Bioinformatics OXFORD

Complete populations of virtual patients for *in silico* clinical trials

Supplementary material

S. Sinisi, V. Alimguzhin, T. Mancini, E. Tronci, B. Leeners

Abstract

This supplementary material is organised as follows. In Section SM1 we give the formal definitions and full details of our methodology. In Section SM2, we show technical details on our case study and on how we instantiated our general methodology to it. In Section SM3, we give formal details about how we conducted our quantitative evaluation of the representativeness of our computed population with respect to the spectrum of behaviours occurring in our datasets. Finally, Section SM4 lists all the acronyms used in the paper.

SM1 Methods

SM1.1 Formal framework

We denote with $\mathbb{R}$, $\mathbb{R}_{0+}$, and $\mathbb{R}_{+}$ the sets of, respectively, all, non-negative, and positive reals, and with $\mathbb{N}$ and $\mathbb{N}_{+}$ the sets of, respectively, non-negative and strictly positive integers. Also, given sets A and B, we denote with A^B the set of functions from B to A.

SM1.1.1 VPH models

As stated in the main paper, we adopt a very general approach to define Virtual Physiological Human (VPH) models and view them as parametric input-output dynamical systems, *i.e.*, parametric functions mapping input time functions (denoting external inputs such as drug administrations) to output time functions (defining the time evolution of observable quantities of interest).

This general definition (Definition 1) is standard in signals and systems (see, *e.g.*, 2), especially when, as in the case of physiological models, the system internal state is not accessible, and only selected outputs (*system observables*) can be measured.

For physical reasons, we also require that our models are *strictly causal*, *i.e.*, their behaviour up to any time instant depends only on *past* inputs. Also, given the presence of parameters, we focus on deterministic systems, in that parameters embody any initial condition which the system output might depend on.

Finally, our definition captures both *continuous-time* models (as, *e.g.*, those defined by means of Ordinary Differential Equations, ODEs) and *discrete-time* models (as, *e.g.*, those defined by means of difference equations).

Definition 1. *A* parametric deterministic input-output dynamical system $\mathcal{S}$ *is defined by a tuple* $(\mathcal{T}, \Lambda, \mathcal{U}, \mathcal{Y}, \mathbf{y})$*, where:* $\mathcal{T} \in \{\mathbb{R}_{0+}, \mathbb{N}\}$ *(or a bounded interval thereof) is the* time-set *and* Λ*,* $\mathcal{U}$ *and* $\mathcal{Y} \neq \emptyset$ *are, respectively, the* parameter*,* input*, and non-empty* output spaces*.*

Function $\mathbf{y}$ *is the* observation function *of* $\mathcal{S}$*: for any* $\mathbf{u} \in \mathcal{U}^{\mathcal{T}}$ *and* $\lambda \in \Lambda$*,* $\mathbf{y}(\mathbf{u}, \lambda) \in \mathcal{Y}^{\mathcal{T}}$ *is a function defining the* time evolution *of the system observables at each time point in* $\mathcal{T}$*, when the system parameters are set to* λ *and the system is fed with input function* $\mathbf{u}$*. For any* $t \in \mathcal{T}$*, we denote with* $\mathbf{y}(t; \mathbf{u}, \lambda)$ *the value of* $\mathbf{y}(\mathbf{u}, \lambda)$ *at time point* t*.*

System $\mathcal{S}$ *is* strictly causal *if, for any* $\lambda \in \Lambda$*,* $t \in \mathcal{T}$*, and any pair* $\mathbf{u}_1, \mathbf{u}_2 \in \mathcal{U}^{\mathcal{T}}$ *such that* $\mathbf{u}_1(t') = \mathbf{u}_2(t')$ *for all* $t' < t$*, it holds* $\mathbf{y}(t; \mathbf{u}_1, \lambda) = \mathbf{y}(t; \mathbf{u}_2, \lambda)$*.*

Strict causality implies that the system initial output, $\mathbf{y}(0; \mathbf{u}, \lambda)$, depends on λ, but not on $\mathbf{u}$, *i.e.*, that λ also embodies information about any initial condition of the system (and, in turn, the output at time 0).

SM1.1.2 Virtual patients, phenotypes, populations

Not all assignments to a VPH model parameters yield behaviours of interest. Many might even yield physiologically meaningless behaviours. Conversely, due to, *e.g.*, system over-parametrisation or non-identifiability, multiple parameter assignments may yield (almost) indistinguishable behaviours (*i.e.*, their associated evolutions are very similar on all inputs). As explained in the main paper, such indistinguishable Virtual Patients (VPs) would increase the computational efforts needed to carry out an *In Silico* Clinical Trial (ISCT) on the entire population, without bringing any advantage in terms of representativeness of the trial.

