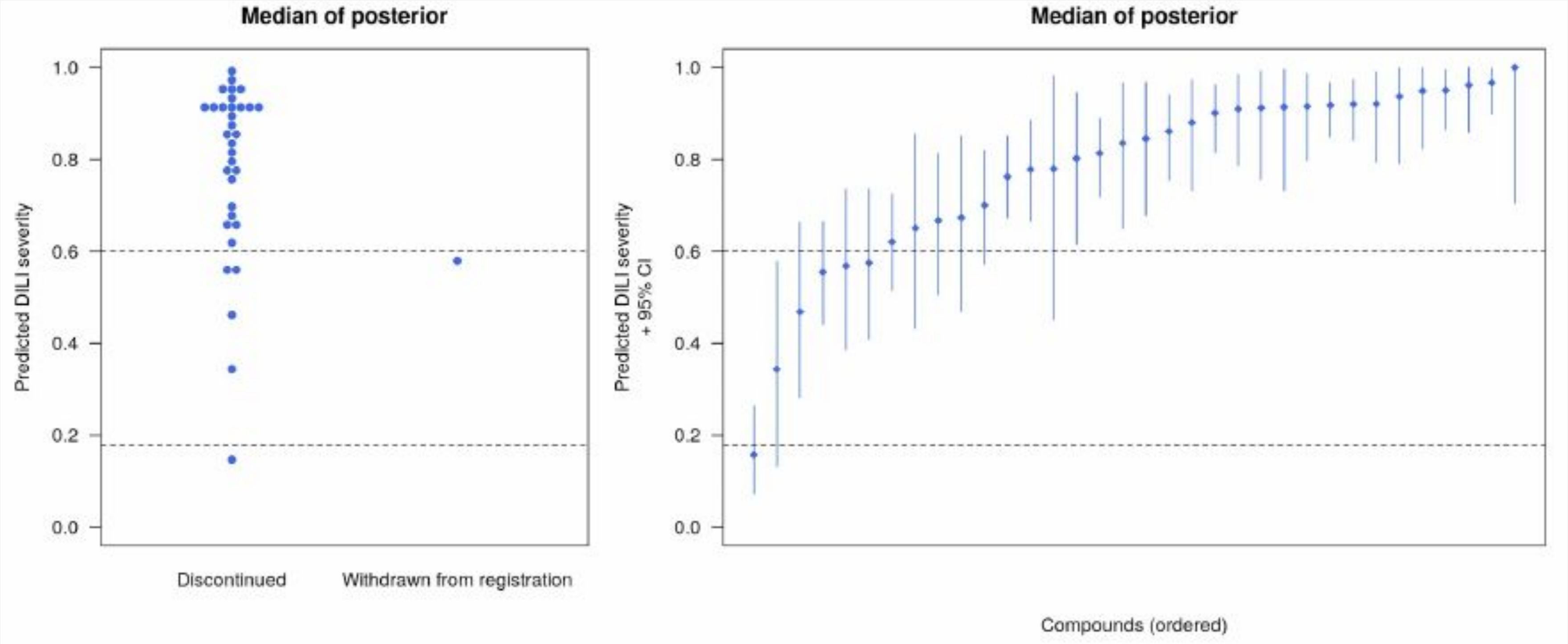
- $1/\sigma$ = “weight of evidence”.
- But it does not propagate to the final prediction.



