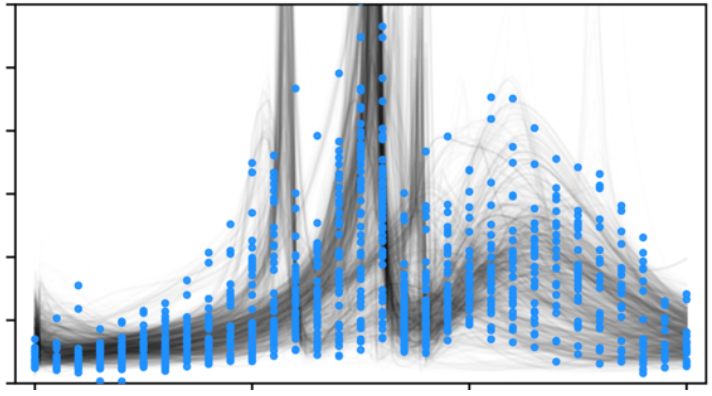
Empirical Bayes methods for prior estimation in systems biology modelling

Susanna Röblitz

Universitetet i Bergen



COMPUTATIONAL BIOLOGY UNIT

BIRS Workshop 2018, November 15

Background

Typical situaton in clinical applications:
Data available from **many** patients, but **few** data per individual.

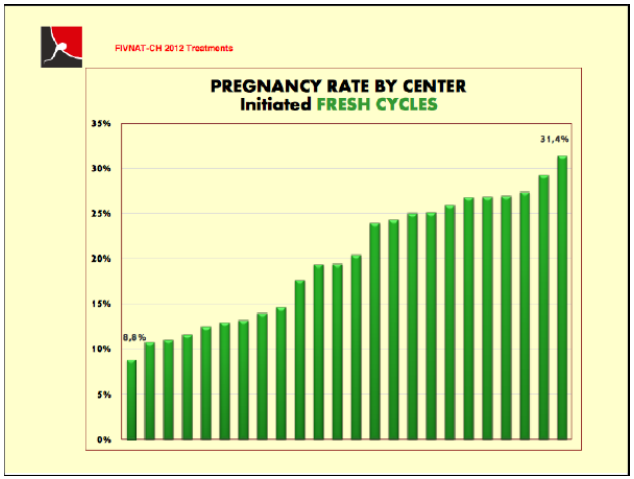
Background

Typical situaton in clinical applications:
Data available from **many** patients, but **few** data per individual.

Example: Fertility treatment in reproductive endocrinology

Large variation in success rates among clinics (8-35%) due to

- different treatment protocols,
- high intra- and inter-individul variability in the female menstrul cycle



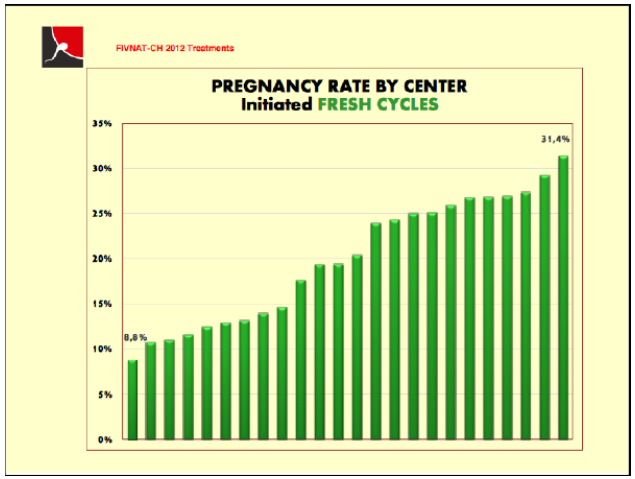
Background

Typical situaton in clinical applications:
Data available from **many** patients, but **few** data per individual.

Example: Fertility treatment in reproductive endocrinology

Large variation in success rates among clinics (8-35%) due to

- different treatment protocols,
- high intra- and inter-individul variability in the female menstrul cycle

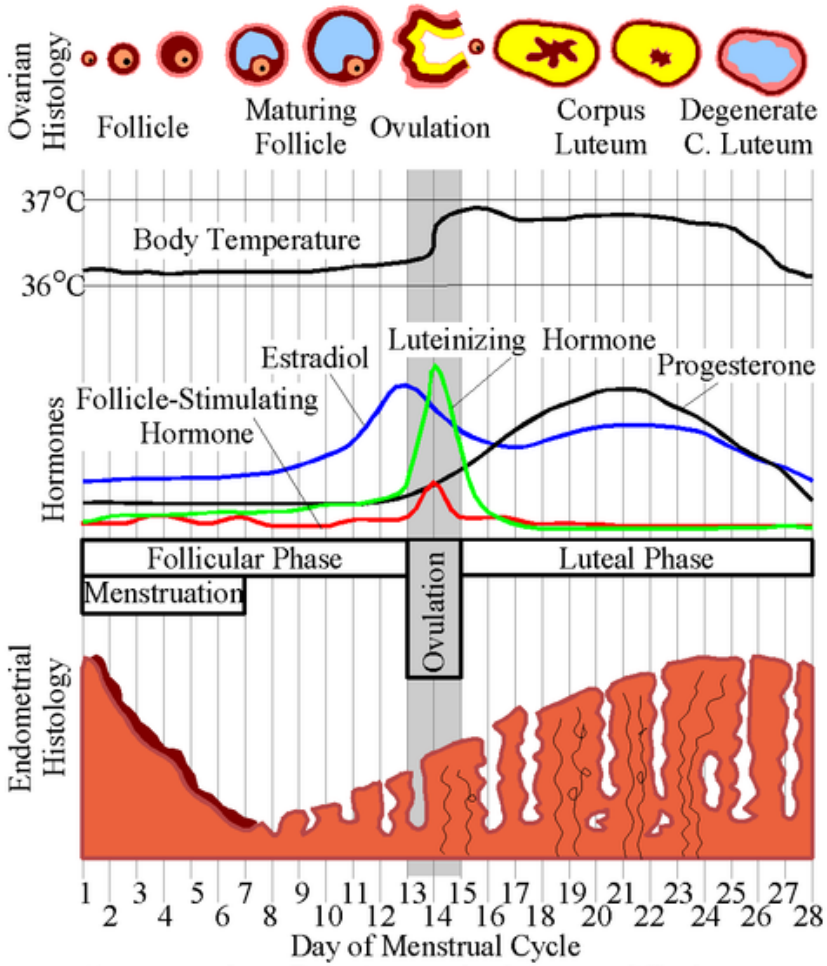


Question: Can we construct a mechanistic model of the menstrual cycle that displays this variability and that can be used to evaluate treatment success rates *in silico*?

Outline

1. Physiological background
2. Data
3. Model construction
4. Model parametrization
5. Model applications

The human menstrual cycle



(Average values. Durations and values may differ between different females or different cycles.)

(http://www.websters-online-dictionary.org/definiions/Menstrual Cycle)

Exactly timed interplay of physiological processes

- follicle development
- ovulation and fertilization
- formation of corpus luteum
- embryonic attachment and growth in the uterus

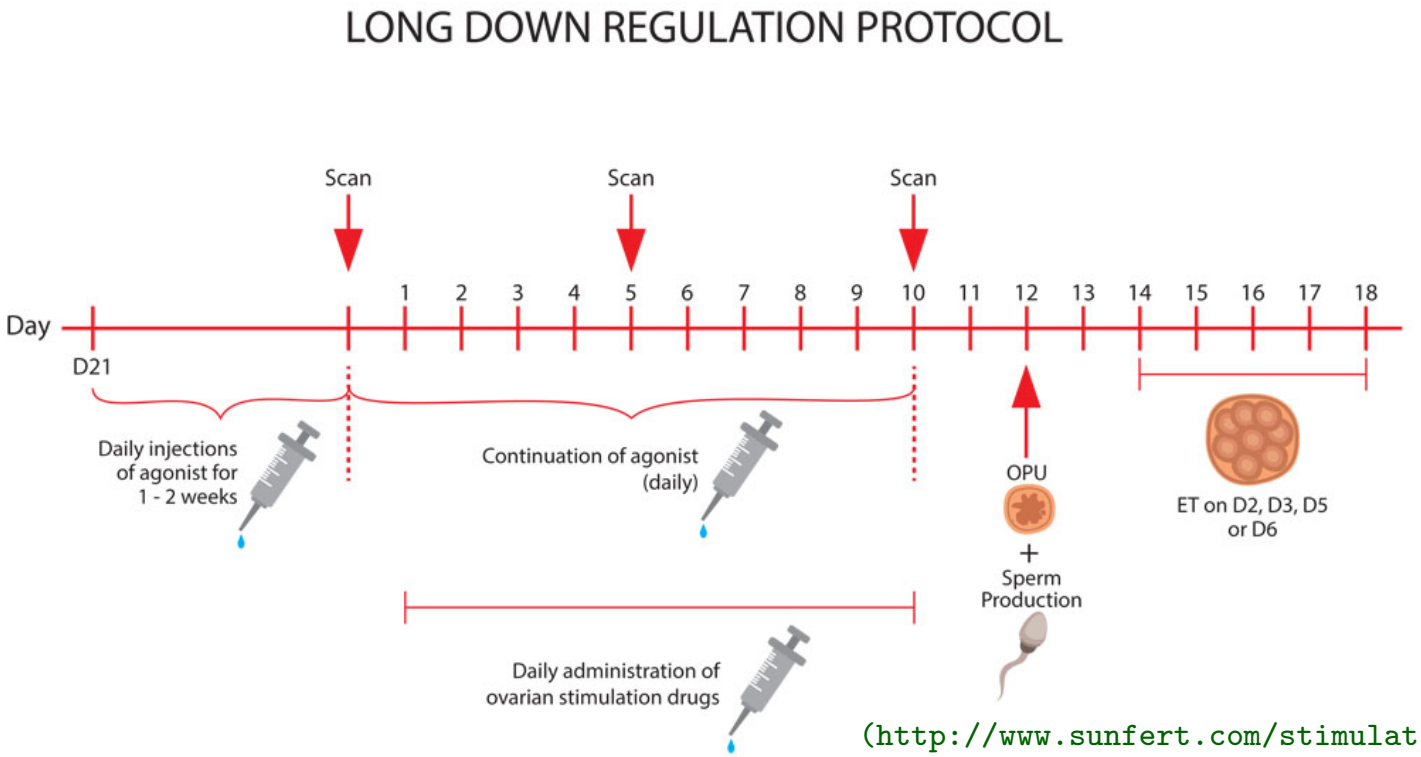
$\Rightarrow$ coordination between neural and endocrine systems

