Building Robust PK/PD Population Models with Bayesian Inference

Michael Betancourt
University of Warwick
Sebastian Weber
Novartis Pharma AG, BDM Oncology, Basel

PAGE 2015
Hersonissos, Crete, Greece
June 4, 2015

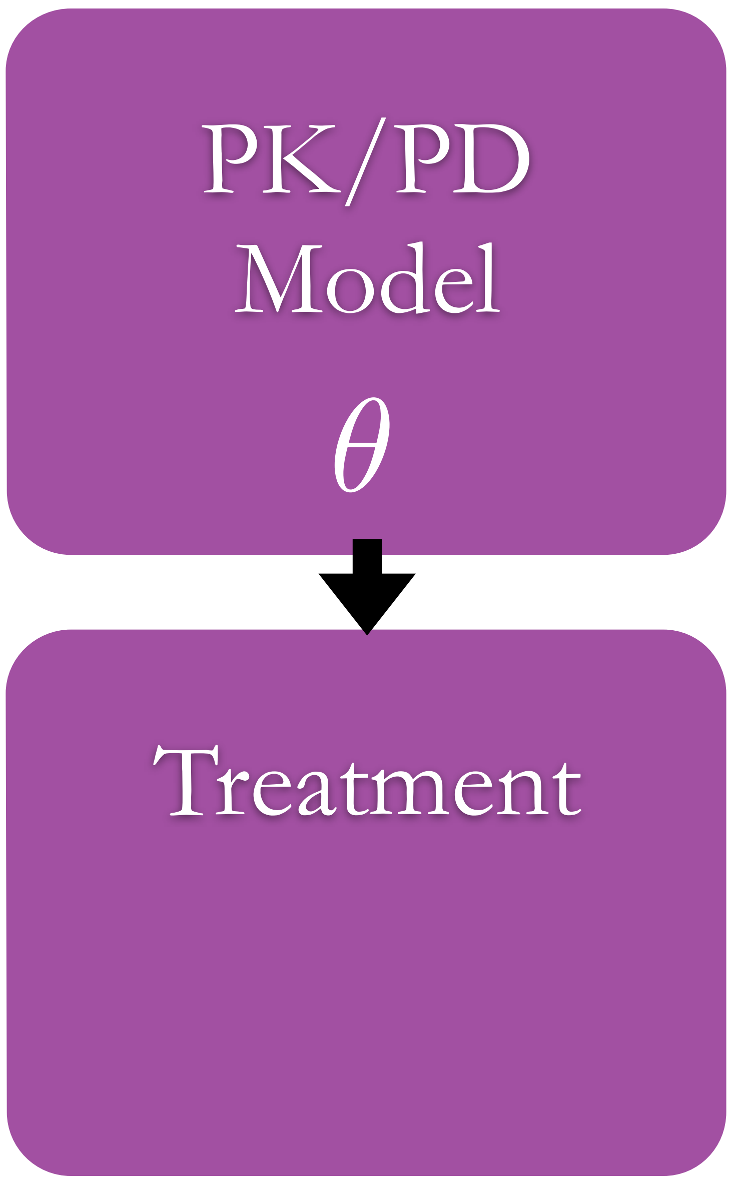
Clinical neonate example in scientific collaboration with M. Pfister, university children's hospital Basel (UKBB)

Because of the clinical applications, precise PK/PD modeling is incredibly important.

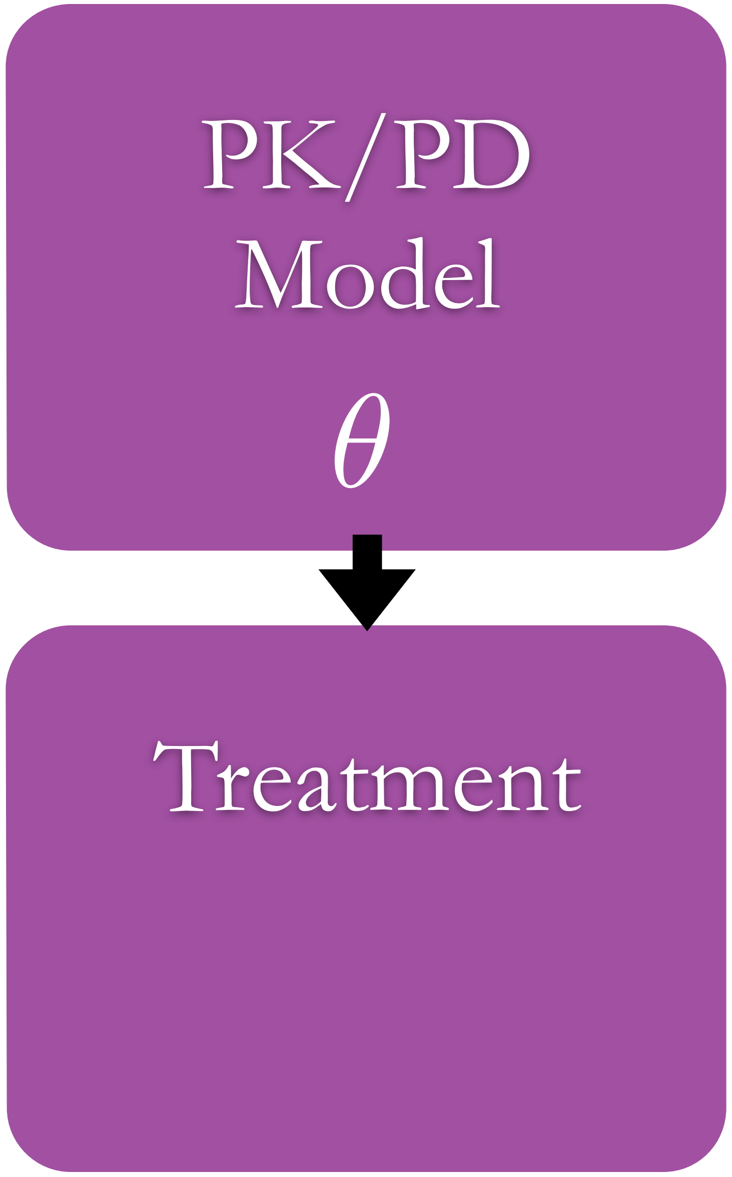
PK/PD Model

θ

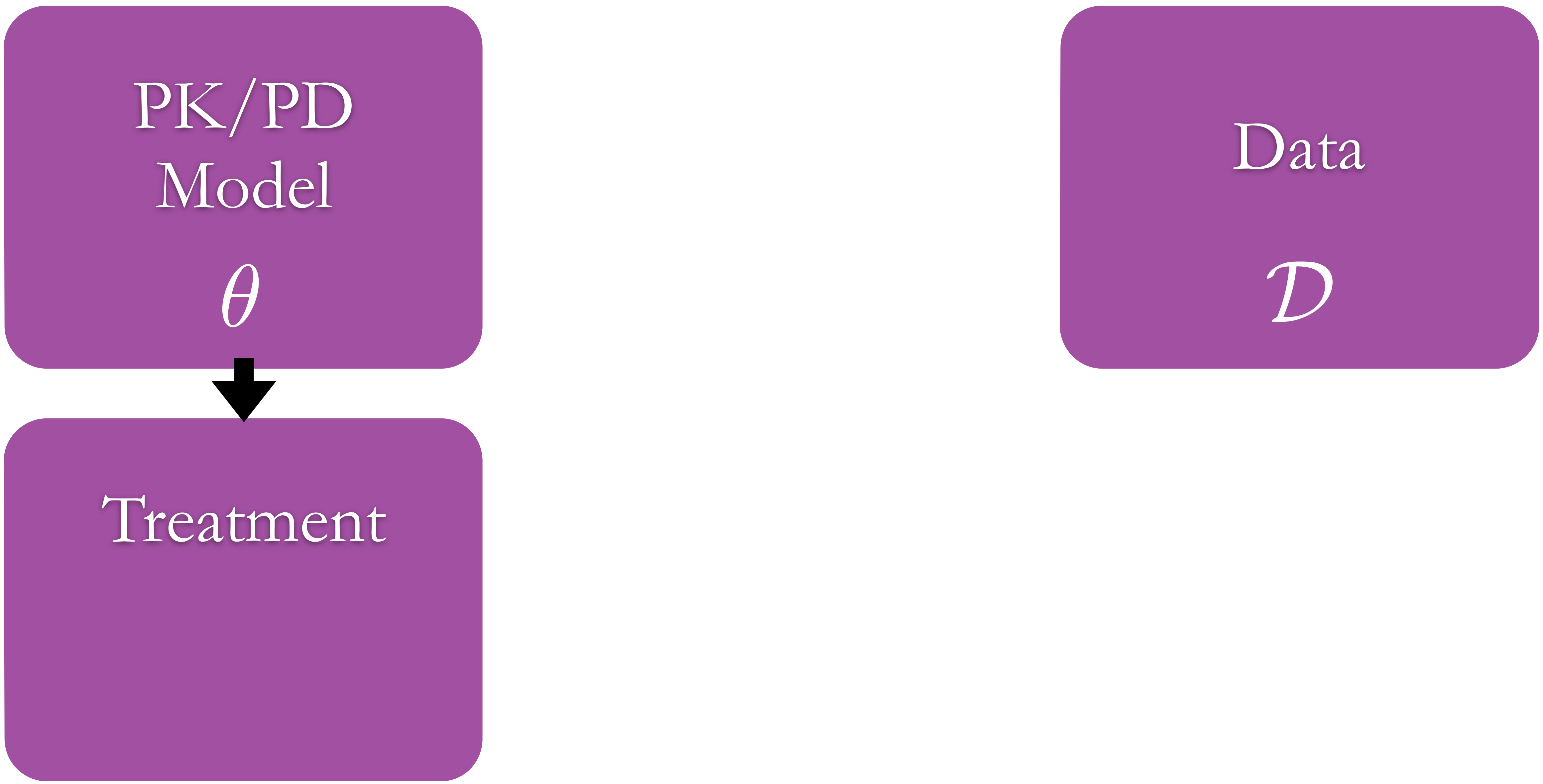
Because of the clinical applications, precise PK/PD modeling is incredibly important.