Forthcoming Definition 2 formalises the concepts of VP, population of VPs, phenotype, and All-Different Phenotype Population (APP) for a given VPH model. As introduced in Section 2.1 of the main paper, these concepts rest on user-provided Boolean function φ and equivalence relation $\sim$. Boolean function φ defines the conditions to be met by any VPH model parameter λ for its trajectories to be considered of interest, for example *physiologically meaningful* (hence λ has to be regarded as a VP). Equivalence relation $\sim$ on the set of VPs defines when two VPs shall

be considered having indistinguishable behaviour (*i.e.*, showing the same *phenotype*): for any two VPs λ and λ', $\lambda \sim \lambda'$ means that the two VPs show the same phenotype, which is represented by their equivalence class $[\lambda]_\sim = [\lambda']_\sim$.

We recall that the definition of both function φ and relation $\sim$ will depend on the VPH model at hand, and has to be made starting from expert knowledge. Also, when the model is subject to external inputs (*e.g.*, drug administrations), both φ and $\sim$ might need to be defined on model evolutions under *different input functions*.

Definition 2 (Virtual Patients, Phenotypes, Populations). *Let $\mathcal{S}$ be a VPH model (as in Definition 1) with parameter space Λ.*

Given a function φ of the form $\varphi : \Lambda \to \{0,1\}$, the population *$\hat{\Lambda}$ of Virtual Patients (VPs) associated to $\mathcal{S}$ with respect to φ is: $\hat{\Lambda} = \{\lambda \mid \lambda \in \Lambda \wedge \varphi(\lambda) = 1\}$.*

Given an equivalence relation $\sim$ over $\hat{\Lambda}$, the phenotype space *$\hat{\Lambda}/\sim$ of $\hat{\Lambda}$ with respect to $\sim$ is the quotient set of $\hat{\Lambda}$ with respect to $\sim$: $\hat{\Lambda}/\sim = \{[\lambda]_\sim \mid \lambda \in \hat{\Lambda}\}$, where $[\lambda]_\sim$ is the equivalence class of λ. Two VPs $\lambda, \lambda' \in \hat{\Lambda}$ such that $\lambda \sim \lambda'$ are said to have the* same phenotype *$[\lambda]_\sim = [\lambda']_\sim$.*

An All-Different Phenotype Population (APP) of VPs *with respect to $\sim$ is any subset $\hat{\Lambda}^\sim$ of $\hat{\Lambda}$ such that no two VPs λ, λ' exist in $\hat{\Lambda}^\sim$ such that $[\lambda]_\sim = [\lambda']_\sim$. An APP $\hat{\Lambda}^\sim$ is said a* Complete APP (CAPP) *if $\hat{\Lambda}^\sim/\sim$ equals $\hat{\Lambda}/\sim$, i.e., $\hat{\Lambda}^\sim$ contains a representative of* all phenotypes *in the phenotype space of $\hat{\Lambda}$ with respect to $\sim$.*

Note that, when $\sim$ is $\mathbf{1}$ (*i.e.*, the equivalence relation defining a distinct class per VP $\lambda \in \hat{\Lambda}$), we have $\hat{\Lambda}/\mathbf{1} = \hat{\Lambda}$. Hence, the entire population of VPs ($\hat{\Lambda}$) can always be regarded as a CAPP.

SM1.2 Computing complete populations

SM1.2.1 The algorithm

Here we state our formal result about the statistical guarantees provided by our algorithm.

Theorem 1. *Let $\mathcal{S}$ be a VPH model as in Definition 1 having finite parameter space Λ, φ and $\hat{\Lambda}$ be as in Definition 2 (with $\hat{\Lambda}$* unknown*), $\sim_1, \ldots, \sim_L$ be $L > 0$ equivalence relations on $\hat{\Lambda}$, and $\delta \in (0,1)$. Also, let $\Lambda_1, \ldots, \Lambda_k$ be a splitting of Λ into $k > 0$ slices.*

For all $i \in [1,k]$, $l \in [1,L]$, whenever the i-th branch of the algorithm in Figure 1 of the main paper (working on Λ_i), outputs $(\sim_l, \hat{\Lambda}_i^{\sim_l}, \varepsilon_l)$, then, with probability $\geq (1-\delta)$, value ε_l is an upper bound to the probability that further sampling will disprove that $\hat{\Lambda}_i^{\sim_l}$ is a CAPP of Λ_i with respect to $\sim_l$.