Unwanted childlessness among couples in Europe: 12-15%

Assisted reproductive technology (ART)

Ovarian stimulation and in-vitro fertilization (IVF):

downregulation + stimulation + oocyte retrieval

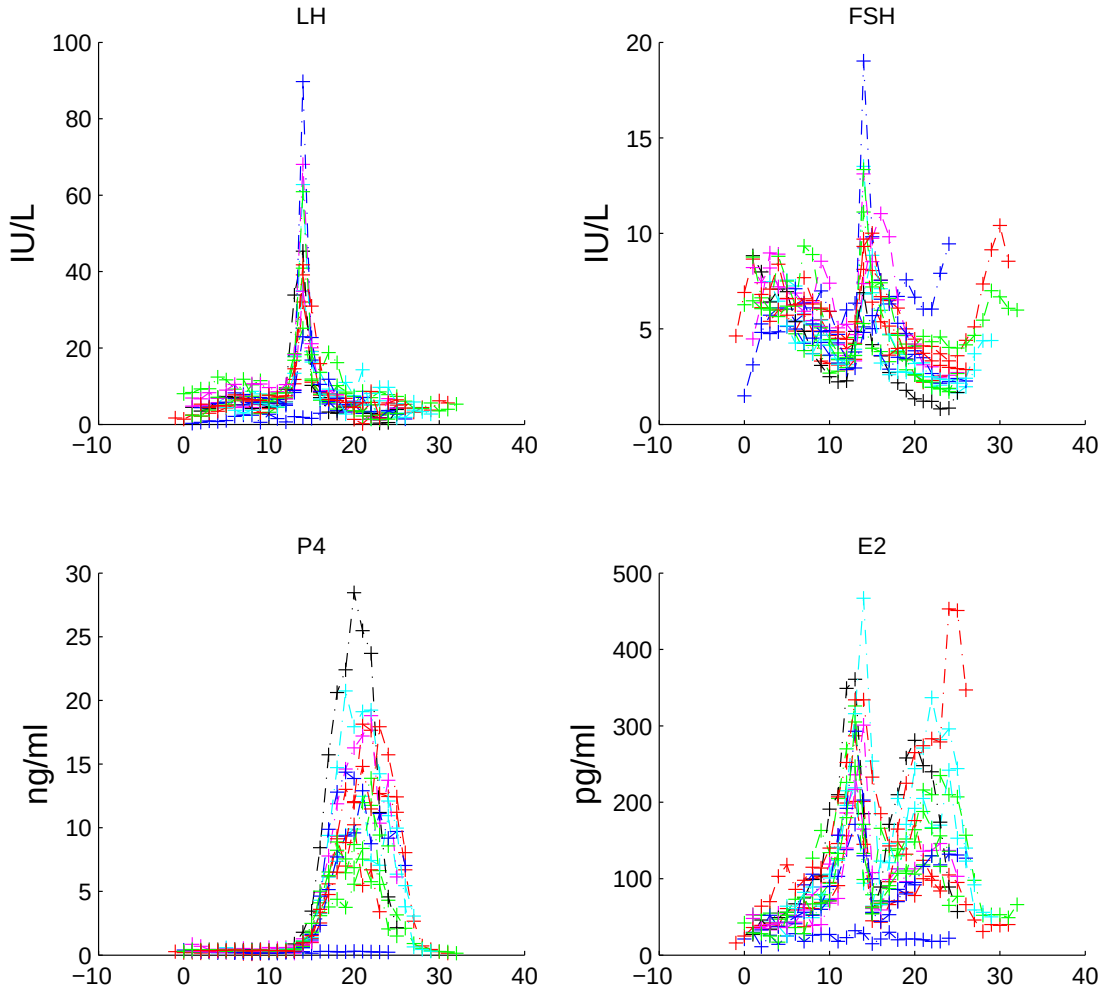


(http://www.sunfert.com/stimulation-protocols/)