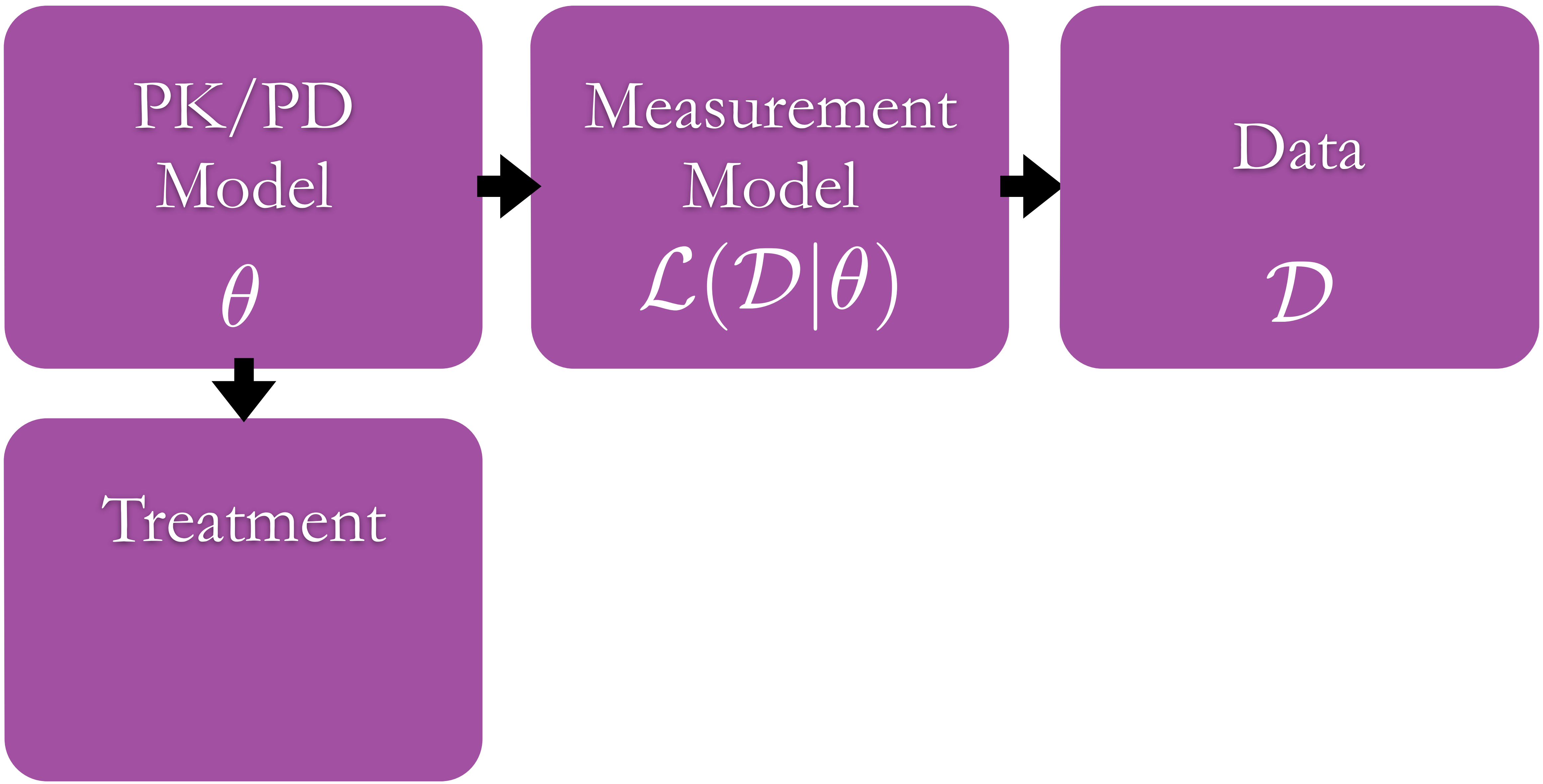
Robust treatment decisions requires modeling both the latent PK/PD process and the measurement.



Robust treatment decisions requires modeling both the latent PK/PD process and the measurement.



Robust treatment decisions requires modeling both the latent PK/PD process and the measurement.

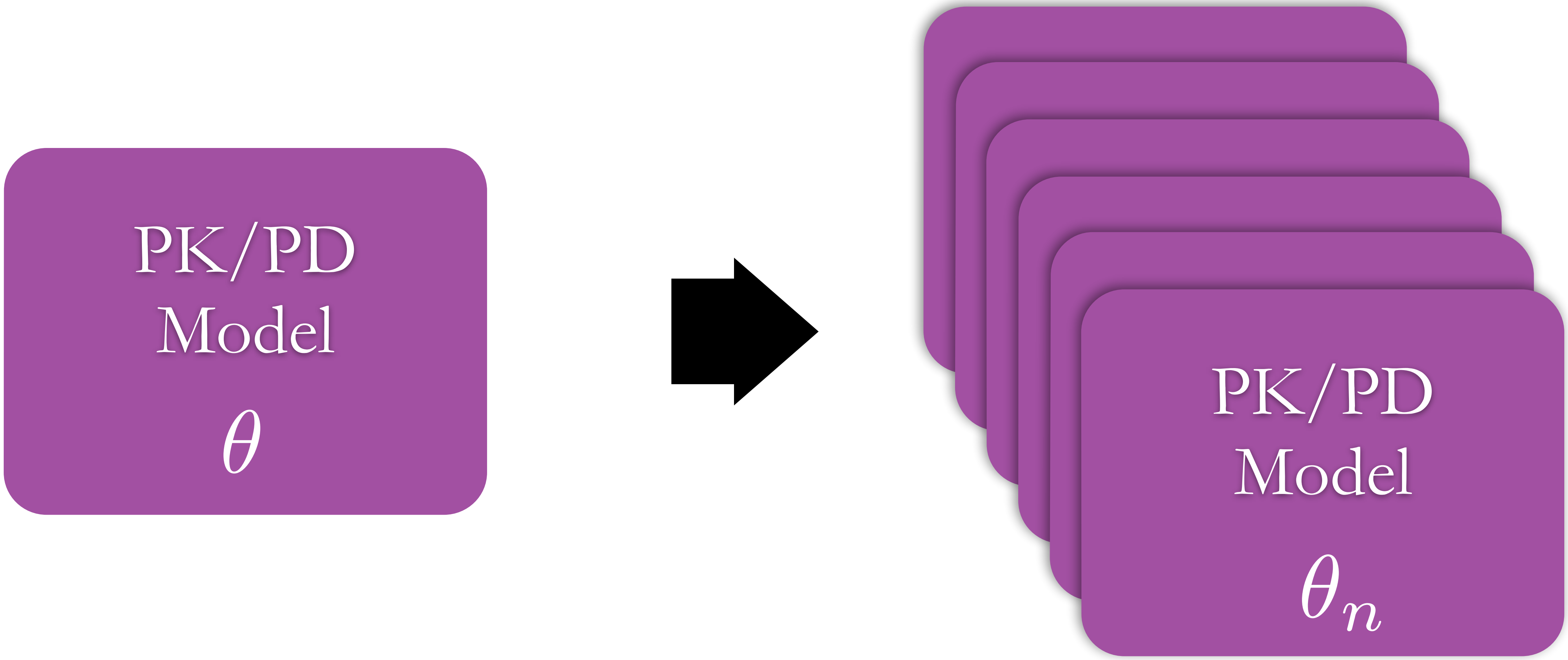


The model becomes substantially more complicated once we generalize to populations of patients.