Drug induced liver injury (DILI) example

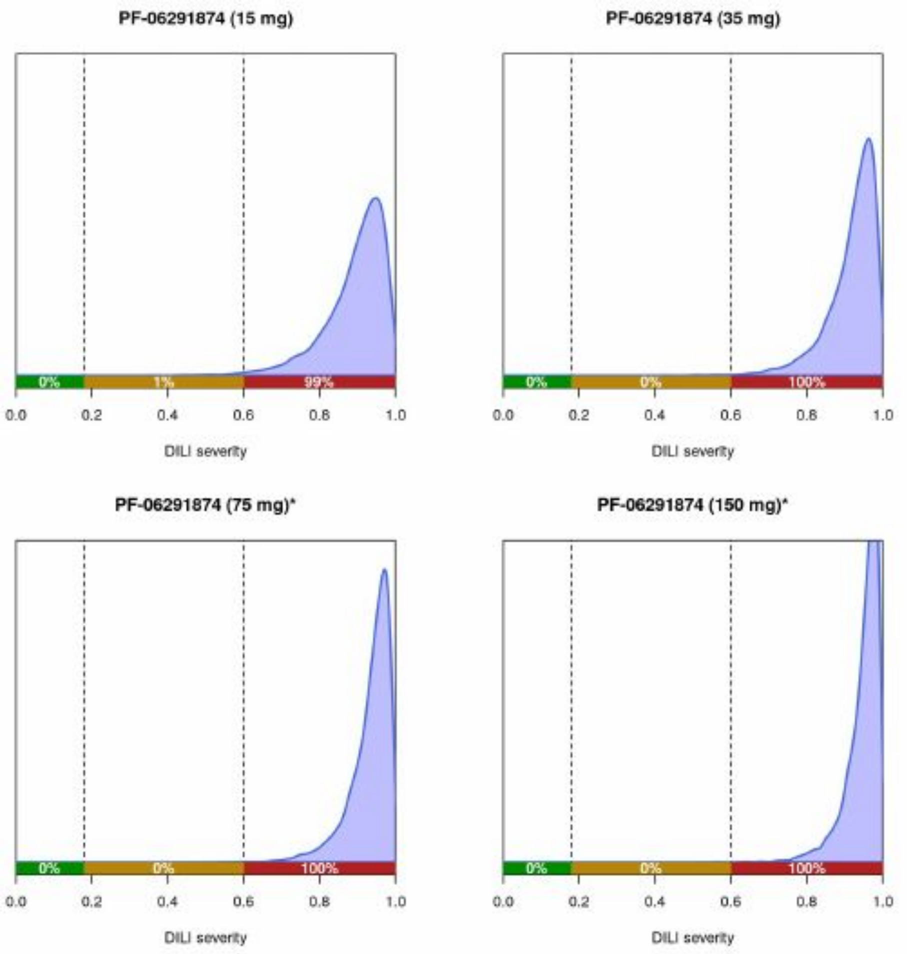


DILI example (test data)



DILI example

The model detects a dose-response relationship with DILI severity for a test compound.



Julia + Turing.jl example

```julia
@model hetero_var_model(x, y) = begin

    θ₁ ~ TruncatedNormal(1, 5, 0, Inf)
    θ₂ ~ TruncatedNormal(0, 5, 0, Inf)
    θ₃ ~ TruncatedNormal(0, 2, 0, Inf)
    σ₀ ~ Normal(0, 10)
    σ₁ ~ Normal(0, 10)

    μ = θ₃ + θ₂ * (1 - exp(-θ₁ * x))
    σ = log(1 + exp(σ₀ + σ₁ * μ))
    y ~ Normal(μ, σ)
end
```

References

1) Semenova E, Guerriero ML, Zhang B, Hock A, Hopcroft P, Kadamur G, Afzal AM, **Lazic SE** (2021). Flexible fitting of PROTAC concentration-response curves with Gaussian Processes. *bioRxiv*

2) Semenova E, Williams DP, Afzal AM, **Lazic SE** (2020). A Bayesian neural network for toxicity prediction. *Computational Toxicology* 16:100133.

3) Williams DP, **Lazic SE**, Foster AJ, Semenova E, Morgan P (2020). Predicting drug-induced liver injury with Bayesian machine learning. *Chemical Research in Toxicology* 33(1):239–248.

4) **Lazic SE**, Edmunds N, Pollard CE (2018). Predicting drug safety and communicating risk: benefits of a Bayesian approach. *Toxicological Sciences* 162(1):89–98.

Thank you!

Questions?

stan.lazic@prioris.ai
@stanlazic