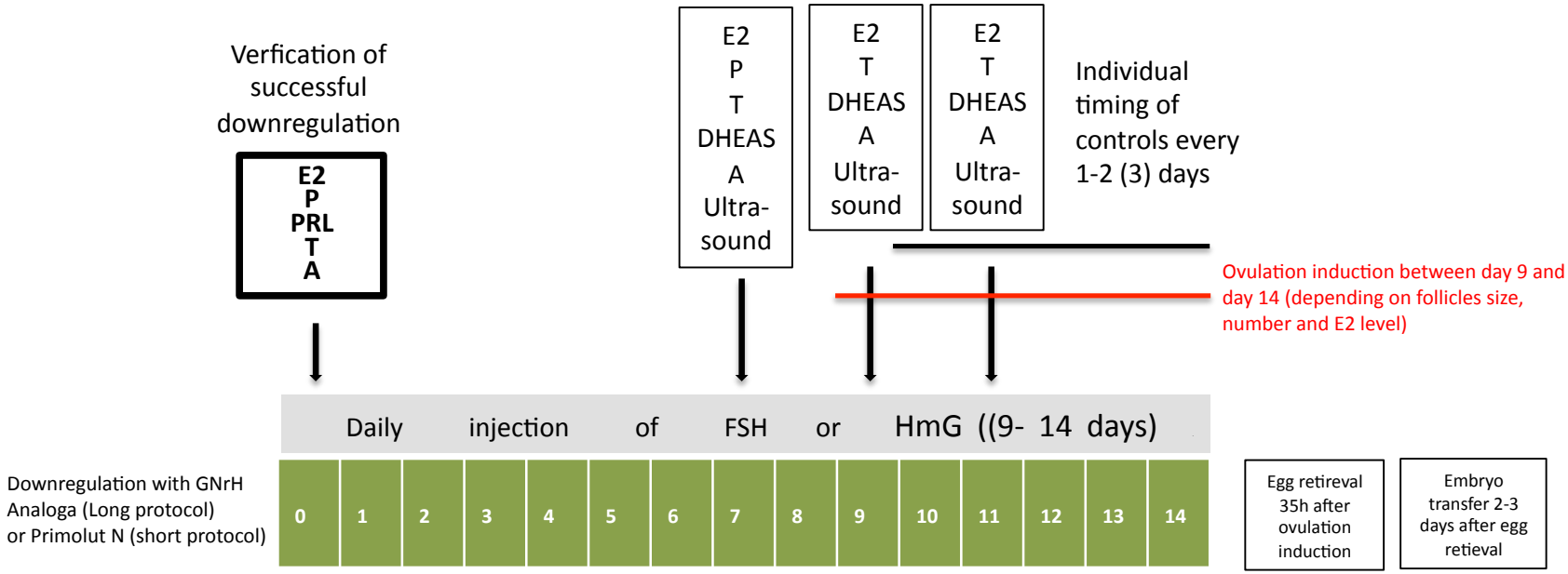
Aim: between 11 and 15 mature oocytes

Success rates: 8 - 35%, depending on the clinic

Cycle data



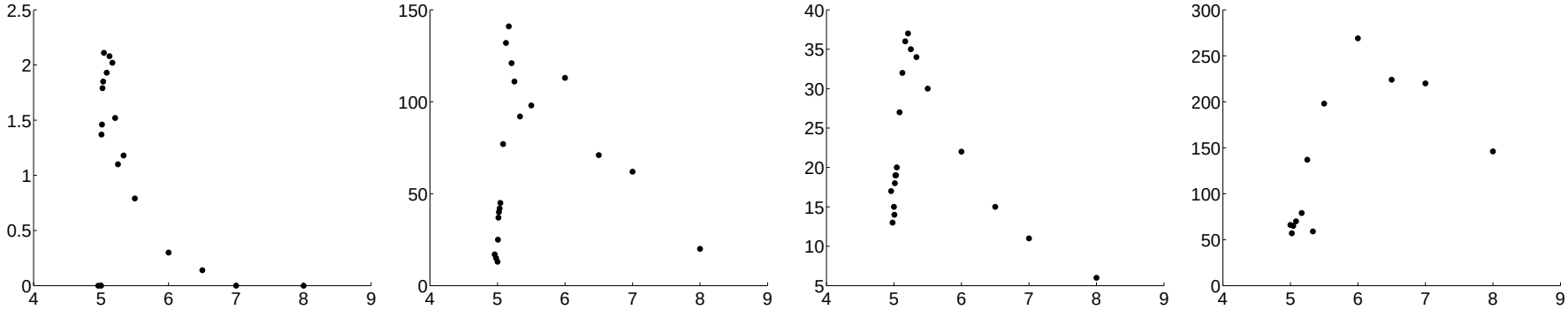
Treatment protocol data



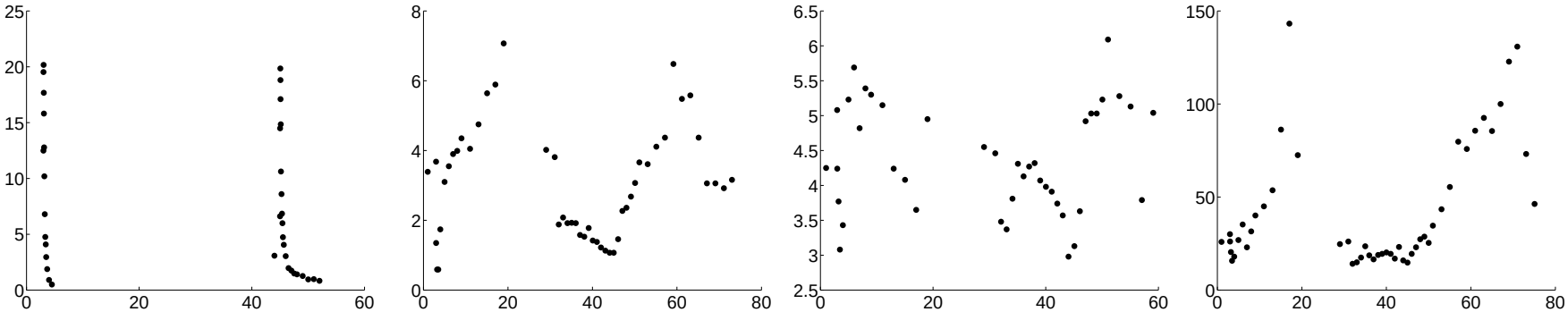
							Ovar rechts							Ovar links						
GnRHa	hMG/ FSH	Tag	Datum	BT	E2 pmol/L	P4 nmol/L	< 10	10 11	12 13	14 15	16 17	18 19	≥ 20	< 10	10 11	12 13	14 15	16 17	18 19	≥ 20
.																				
.																				
.																				
1	225	Fr	07.06.13	8	2841		4	1	1					5	1	1				
1	225	Sa	08.06.13	9																
1	225	So	09.06.13	10																
1	225	Mo	10.06.13	11	6062		2		1		1	1		2	3	5	3	3		

Drug data

Single dose nafarelin (GnRH agonist)

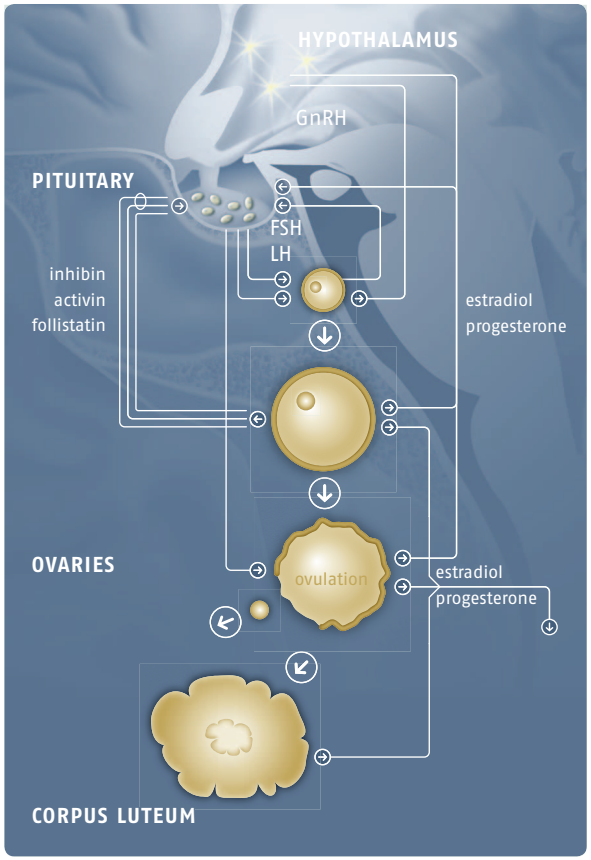


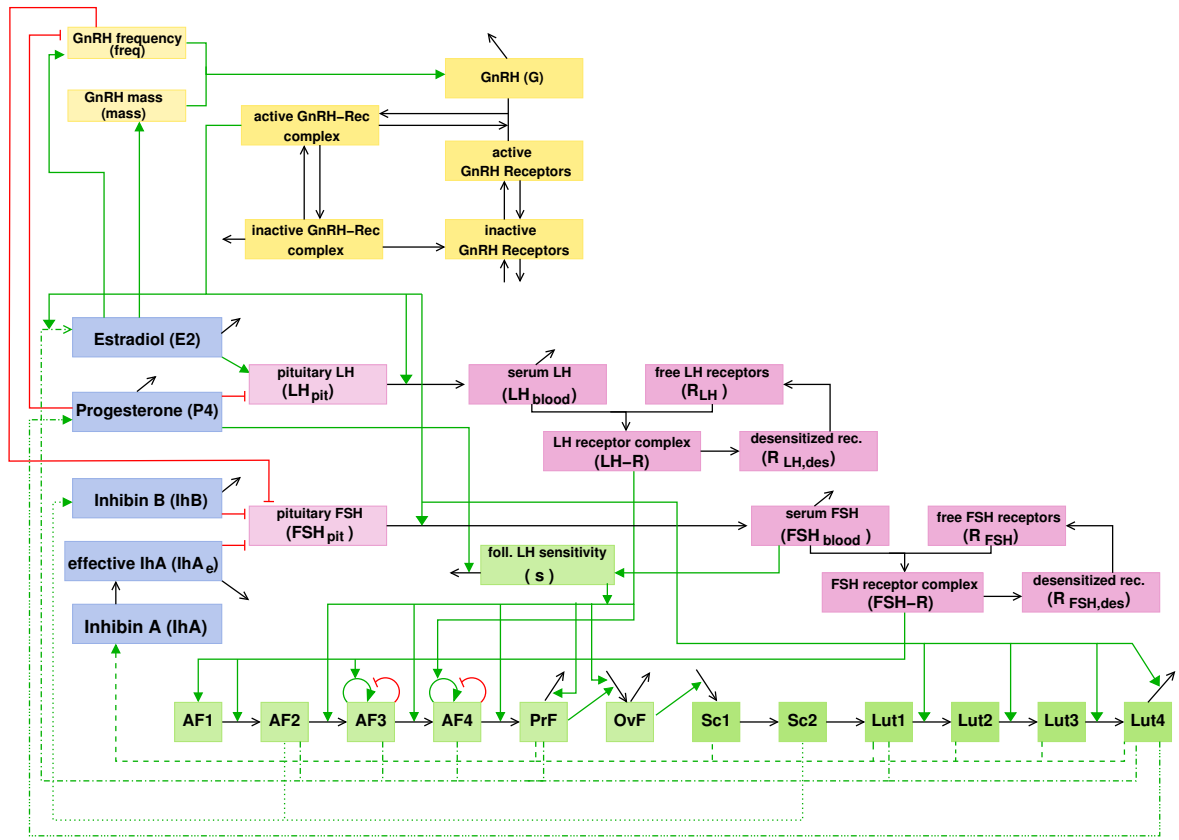
Single and multiple dose cetrorelix (GnRH antagonist)



measurements: drug, LH, FSH, E2

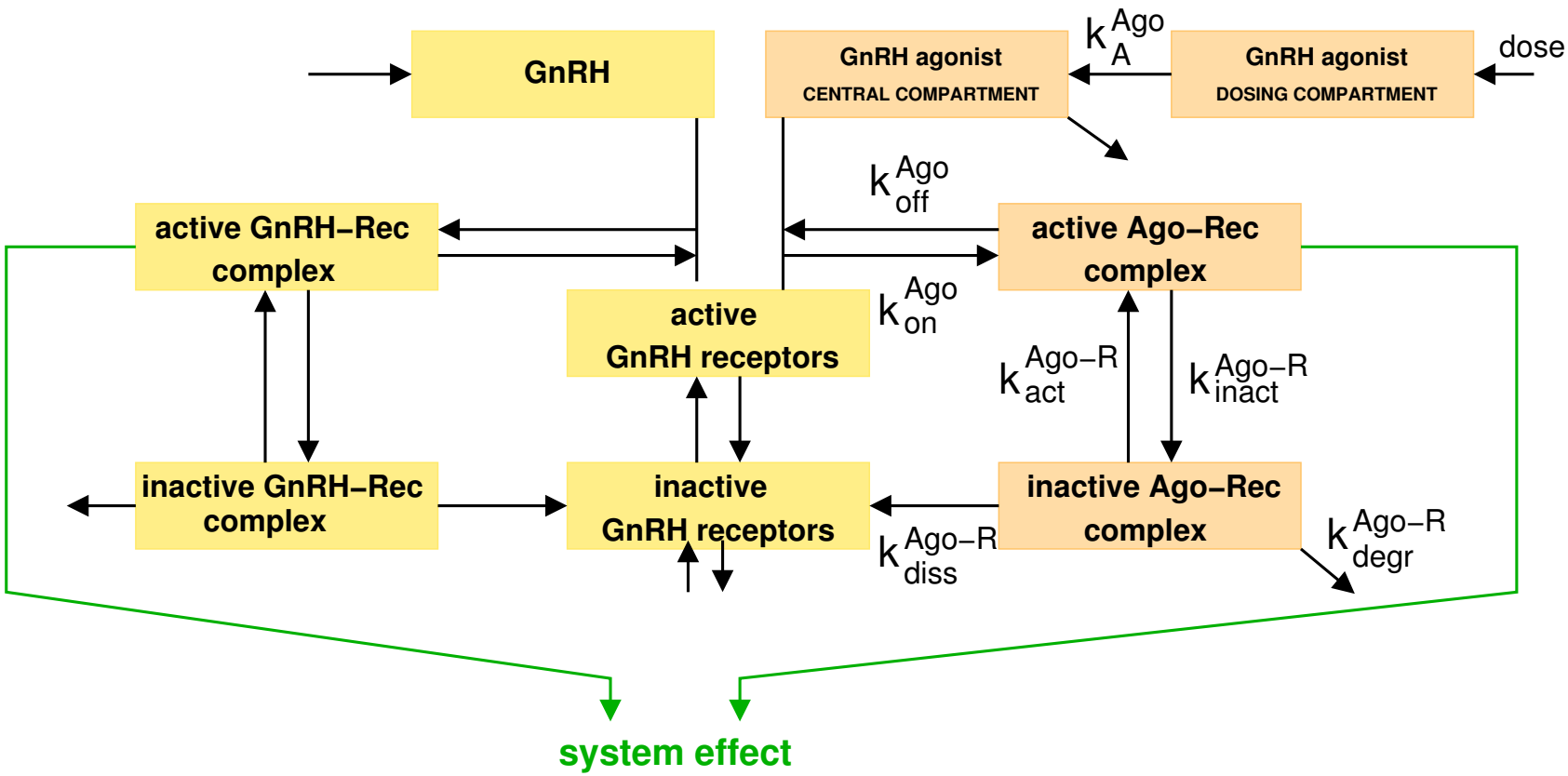
GynCycle





GynCycle: 33 ODEs, 114 parameters [Röblitz et al., J Theoret Biol 2013]
computation of hormone profiles and follicle development over time