PK/PD Model

θ

The model becomes substantially more complicated once we generalize to populations of patients.



No pooling of data avoids any bias, but at the expense of sparse data and large uncertainties.

θ_1	$\cdots$	θ_n	$\cdots$	θ_N

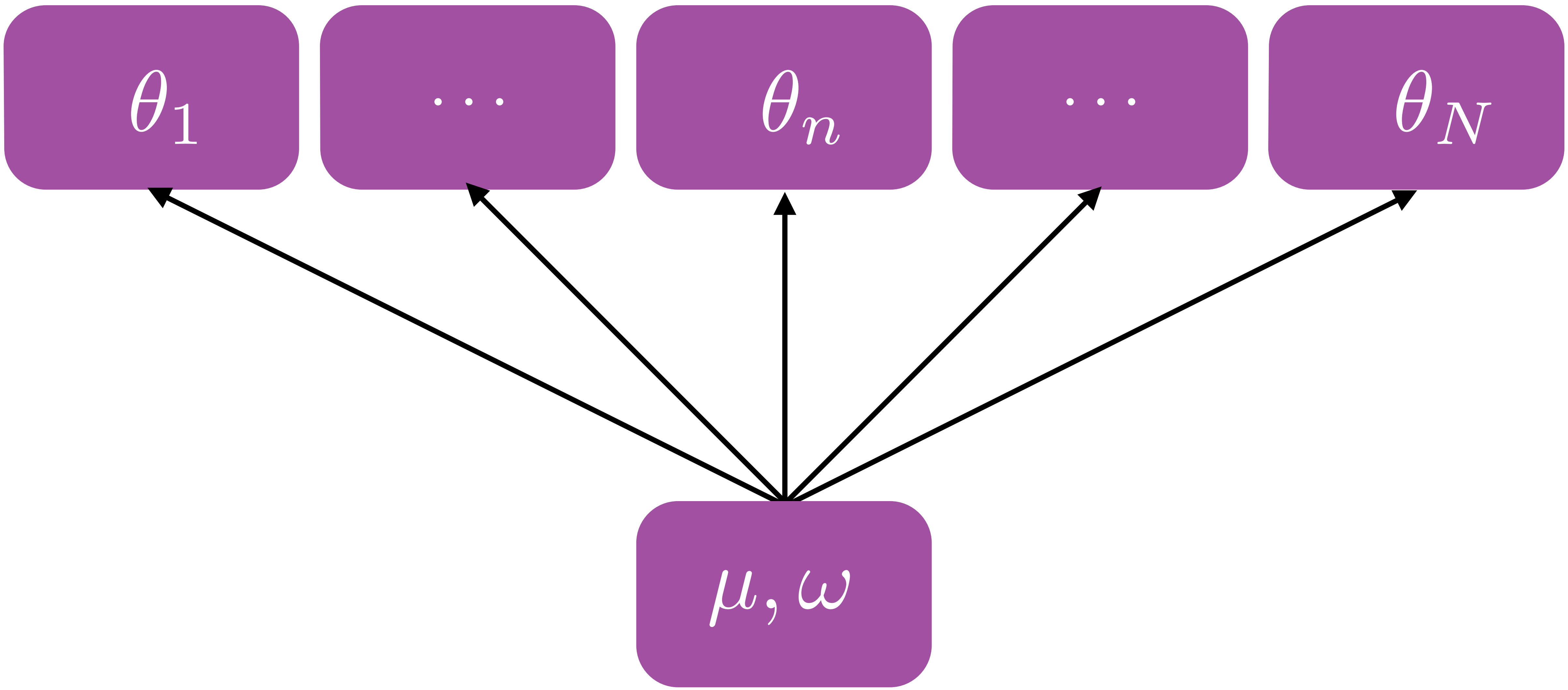
Complete pooling assumes that all patients are identical, reducing variance at the expense of bias.

Complete pooling assumes that all patients are identical, reducing variance at the expense of bias.

Hierarchical models explicitly model the population, allowing for partial pooling of the data.

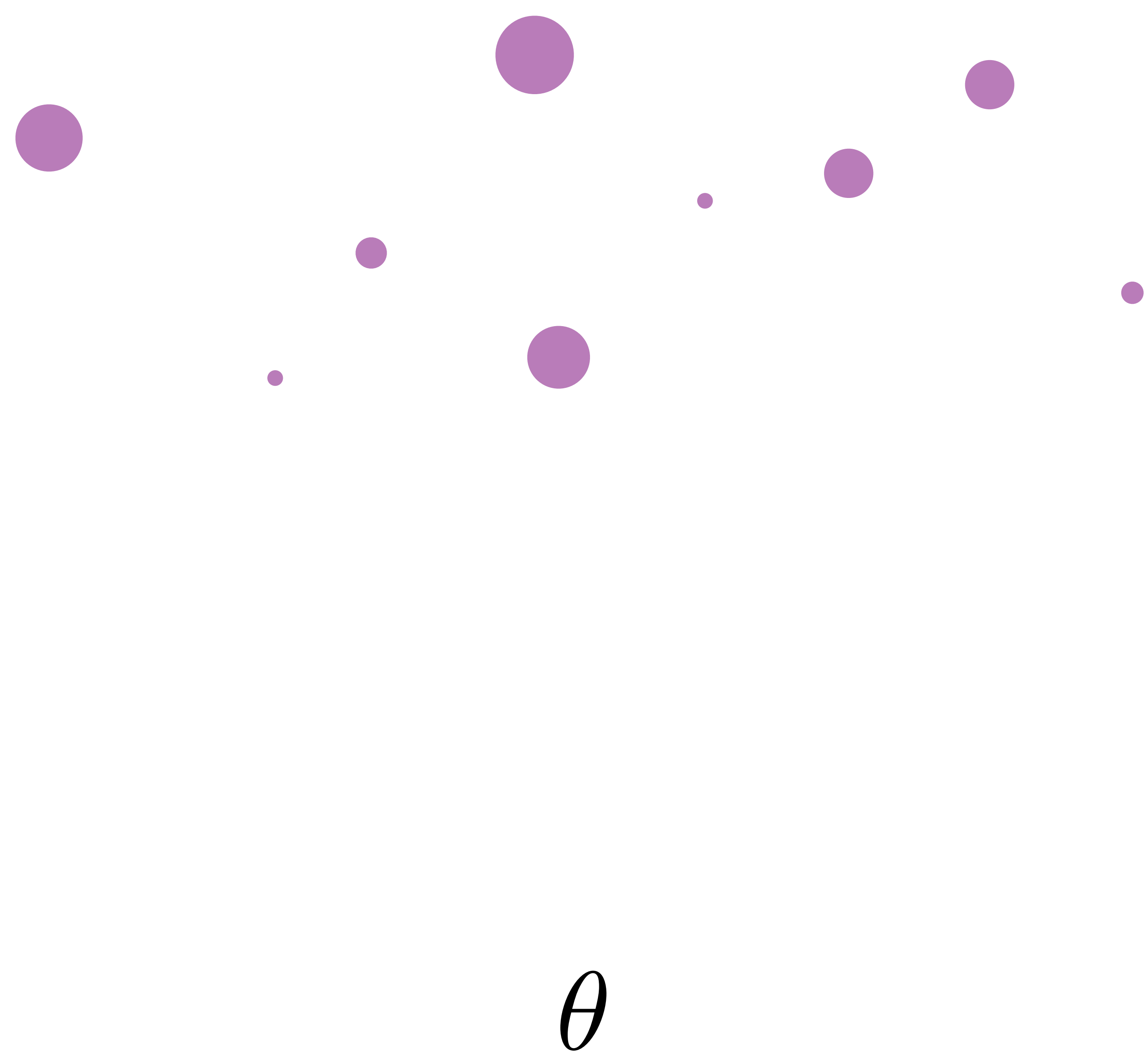
$$\theta_1 \quad \cdots \quad \theta_n \quad \cdots \quad \theta_N$$

Hierarchical models explicitly model the population, allowing for partial pooling of the data.