PKPD Modelling: GnRH agonists

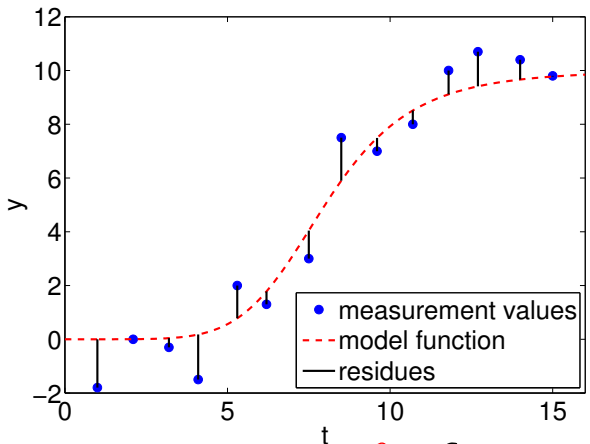


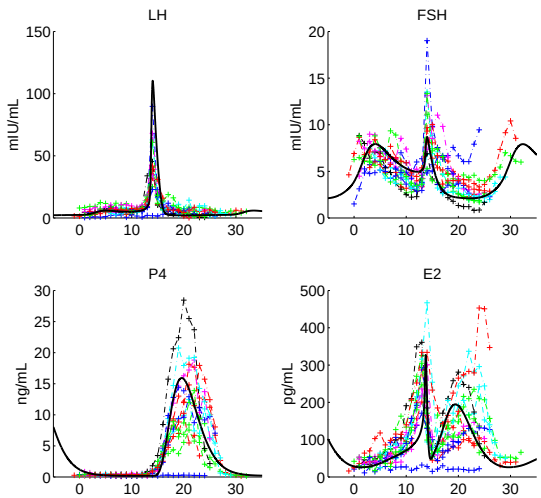
G protein-coupled receptor (GPCR) model coupled to PK model for GnRH agonists

Frequentists' approach

Assumption: there exists a fixed but unknown parametrization θ
$\rightarrow$ estimate θ (ill-posed!) and generate predictions:
least squares minimization by error-oriented Gauss-Newton method
NLSCON $\rightarrow$ restriction to subspace of identifiable parameters

$$\|F(\theta)\|_2^2 = \sum_{k=1}^{n} \sum_{l=1}^{m_k} \frac{(x_{kl} - y_k(t_l, \theta))^2}{2\sigma_{kl}^2} \xrightarrow{\theta} \min$$



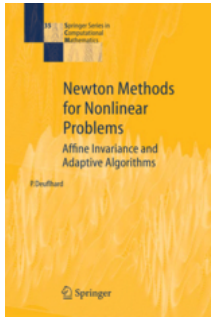


Single parametrization θ_0 from normal cycle and drug data

Gauss-Newton algorithm NLSCON

- solution of the nonlinear least squares problem by a global adaptive Gauss-Newton method

$$\|F'(\theta^{(k)})\Delta\theta^{(k)} + F(\theta^{(k)})\|^2 \to min$$
$$\theta^{(k+1)} = \theta^{(k)} + \lambda_k \Delta\theta^{(k)}, k = 0, 1, 2, \ldots$$



[Deuflhard: Newton Methods for Nonlinear Problems, 2004]

- sequence of linear least squares problems with Jacobian $F'(p)$

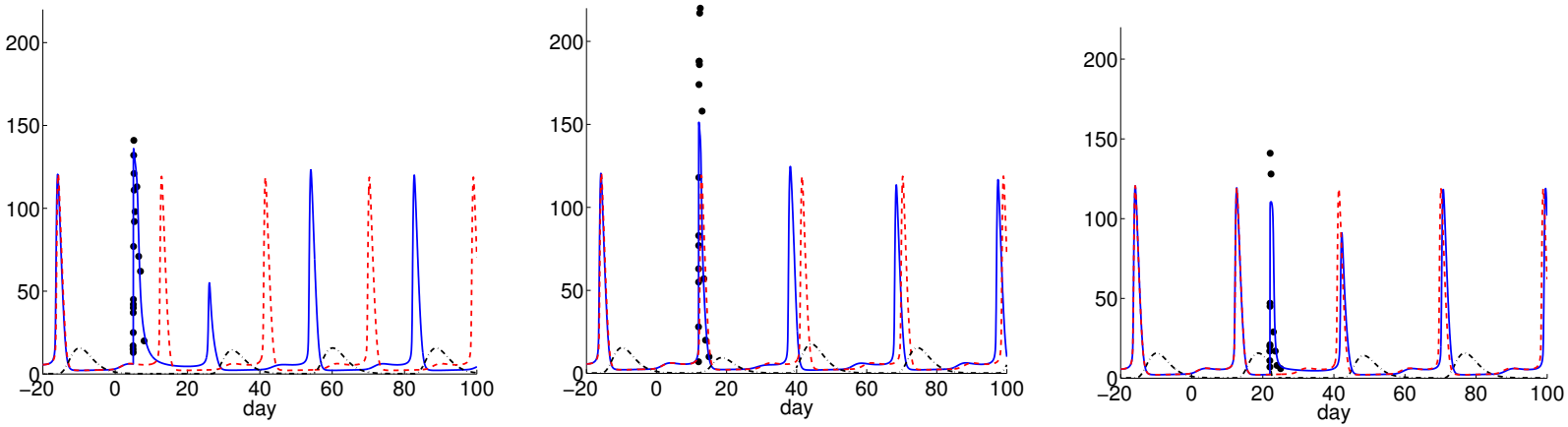
$$F'_{ij}(\theta) = \frac{\partial}{\partial\theta_j} y_{k_i}(t_i, \theta), \quad i = 1, \ldots, m, \quad k_i \in \{1, \ldots, n\}$$

- detection of linear dependencies by monitoring the subcondition numbers

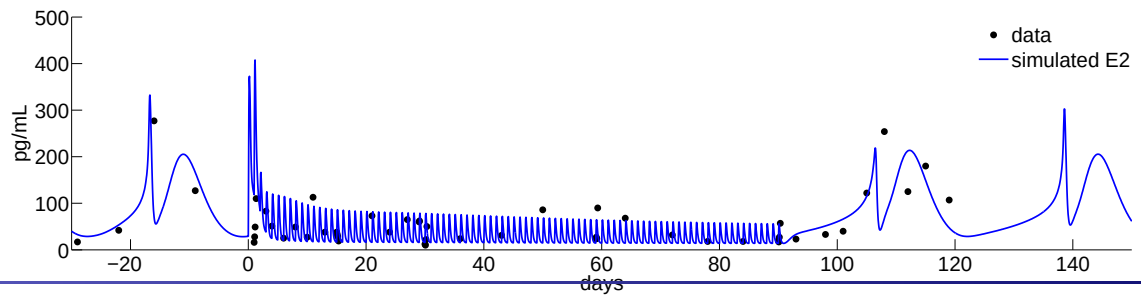
$$F'(\theta)\Pi = QR, \quad r_{11} \geq r_{22} \geq \ldots \geq r_{qq}, \quad \text{sc}_j = r_{11}/r_{jj} < 1/\varepsilon_\theta$$

Generic model

Study influence of **drug, dose and timing of administration** on hormonal profiles in a "normal" menstrual cycle [Röblitz et al. (2013)]

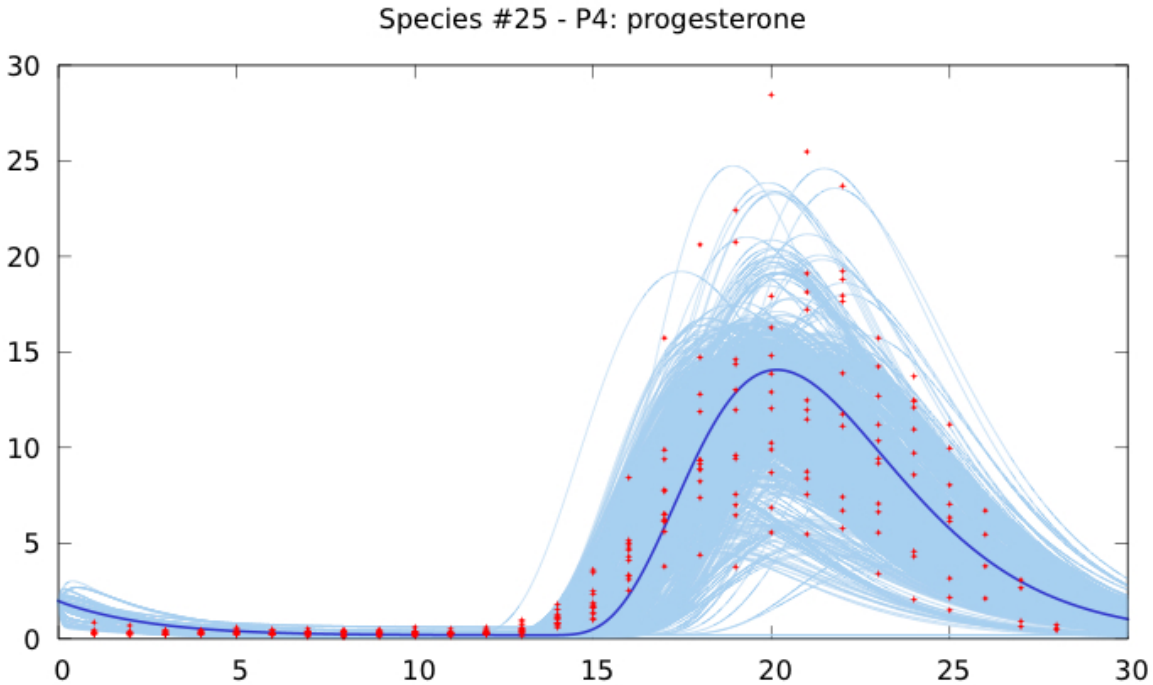


Study **long-time effect** of drug administration under different **compliance** behaviors



Virtual patients

Generate model instances (parametrizations) compatible with **real patient data** for the normal cycle [Mancini et al. (2014)].



Real Patient

(a) Medical level

Virtual patient

(b) Computation level

finite set of biologically admissable parameter sets

offline $\longrightarrow$ *online*

Bayesian Inference

- $\theta \in \Theta \subset \mathbb{R}^d$ unknown parameters
- $x \in \mathcal{X} \subset \mathbb{R}^n$ measurements
- likelihood model $\mathcal{M} = \{p(x|\theta),\, x \in \mathcal{X},\, \theta \in \Theta\}$

Bayesian Inference

- Input:
  - prior $\pi(\theta)$
  - measurement(s) $x \in \mathcal{X}$
  - likelihood $p(x \,|\, \theta)$
- Output:
  - posterior:
    $p(\theta \mid x) \propto p(x \mid \theta)\pi(\theta)$

Bayesian Inference

- $\theta \in \Theta \subset \mathbb{R}^d$ unknown parameters
- $x \in \mathcal{X} \subset \mathbb{R}^n$ measurements
- likelihood model $\mathcal{M} = \{p(x|\theta),\, x \in \mathcal{X},\, \theta \in \Theta\}$

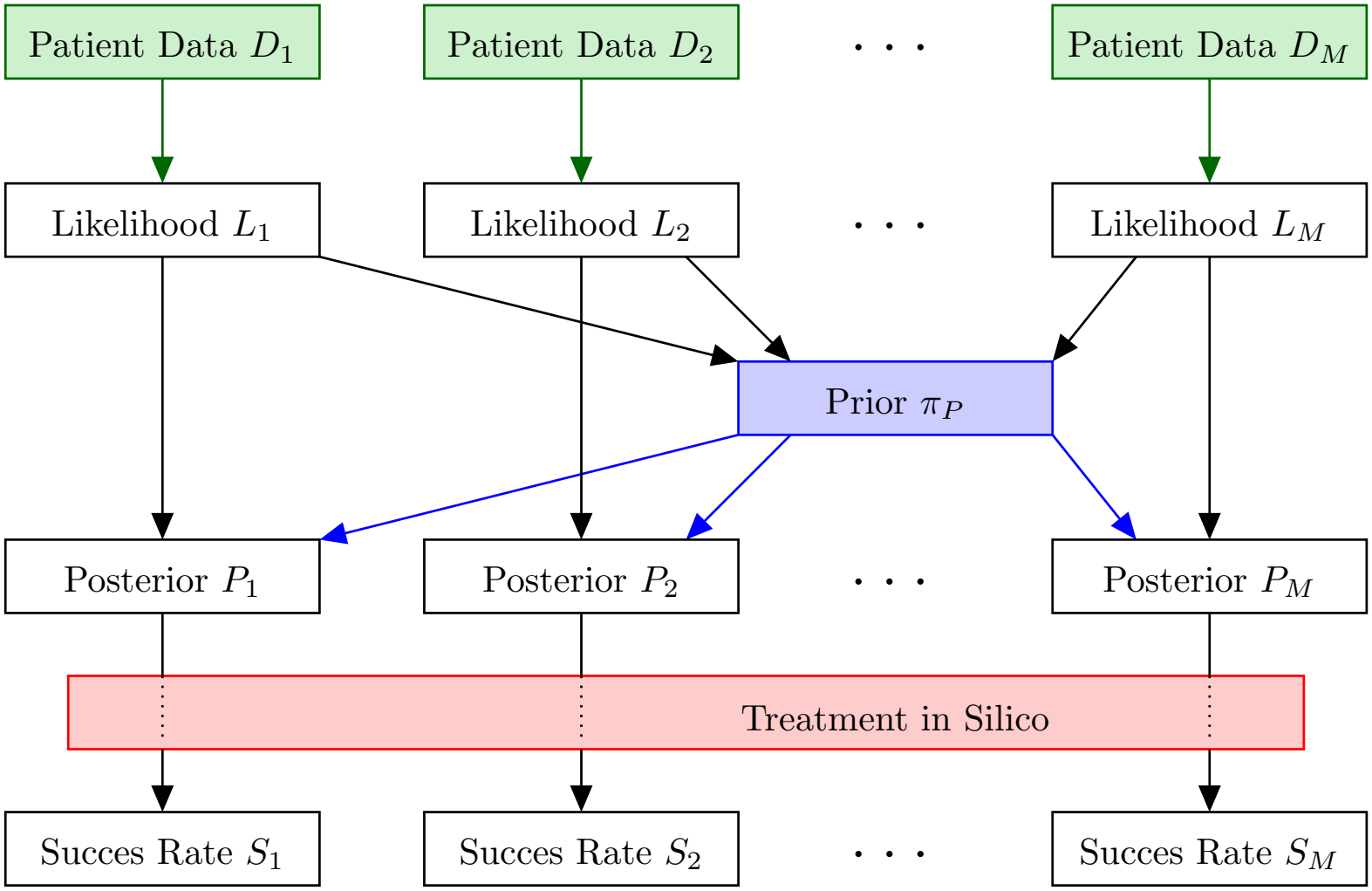
Bayesian Inference

- Input:
  - prior $\pi(\theta)$
  - measurement(s) $x \in \mathcal{X}$
  - likelihood $p(x \,|\, \theta)$
- Output:
  - posterior:
    $p(\theta \mid x) \propto p(x \mid \theta)\pi(\theta)$

Empirical Bayes Methods

- Input:
  - ~~prior $\pi(\theta)$~~
  - measurements $X = (x_1, \ldots, x_m)$ for **several** individuals with **individual** parameters $\theta_m \in \Theta$
  - likelihoods $p(x_m \,|\, \theta_m)$
- Output:
  - density estimate $\pi(\theta)$
  - individual posteriors:
    $p(\theta_m \mid x_m) \propto p(x_m \mid \theta_m)\pi(\theta_m)$

Empirical Bayes: Workflow



Empirical Bayes: Typical approach

- View the unknown prior as a hyperparameter π
- Apply a parameter estimation technique to infer it from its likelihood for the measurement $X = (x_1, \ldots, x_M)$, e.g.

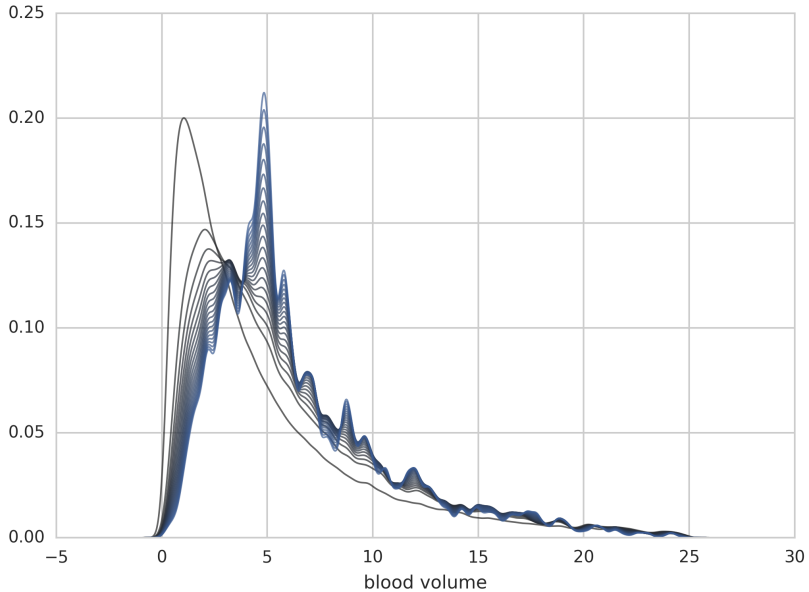
$$\pi_{\mathrm{NPMLE}} = \underset{\pi}{\operatorname{argmax}} \log L(\pi), \quad L(\pi) = \prod_{m=1}^{M} p(x_m \mid \pi), \; p(x_m \mid \pi) = \int p(x_m \mid \theta)\pi(\theta)d\theta$$

Empirical Bayes: Typical approach

- View the unknown prior as a hyperparameter π
- Apply a parameter estimation technique to infer it from its likelihood for the measurement $X = (x_1, \ldots, x_M)$, e.g.

$$\pi_{\mathrm{NPMLE}} = \operatorname*{argmax}_{\pi} \log L(\pi), \quad L(\pi) = \prod_{m=1}^{M} p(x_m \mid \pi), \; p(x_m \mid \pi) = \int p(x_m \mid \theta)\pi(\theta)d\theta$$

- π_{NPMLE} turns out to be a discrete distribution with at most M nodes



Empirical Bayes: Typical approach

- View the unknown prior as a hyperparameter π
- Apply a parameter estimation technique to infer it from its likelihood for the measurement $X = (x_1, \ldots, x_M)$, e.g.

$$\pi_{\mathrm{NPMLE}} = \operatorname*{argmax}_{\pi} \log L(\pi), \quad L(\pi) = \prod_{m=1}^{M} p(x_m \mid \pi), \; p(x_m \mid \pi) = \int p(x_m \mid \theta)\pi(\theta)d\theta$$

- π_{NPMLE} turns out to be a discrete distribution with at most M nodes
- **Regularization** of π_{NPMLE} using a penalty $\Phi(\pi)$:

$$\pi_{\mathrm{MPLE}} = \operatorname*{argmax}_{\pi} \log L(\pi) - \gamma\Phi(\pi)$$

$\gamma > 0$ balances the trade-off between goodness of fit and smoothness or non-informativity of the prior

Empirical Bayes: Typical approach

- View the unknown prior as a hyperparameter π
- Apply a parameter estimation technique to infer it from its likelihood for the measurement $X = (x_1, \ldots, x_M)$, e.g.

$$\pi_{\mathrm{NPMLE}} = \operatorname*{argmax}_{\pi} \log L(\pi), \quad L(\pi) = \prod_{m=1}^{M} p(x_m \mid \pi), \; p(x_m \mid \pi) = \int p(x_m \mid \theta)\pi(\theta)d\theta$$

- π_{NPMLE} turns out to be a discrete distribution with at most M nodes
- **Regularization** of π_{NPMLE} using a penalty $\Phi(\pi)$:

$$\pi_{\mathrm{MPLE}} = \operatorname*{argmax}_{\pi} \log L(\pi) - \gamma\Phi(\pi)$$

  $\gamma > 0$ balances the trade-off between goodness of fit and smoothness or non-informativity of the prior
- Most penalties in use are variant under transformations of X
  $\Longrightarrow$ Results depend on the choice of the parametrization!