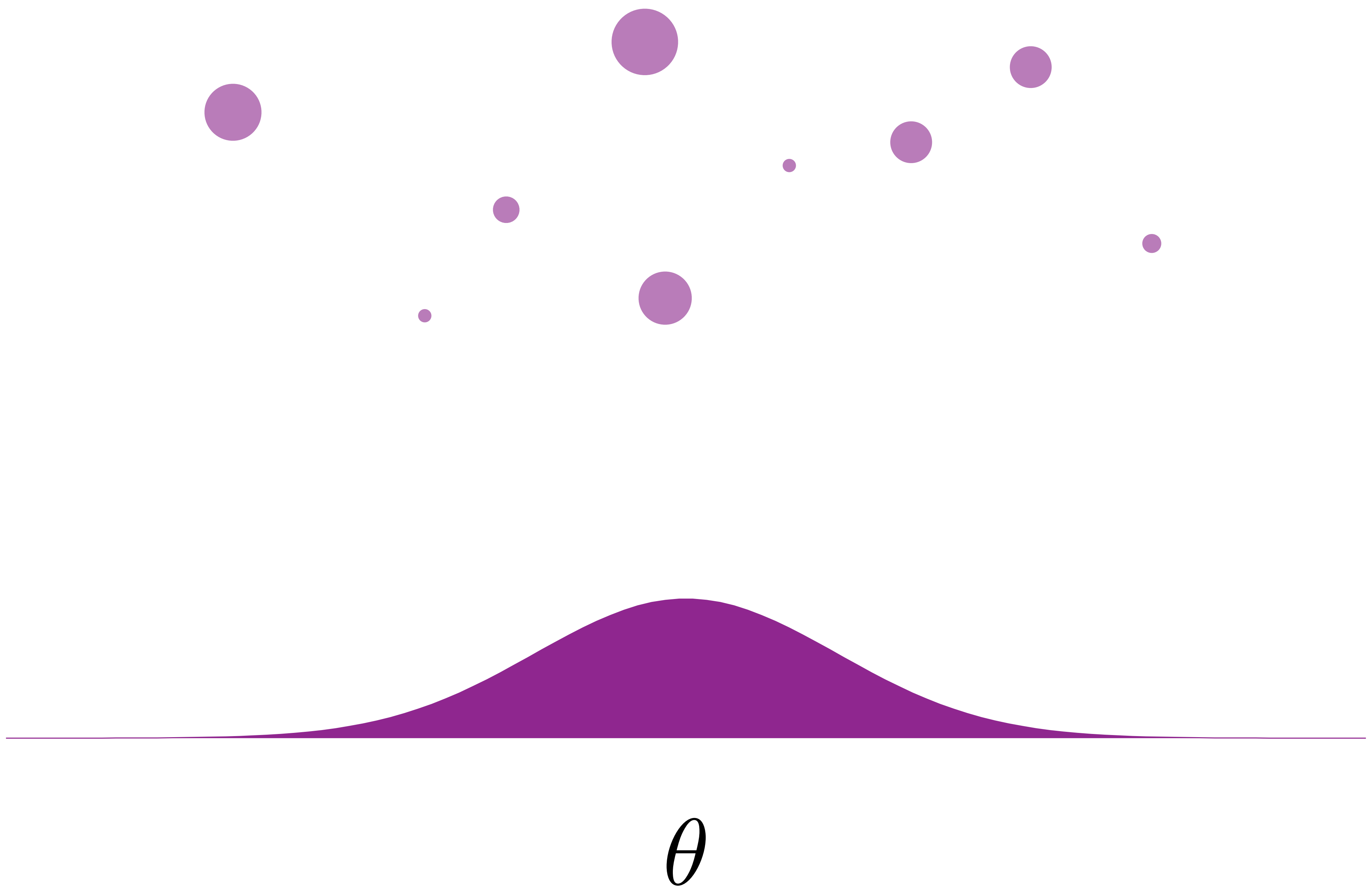


$$\theta_n \sim \mathcal{N}(\mu, \omega)$$

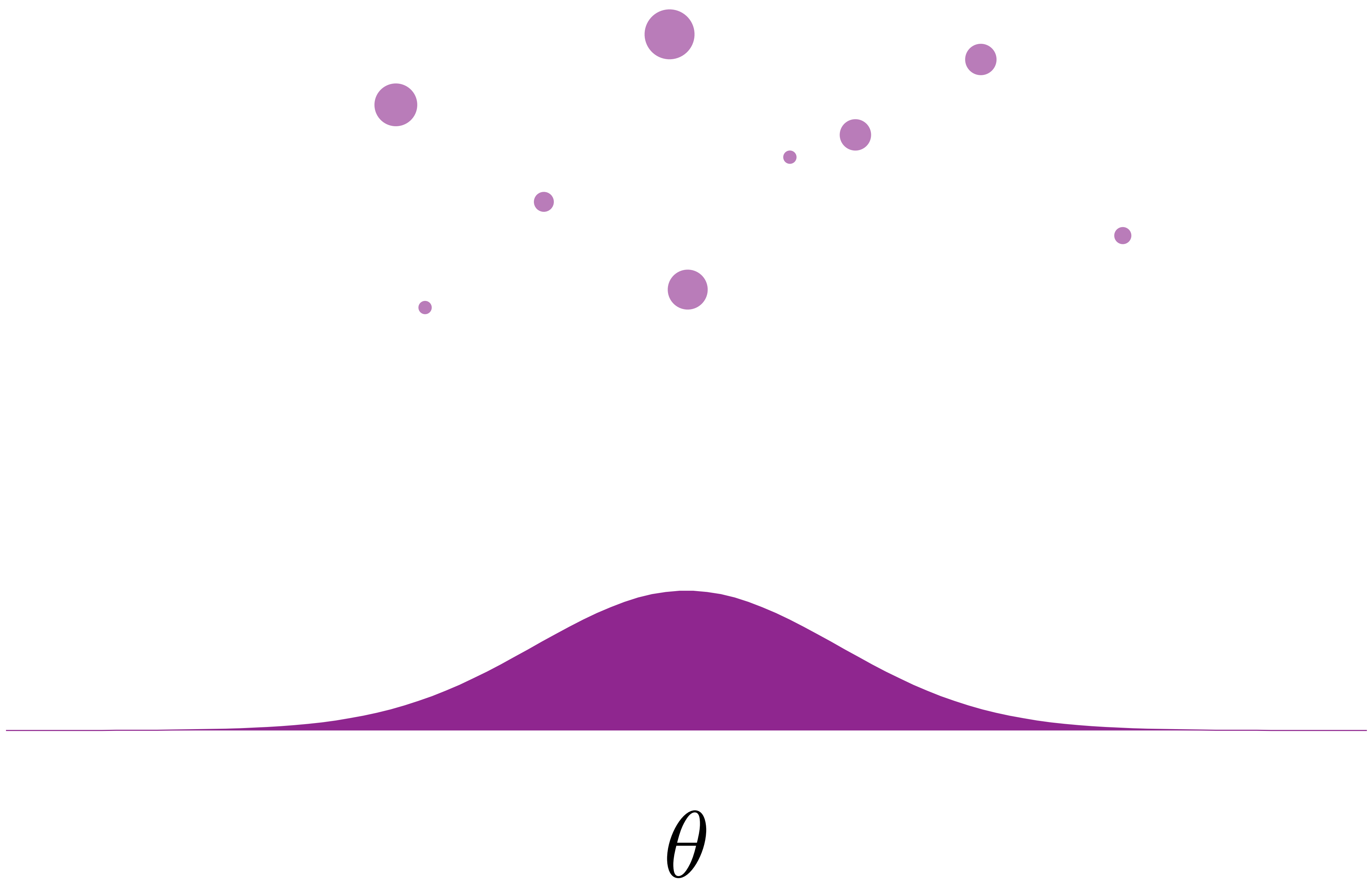
Partial pooling dynamically trades off between bias and variance to achieve improved inferences.