Consistent statistical inference

If equivalent models are used to explain equivalent data, the results of statistical inference must be consistent!

Consistent statistical inference

If equivalent models are used to explain equivalent data, the results of statistical inference must be consistent!

$$-\Phi_{\mathcal{I}} = \mathcal{I}[\pi \mid \mathcal{M}] = \int_{\Theta} \int_{\mathcal{X}} \pi(\theta) p(x \mid \theta) \log \left(\frac{p(x \mid \theta)}{p(x \mid \pi)} \right) dx d\theta$$

The mutual information (expected information gained from one observation of the model $\mathcal{M}$ on a parameter θ with prior $\pi(\theta)$) is invariant under transformations of x and θ and concave in π.

Consistent statistical inference

If equivalent models are used to explain equivalent data, the results of statistical inference must be consistent!

$$-\Phi_{\mathcal{I}} = \mathcal{I}[\pi \mid \mathcal{M}] = \int_{\Theta}\int_{\mathcal{X}} \pi(\theta)p(x \mid \theta)\log\left(\frac{p(x \mid \theta)}{p(x \mid \pi)}\right) dxd\theta$$

The mutual information (expected information gained from one observation of the model $\mathcal{M}$ on a parameter θ with prior $\pi(\theta)$) is invariant under transformations of x and θ and concave in π.

- If the Bayesian inverse problem takes the form $x = \varphi(\theta) + E$, $\Phi_{\mathcal{I}}$ is equivalent to the entropy $\mathcal{H}_X(\pi)$ in measurement space,

$$\mathcal{H}_X(\pi) = -\int p(x \mid \pi)\,\log p(x \mid \pi)\,\mathrm{d}x,$$

  and thereby has a natural interpretation (non-informativity)

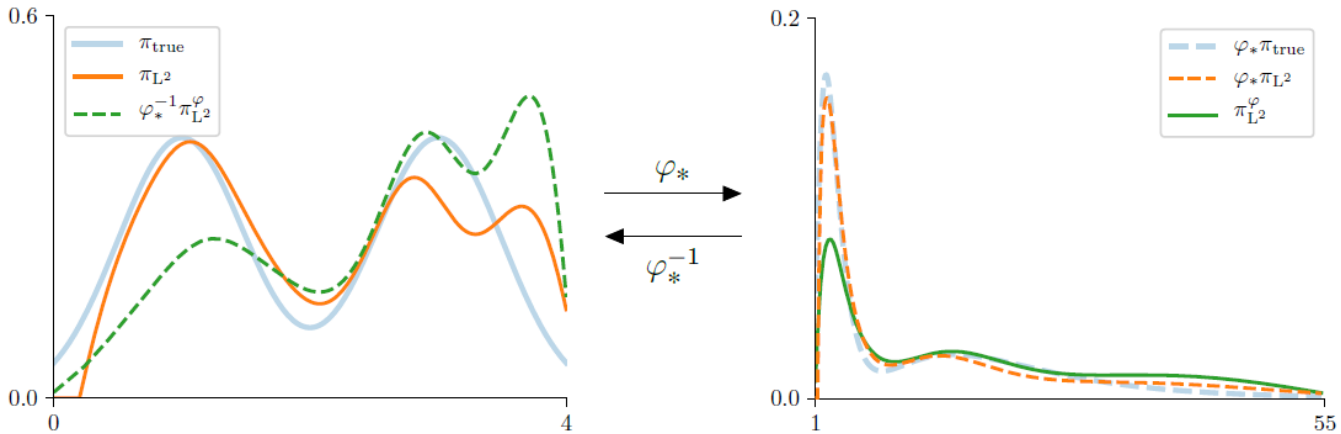
  [Klebanov, Sikorski, Schütte, Röblitz: Objective Priors in the Empirical Bayes Framework, `https://arxiv.org/abs/1612.00064`]

Transformation invariance

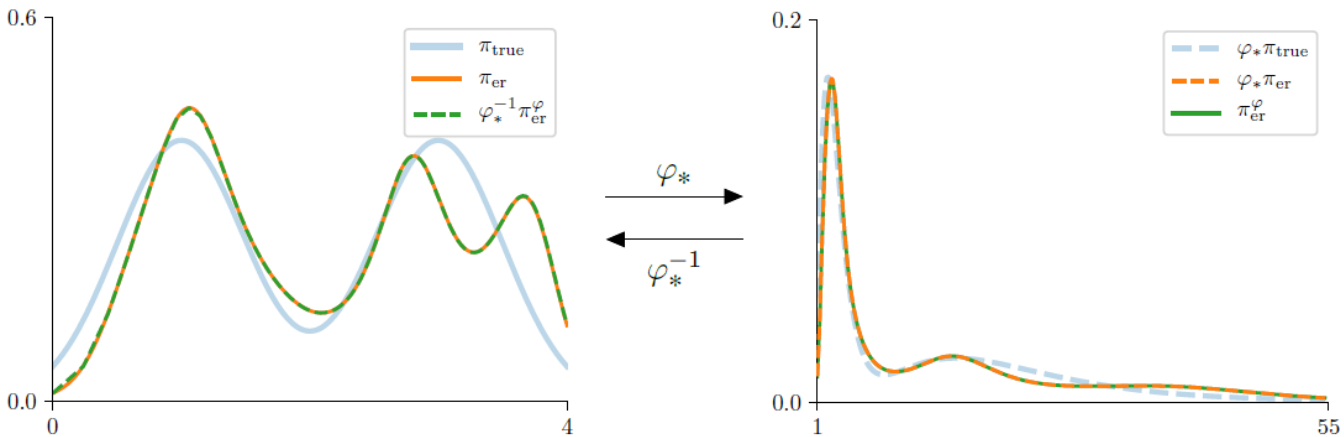
$x|\theta \sim \mathcal{N}(\theta, 0.3^2)$, $\theta \sim \pi_{\text{true}} = 0.5 \cdot (\mathcal{N}(1, 0.5^2) + \mathcal{N}(3, 0.5^2))|_{[0,4]}$, $\varphi : \theta \mapsto \tilde{\theta} = \exp(\theta)$

$M = 100$, MMA algorithm for gradient-based local optimization

Thikonov:



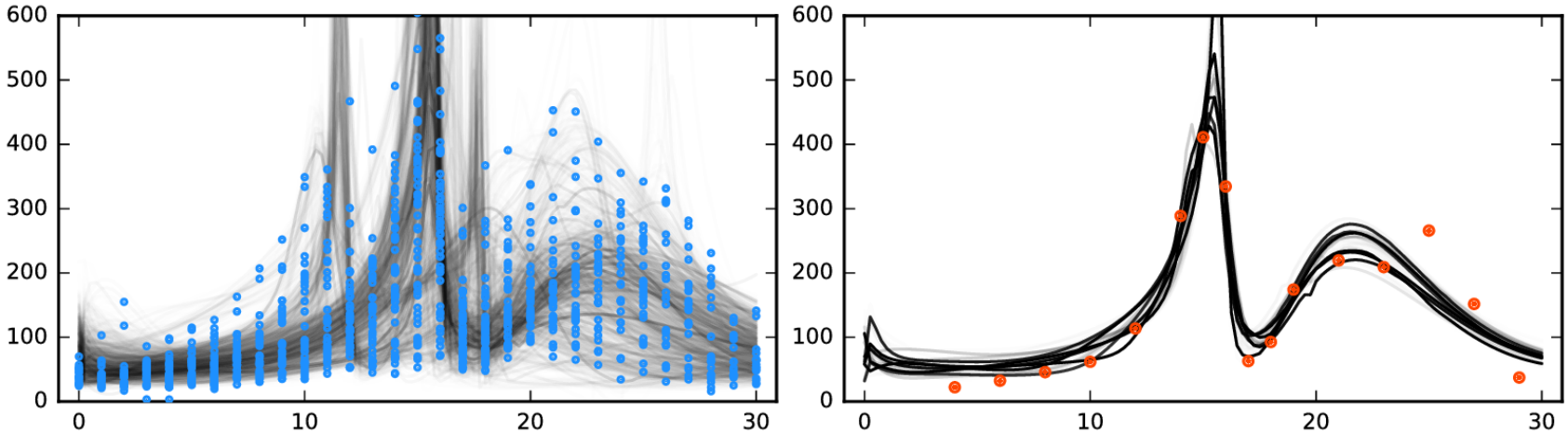
information penalty:



Parameter inference for GynCycle

Data: 53 healthy women, 4 hormones, measured roughly every second day over 30 days

Inference: 82 parameters, 33 initial values



GynCycle: sample trajectories from estimated prior and one individual posterior

[Klebanov et al., ZR 16-56, 2016]