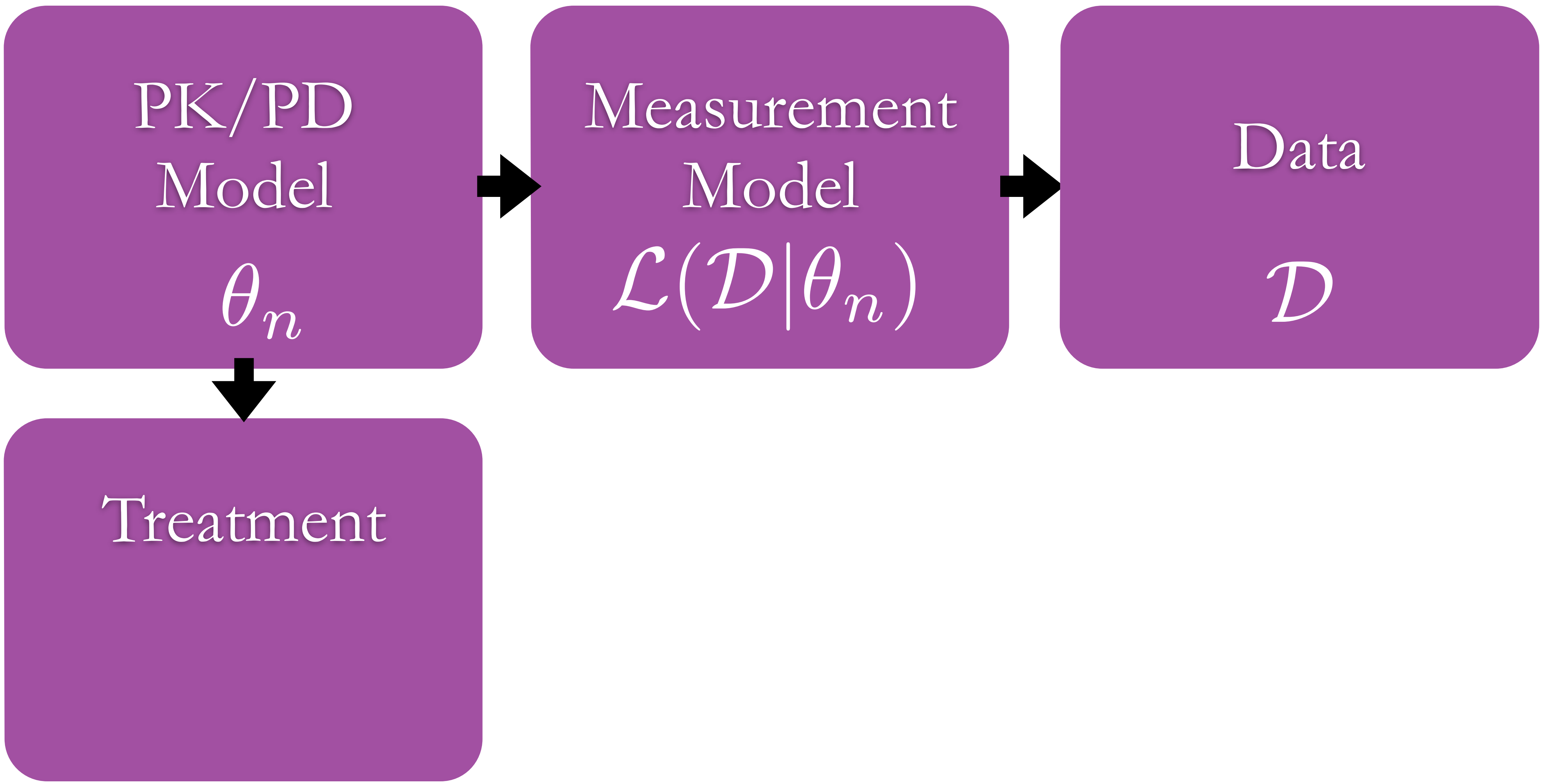
Partial pooling dynamically trades off between bias and variance to achieve improved inferences.



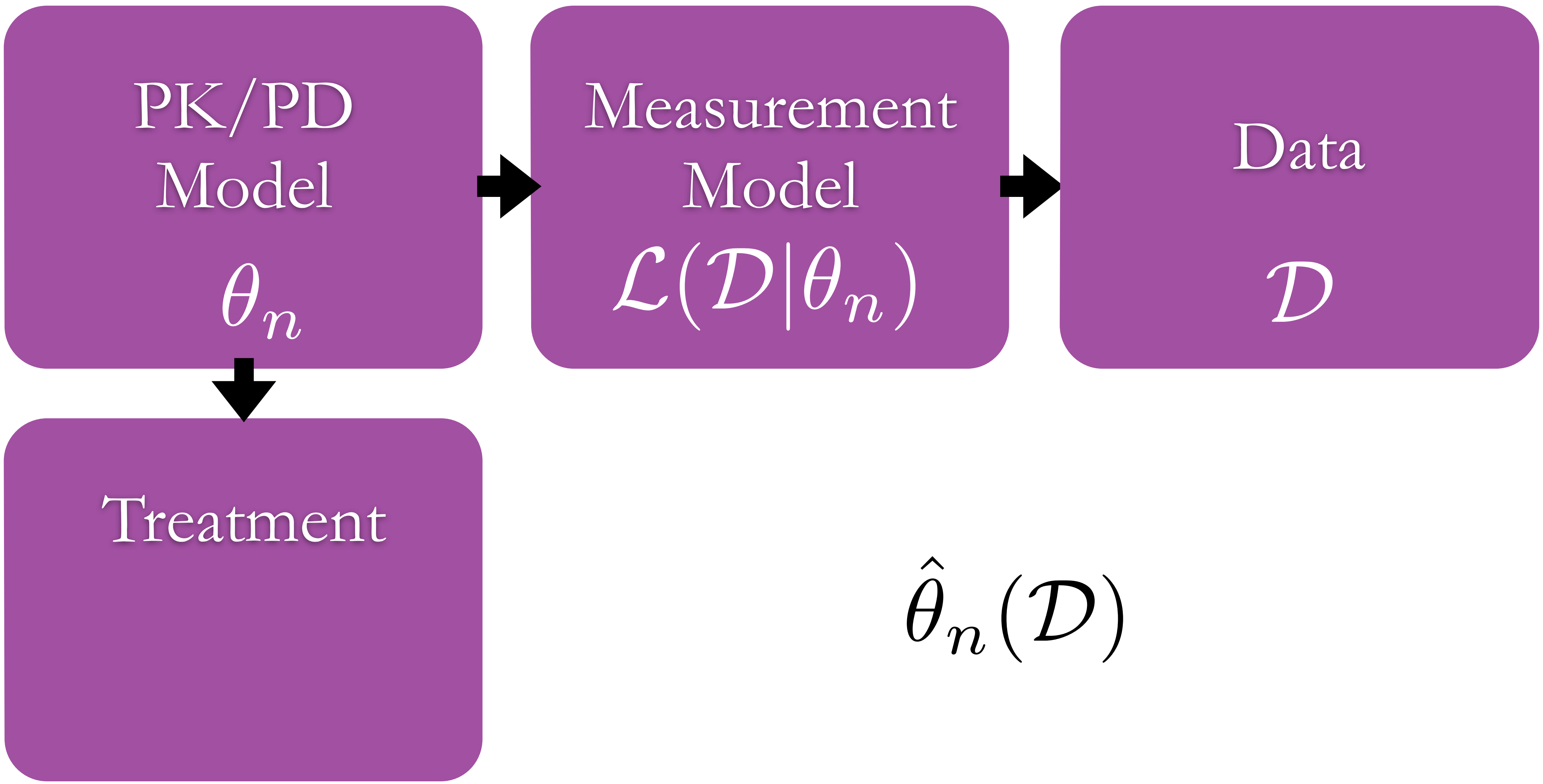
Partial pooling dynamically trades off between bias and variance to achieve improved inferences.



The ultimate utility of these population models, however, depends on how we learn from the data.



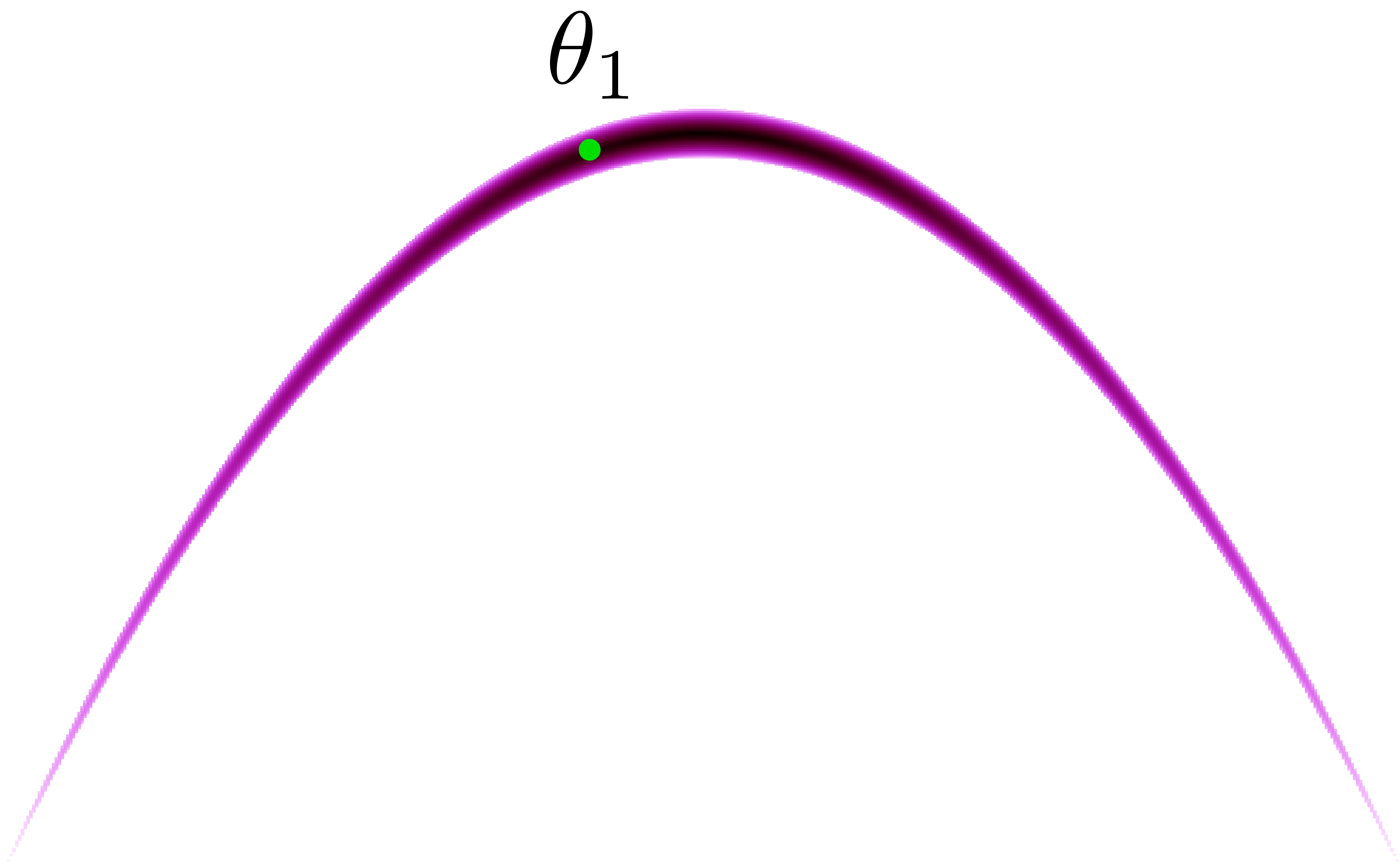
In frequentist statistics we construct *estimators*, or functions of the data, that resemble the true parameters.