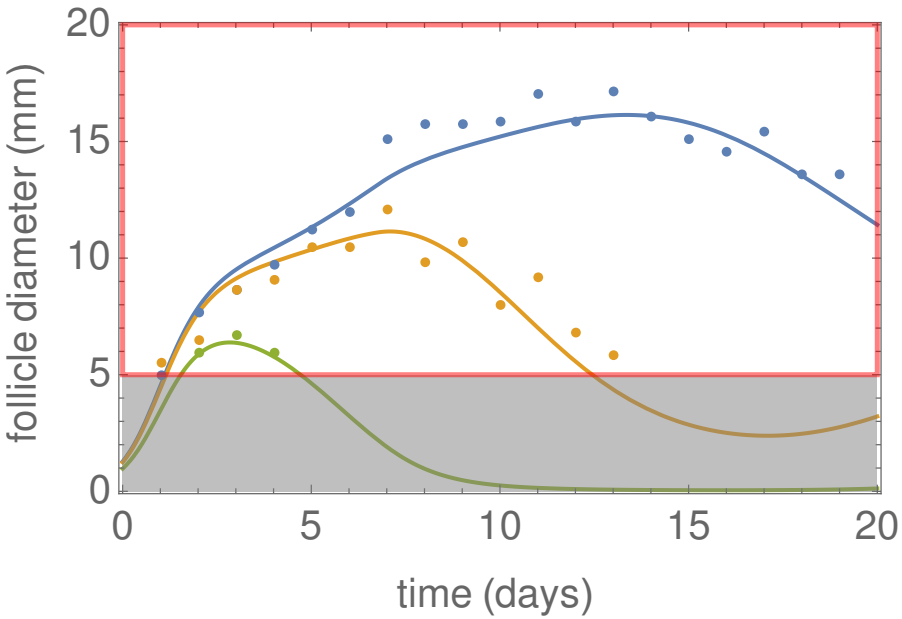
Julia implementation: `https://github.com/axsk/GynC.jl`

A new model for follicular development

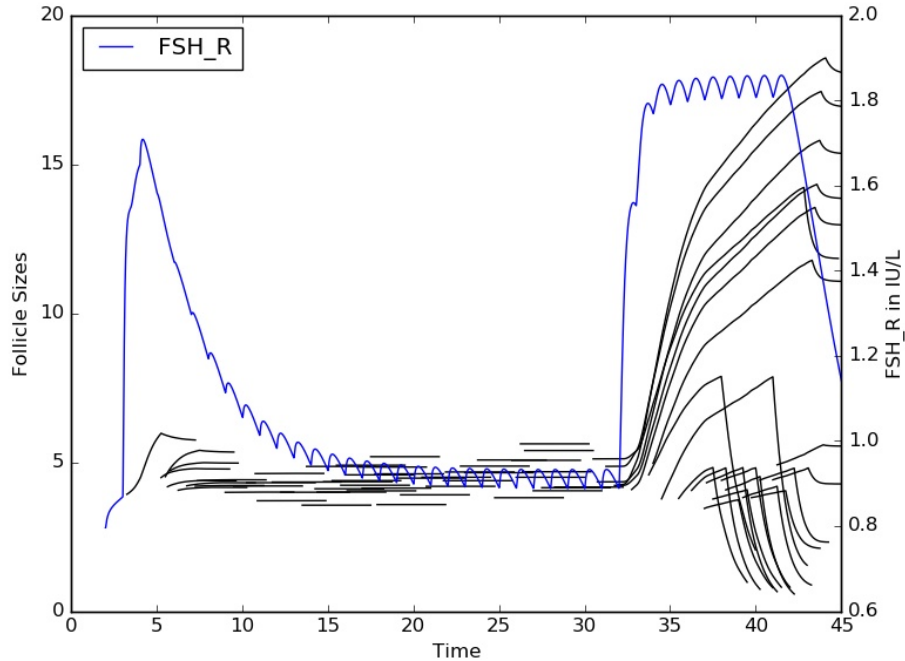
$$x_i' = H^+(FSH, \delta_i, \eta) x_i (\xi - x_i)(\gamma - \kappa(\Sigma x - x_i))$$

$\delta_i \sim \mathcal{N}(\sigma, \mu)$, $x_i(t_{i0}) \sim \mathcal{N}(5mm, 1.5mm)$, Poisson process for t_{i0}

Fit to ultrasound data:



Behavior under treatment:



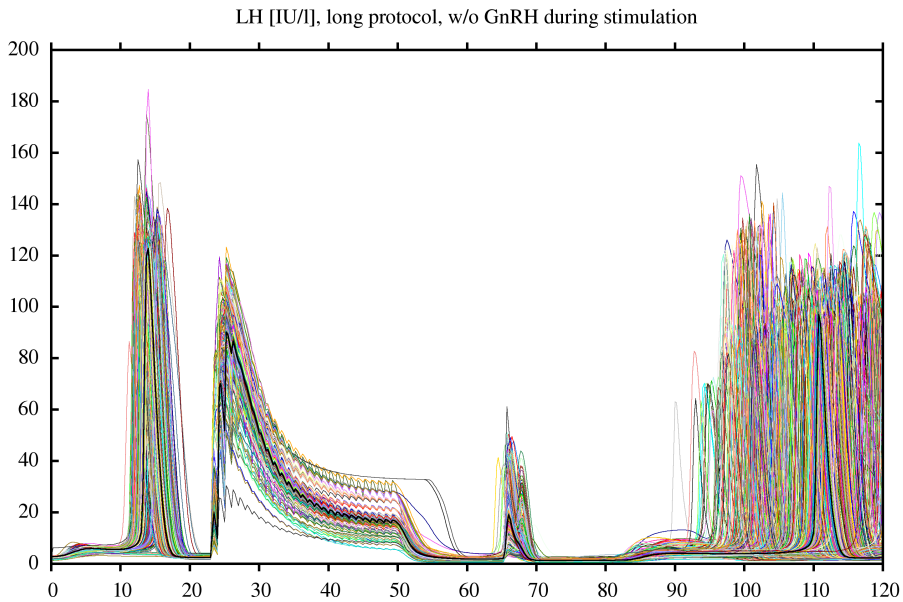
Number of dominant follicles: $d = 1 + \left[\frac{\gamma}{\kappa\xi}\right]$

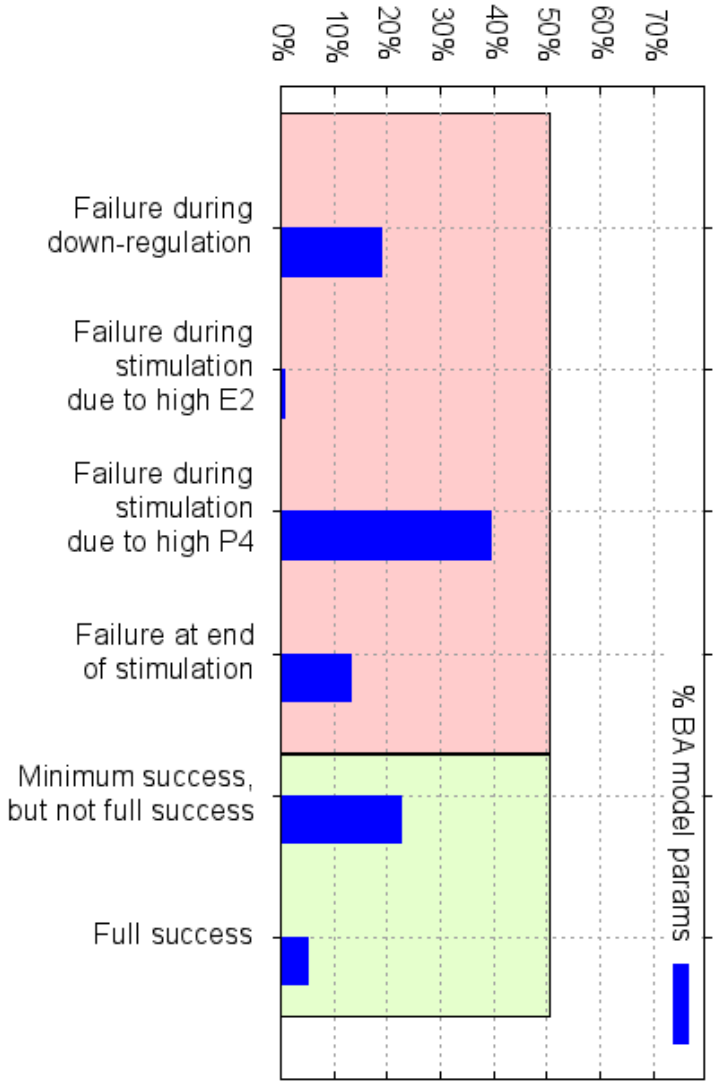
[Lange et al., J Math Biol (2018)]

Model-based treatment verification

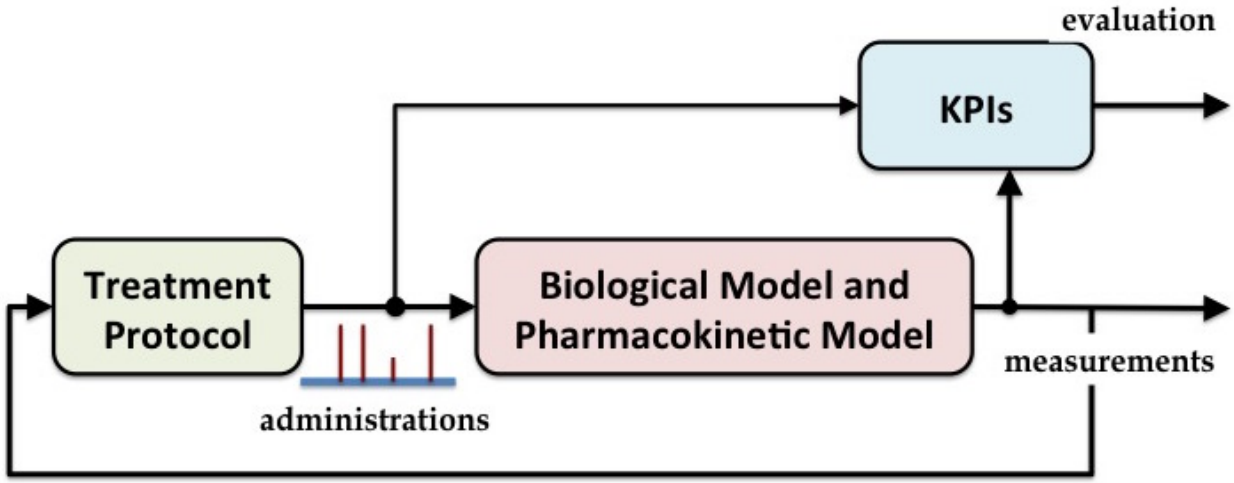
Verify that a given treatment protocol reaches its goal for the largest possible number of (virtual) patients