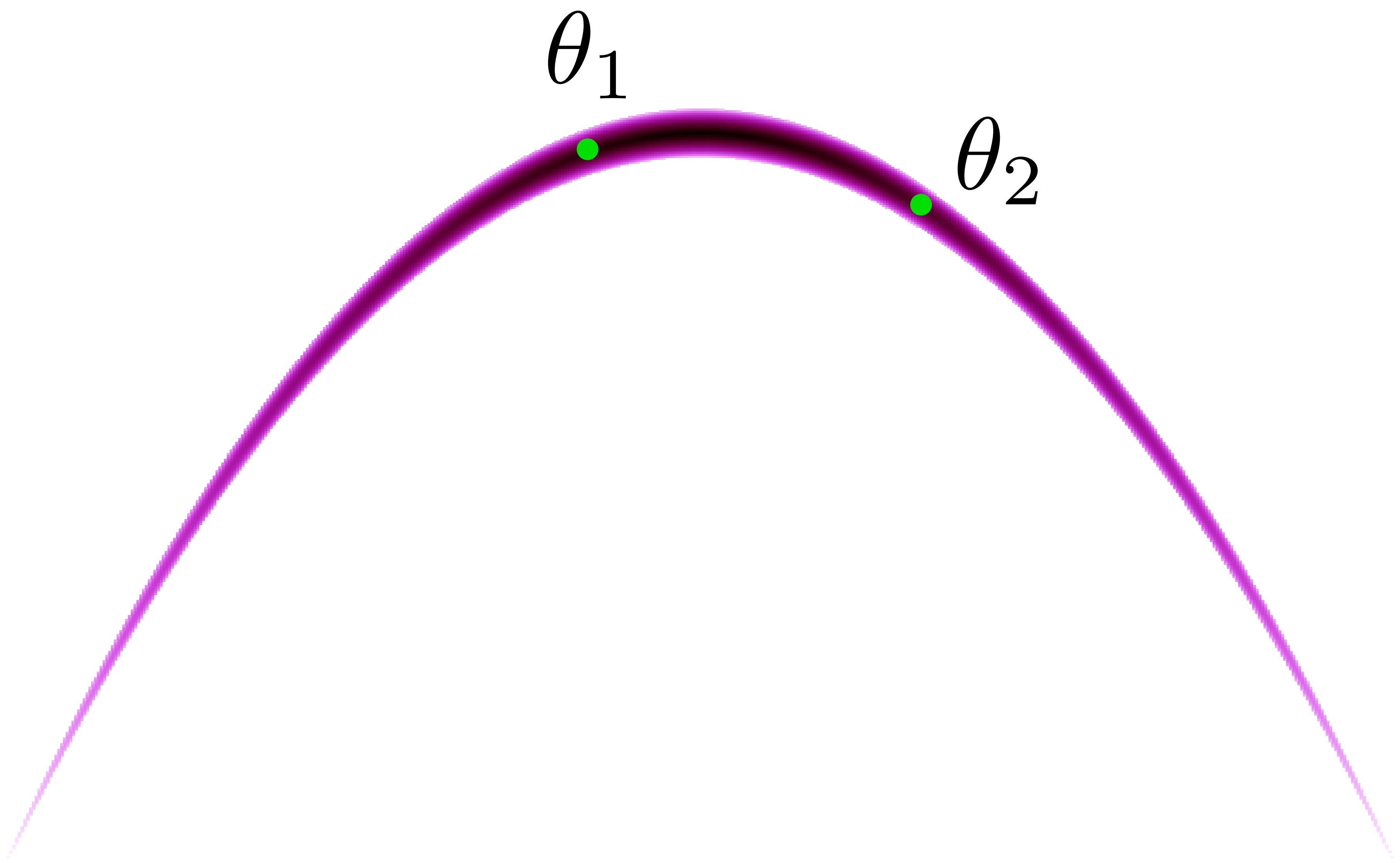
$$\hat{\theta}_n(\mathcal{D})$$

Estimators can be dangerous because they have trouble incorporating uncertainty, leading to poor decisions.

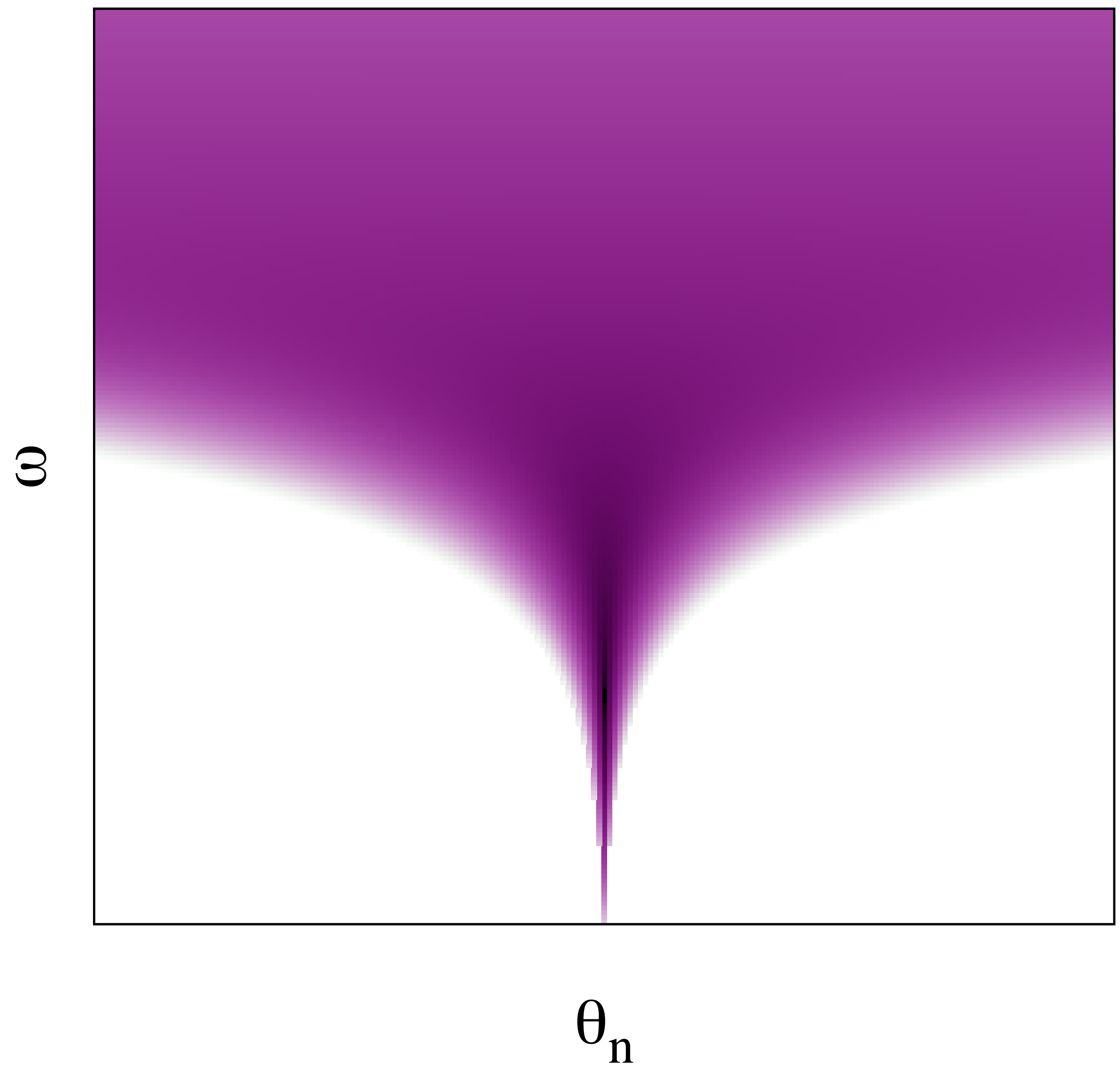
Estimators can be dangerous because they have trouble incorporating uncertainty, leading to poor decisions.



Estimators can be dangerous because they have trouble incorporating uncertainty, leading to poor decisions.



This is particularly dangerous in population models where the clinical data is sparse.



Bayesian inference embraces uncertainty, constructing an entire distribution instead of just a point estimate.

Bayesian inference embraces uncertainty, constructing an entire distribution instead of just a point estimate.

$$\pi(\theta)$$

Bayesian inference embraces uncertainty, constructing an entire distribution instead of just a point estimate.

$$\pi(\mathcal{D}|\theta)\,\pi(\theta)$$

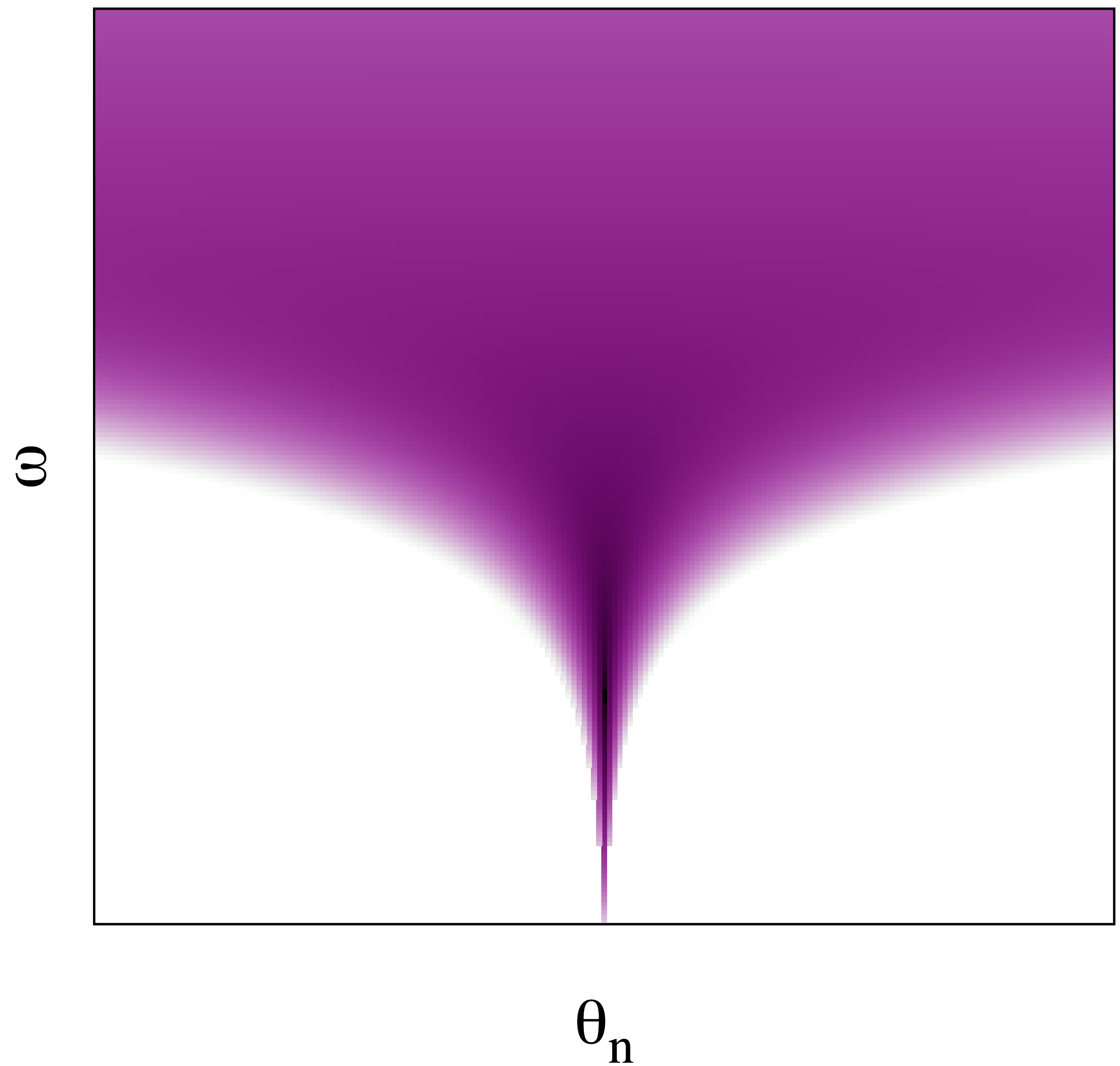
Bayesian inference embraces uncertainty, constructing an entire distribution instead of just a point estimate.

$$\pi(\theta|\mathcal{D}) \propto \pi(\mathcal{D}|\theta)\,\pi(\theta)$$

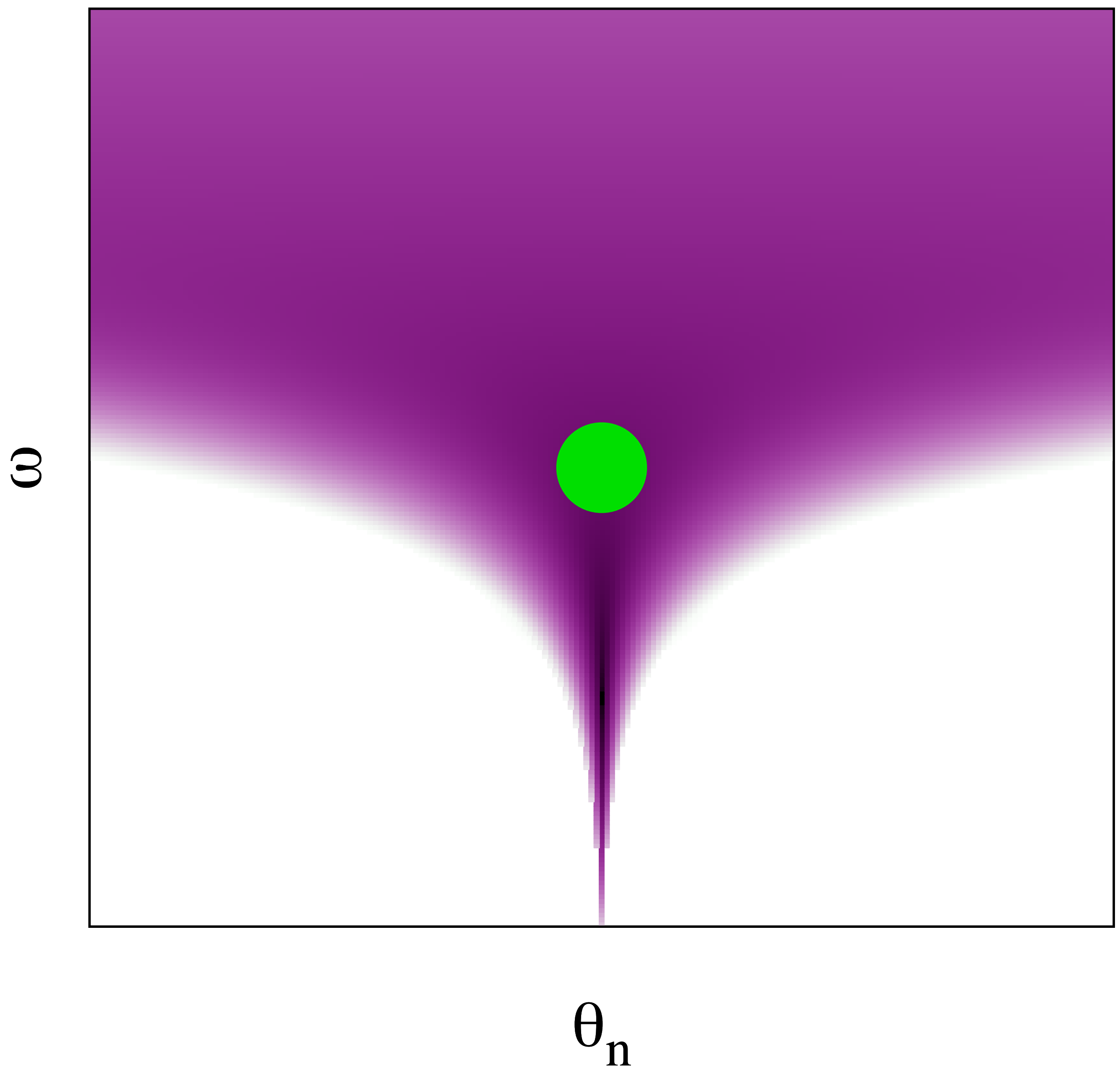
The prior distribution defines a coherent means of incorporating additional, regularizating information.