$\rightarrow$ evaluate treatment success rate





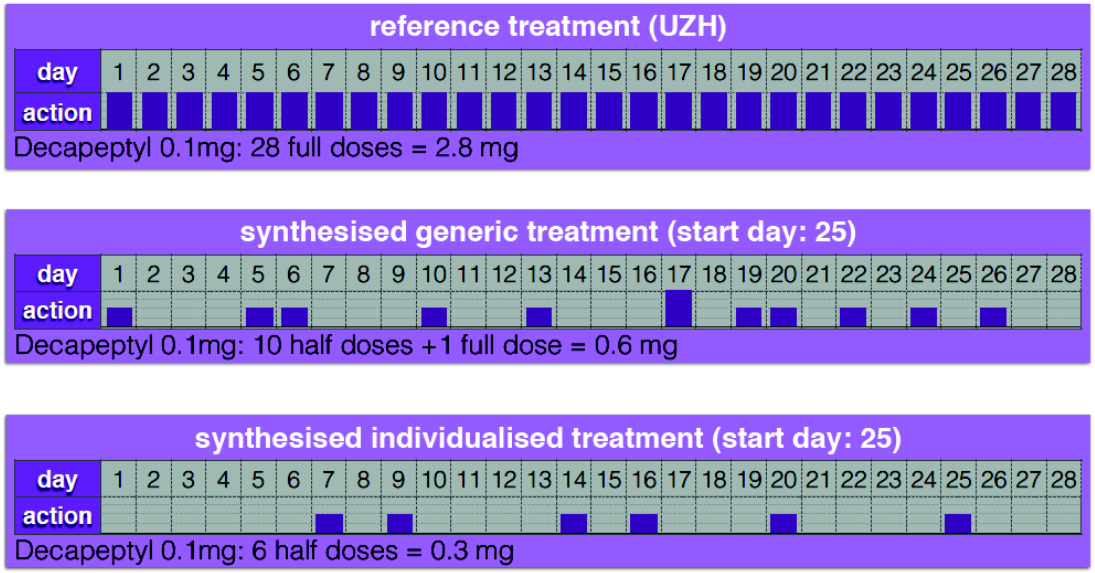
Model-based treatment optimization & design

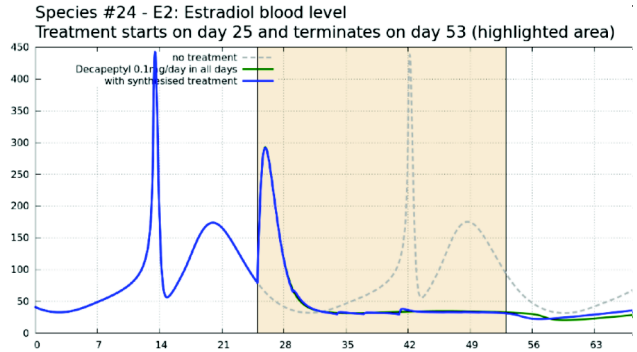


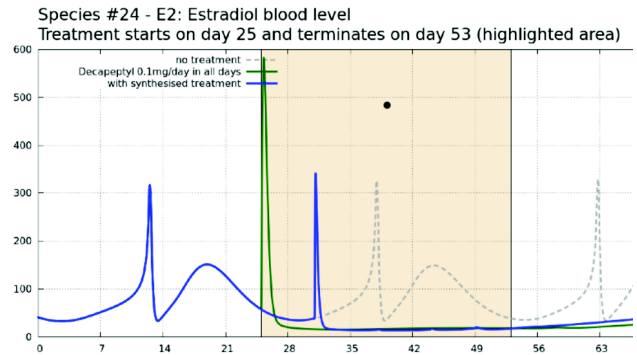
Treatment optimization: finding values for treatment parameters (type, dose and time of drug) that optimize some KPIs

- number and sizes of dominant follicles (efficacy)
- total amount of drug used (costs)
- range of hormone concentrations, e.g. E2 (safety)

Model-based treatment optimization







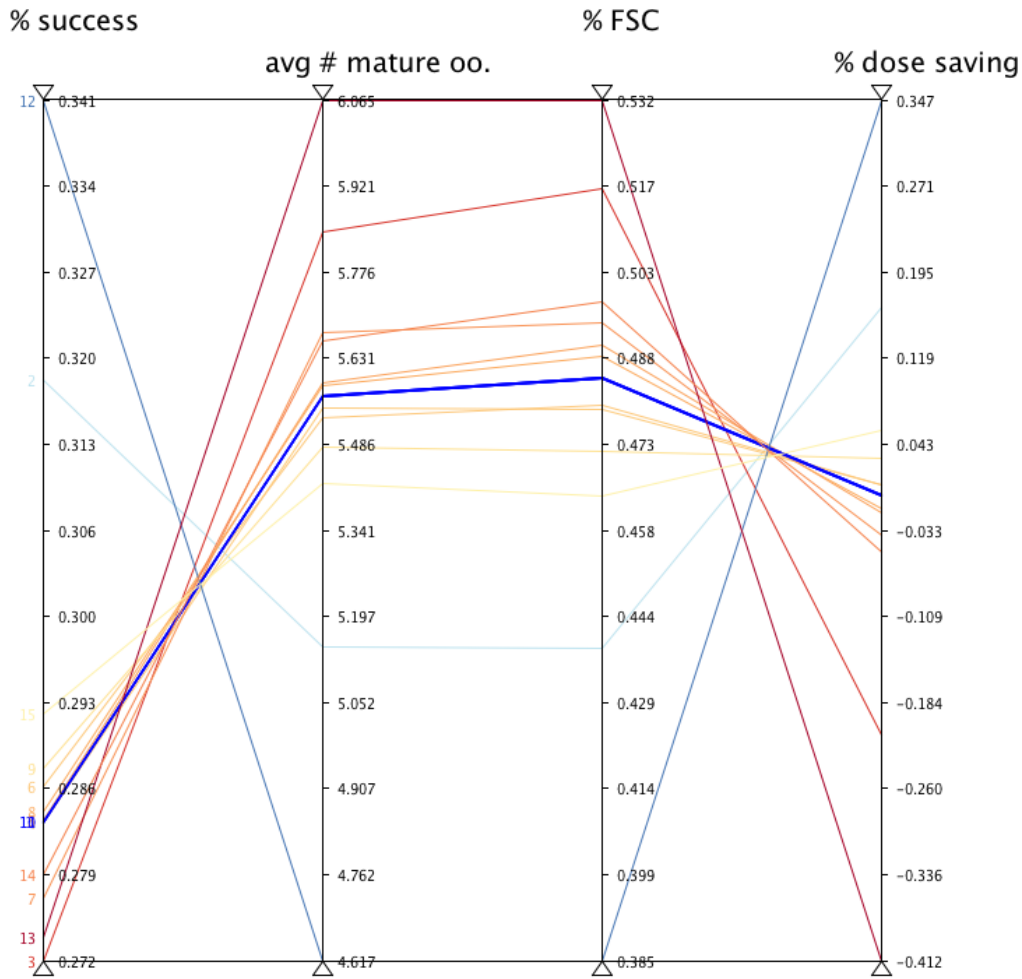
Synthesised generic down-regulation treatments require 40% of the injections and $<25\%$ of the overall Decapeptyl amount required by reference treatment. Individualised treatments even lighter, still achieving clinical goals!

Model-based treatment design

incremental change of treatment parameters:

- age class
- AMH level
- AFC class
- dose of stimulation drug

$\rightarrow$ set of Pareto-optimal treatments, in which at least one performance indicator is better



Conclusion

- We constructed a mechanistic model of the human menstrual cycle.
- Randomness in the follicle model introduces intra-individual variability.
- The empirical Bayes approach allowed us to construct a virtual patient population that displays inter-individual variability.
- The virtual patient population enables treatment verification, optimization, and design *in silico*.

Conclusion

- We constructed a mechanistic model of the human menstrual cycle.
- Randomness in the follicle model introduces intra-individual variability.
- The empirical Bayes approach allowed us to construct a virtual patient population that displays inter-individual variability.
- The virtual patient population enables treatment verification, optimization, and design *in silico*.

Future work:

- Patient-specific parametrization of the follicle model
- Patient-specific treatment planning
- Integration of data and algorithms into a "virtual hospital"

Acknowledgement

Computational Systems Biology Group

http://www.zib.de/numeric/csb

http://www.cbu.uib.no/roblitz-group

in particular:

R. Ehrig, I. Klebanov, A. Sikorski, A. Lange

Collaboration partners

Enrico Tronci (La Sapienza Rome)
Brigitte Leeners (University Hospital Zurich)
Tillmann Krüger (Hannover Medical School)
Marcel Egli (University Lucerne)

http://paeon.di.uniroma1.it

Funding

Research Center Matheon
Mathematics for Key Technologies

ECMath
Einstein Center for Mathematics Berlin

Bundesministerium für Bildung und Forschung

SEVENTH FRAMEWORK PROGRAMME