$$\pi(\theta|\mathcal{D}) \propto \pi(\mathcal{D}|\theta)\, \pi(\theta)$$

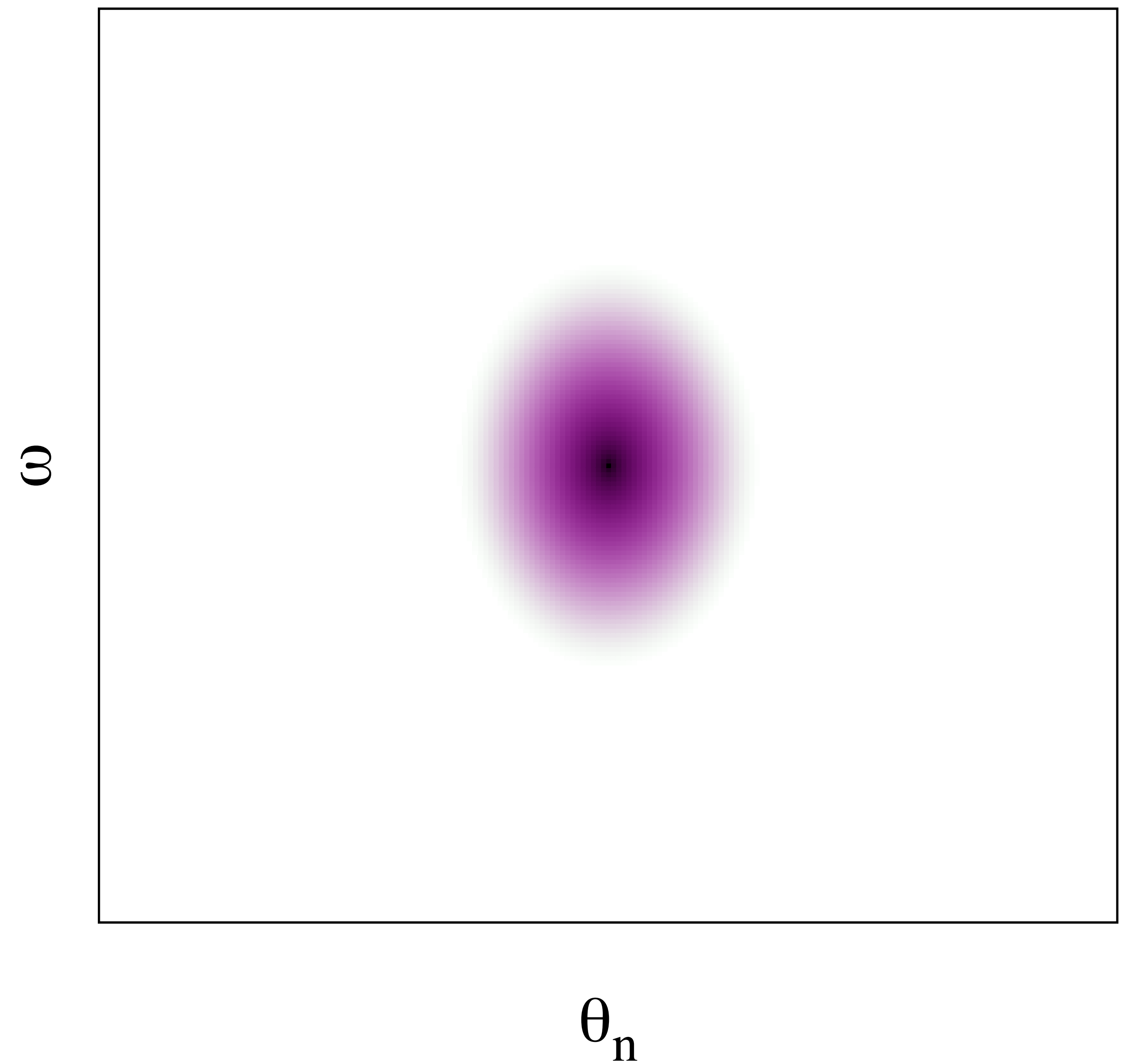
The prior distribution defines a coherent means of incorporating additional, regularizating information.



The prior distribution defines a coherent means of incorporating additional, regularizating information.



The prior distribution defines a coherent means of incorporating additional, regularizating information.



Moreover, given a posterior we can incorporate all of the uncertainty into our decision with the *expected risk*.

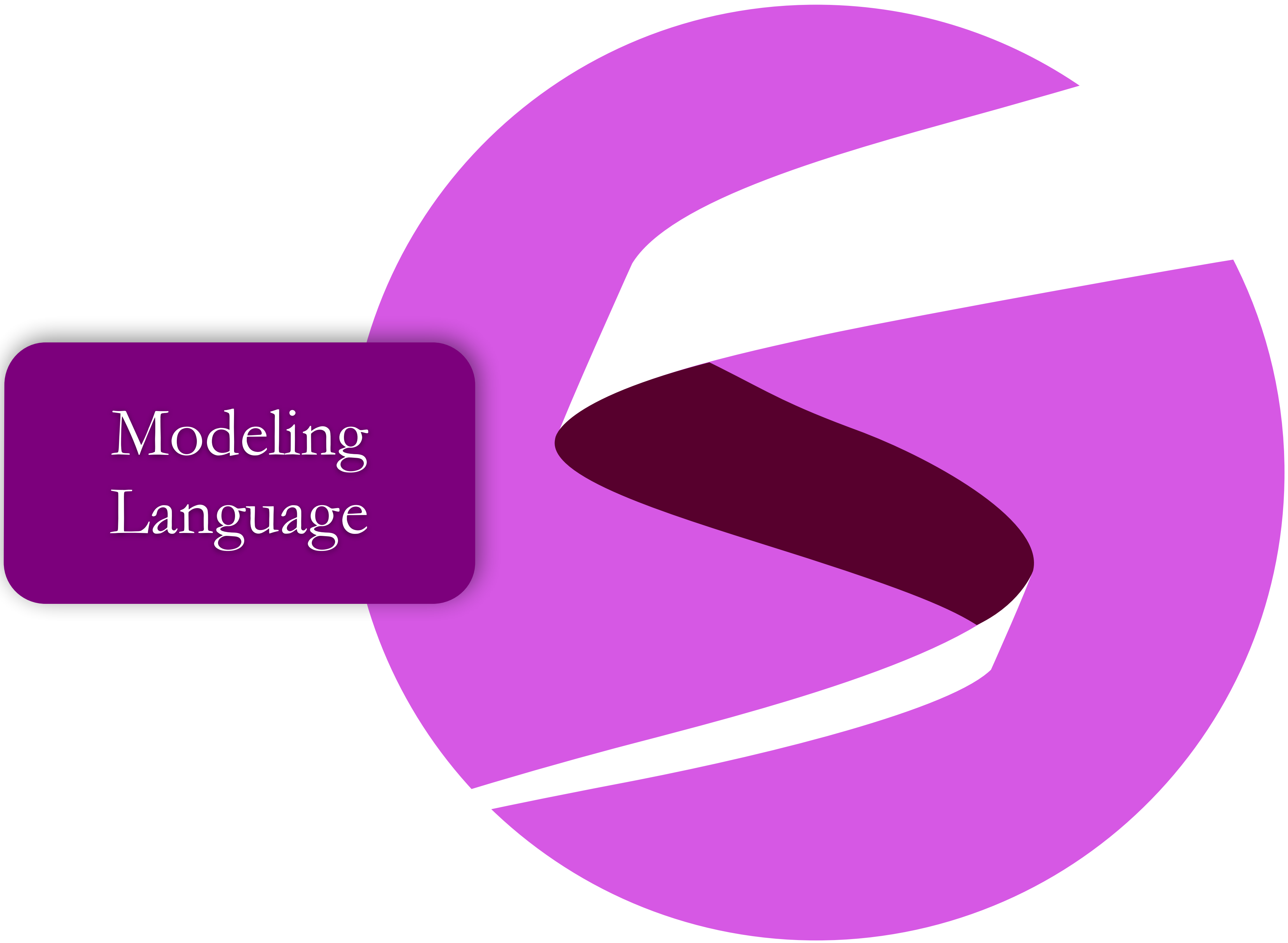
$$\mathbb{E}\left[f\right] = \int \mathrm{d}\theta \, \pi(\theta | \mathcal{D}) \, f(\theta)$$

The computational challenge with these Bayesian methods is that we have to compute expectations.

$$\mathbb{E}\left[f\right] = \int \mathrm{d}\theta \, \pi(\theta|\mathcal{D}) \, f(\theta)$$

Stan provides state-of-the-art statistical tools for efficiently estimating these expectations.

Stan provides state-of-the-art statistical tools for efficiently estimating these expectations.



Stan provides state-of-the-art statistical tools for efficiently estimating these expectations.