Proof. Assume that the branch of the algorithm of Figure 1 (main paper) working on slice Λ_i ($i \in [1,k]$) outputs tuple $(\sim_l, \hat{\Lambda}_i^{\sim_l}, \varepsilon_l)$ with $l \in [1,L]$ and $\varepsilon_l < 1$ (in case $\varepsilon_l = 1$ the thesis trivially follows).

From the algorithm of Figure 1, ε_l is $1 - \delta^{1/N_l}$, where N_l is the number of the last consecutively drawn random samples that did not result in a VP with a phenotype not represented in the current APP $\hat{\Lambda}_i^{\sim_l}$.

Let $H_0(\hat{\Lambda}_i^{\sim_l}; \varepsilon_l)$ be the following *null hypothesis*: "the probability to disprove, by further sampling, that $\hat{\Lambda}_i^{\sim_l}$ is a CAPP is $\geq \varepsilon_l$".

To prove the thesis, it suffices to show that the probability that the algorithm rejects $H_0(\hat{\Lambda}_i^{\sim_l}; \varepsilon_l)$ when it actually holds (*type-I error*) is at most δ.

Indeed, if $H_0(\hat{\Lambda}_i^{\sim_l}; \varepsilon_l)$ holds, the probability to sample a VP of a phenotype not represented in $\hat{\Lambda}_i^{\sim_l}$ would be (by the hypothesis itself) $\geq \varepsilon_l$.

Hence, the probability that the experienced N_l failures to draw such a VP by means of N_l independent and identically distributed (iid) samples actually occur is $\leq (1-\varepsilon_l)^{N_l}$.

Since $\varepsilon_l = 1 - \delta^{1/N_l}$ (see algorithm in Figure 1), we have $(1 - \varepsilon_l)^{N_l} = (\delta^{1/N_l})^{N_l} = \delta$. The thesis follows. □

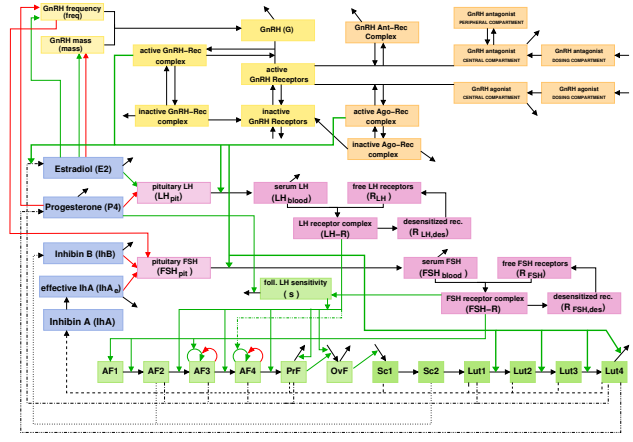

Fig. 1. Components of the human female menstrual cycle and their relationships, as modelled within GynCycle.

SM2 Computing complete stratified populations for a VPH model of the HPG axis

SM2.1 The GynCycle model

GynCycle [1] is a VPH model of the human female Hypothalamic-Pituitary-Gonadal (HPG) axis, focussing on the interactions and feedback mechanisms between components involved in the menstrual cycle like GnRH, FSH, LH, development of follicles and corpus luteum, and the production of E2, P4, Inhibin A and Inhibin B.

GynCycle also comprises Pharmacokinetics/Pharmacodynamics (PK/PD) of GnRH analogues and their effects on the menstrual cycle, such as the suppression of gonadotropins.

Figure 1 shows, as a flowchart, a high-level view of the main components of the human female menstrual cycle as well as of their relationships, as modelled within GynCycle.

For each modelled biological species, GynCycle defines a parametric ODE. Below we show the ODEs defining the time evolution of the 4 hormones we regarded as *model observables* in our case-study, *i.e.*, LH, FSH, E2 and P4:

$$\begin{aligned}
\dot{\mathrm{LH}}_{blood}(t) &= \frac{\mathrm{LH_{pit}}(t)}{V_{blood}} \cdot \left(b^{\mathrm{LH}}_{Rel} + k^{\mathrm{LH}}_{\text{G-R}} \cdot H^{+}(\text{G-R}(t), T^{\mathrm{LH}}_{\text{G-R}}; n^{\mathrm{LH}}_{\text{G-R}})\right) \\
&\quad - (k^{\mathrm{LH}}_{on} \cdot \mathrm{R_{LH}}(t) + k^{\mathrm{LH}}_{cl}) \cdot \mathrm{LH}_{blood}(t) \\
\dot{\mathrm{FSH}}_{blood}(t) &= \frac{\mathrm{FSH_{pit}}(t)}{V_{blood}} \cdot \left(b^{\mathrm{FSH}}_{Rel} + k^{\mathrm{FSH}}_{\text{G-R}} \cdot H^{+}(\text{G-R}(t), T^{\mathrm{FSH}}_{\text{G-R}}; n^{\mathrm{FSH}}_{\text{G-R}})\right) \\
&\quad - (k^{\mathrm{FSH}}_{on} \cdot \mathrm{R_{FSH}}(t) + k^{\mathrm{FSH}}_{cl}) \cdot \mathrm{FSH}_{blood}(t) \\
\dot{\mathrm{E2}}(t) &= b^{\mathrm{E2}} + k^{\mathrm{E2}}_{\mathrm{AF2}} \cdot \mathrm{AF2}(t) + k^{\mathrm{E2}}_{\mathrm{AF3}} \cdot \mathrm{LH}(t) \cdot \mathrm{AF3}(t) \\
&\quad + k^{\mathrm{E2}}_{\mathrm{AF4}} \cdot \mathrm{AF4}(t) \quad + k^{\mathrm{E2}}_{\mathrm{PrF}} \cdot \mathrm{LH}(t) \cdot \mathrm{PrF} \\
&\quad + k^{\mathrm{E2}}_{\mathrm{Lut1}} \cdot \mathrm{Lut1}(t) + k^{\mathrm{E2}}_{\mathrm{Lut4}} \cdot \mathrm{Lut4}(t) - k^{\mathrm{E2}}_{cl} \cdot \mathrm{E2}(t) \\
\dot{\mathrm{P4}}(t) &= b^{\mathrm{P4}} + k^{\mathrm{P4}}_{\mathrm{Lut4}} \cdot \mathrm{Lut4}(t) - k^{\mathrm{P4}}_{cl} \cdot \mathrm{P4}(t)
\end{aligned}$$

In the differential equations above, $\mathrm{LH}(t)$, $\mathrm{FSH}(t)$ etc. denote species of interest, while b^{LH}_{Rel}, k^{LH}_{cl}, etc. denote *model parameters*, which typically define patient-specific quantities like decay or reaction rates.

Furthermore, as it often happens in complex VPH models, some represented biological mechanisms are not known exactly. This is handled by introducing additional parameters.

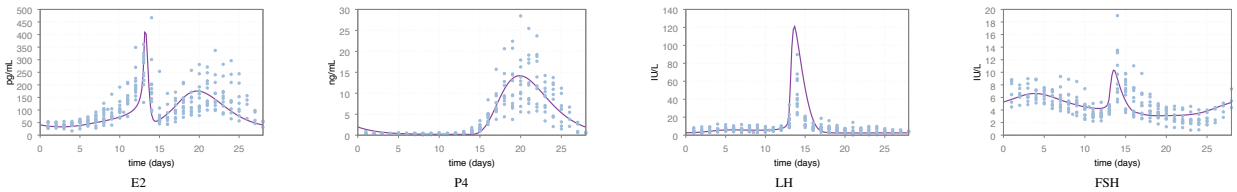

Fig. 2. GynCycle evolution under the reference VP (curves) averaging data of 12 patients (points) [1].

For example, in order to model stimulatory or inhibitory effects, additional parameters are introduced, together with *Hill functions* of the form:

$$H^+(S(t), T; n) = \frac{(S(t)/T)^n}{1 + (S(t)/T)^n}$$

$$H^-(S(t), T; n) = \frac{1}{1 + (S(t)/T)^n}$$

Here, $S(t) \geq 0$ denotes the influencing substance, $T > 0$ is the amount of S that causes 50% of the maximum of H^+ or H^-, and $n \geq 1$ is the Hill coefficient, which determines the rate of switching. The exact values for T and n are not easily observable and hence only a suitable range is known, typically derived from indirect arguments.

The whole GynCycle comprises a total of 33 parametric ODEs and 76 model parameters whose values (real numbers) are patient-specific. Hence, for our goals, the model parameter space is a subset of $\mathbb{R}^{76}$, which we finitised into a set Λ of 10^{76} elements (see Section 3.1 of the main paper).

SM2.2 Physiological meaningfulness

The GynCycle model is equipped with a *reference virtual patient*, $\lambda^{(0)} \in \Lambda$, which has been computed in [1] using a Pfizer database comprising 20–25 measures for 4 observed hormones (E2, P4, FSH, and LH) on 12 healthy women, totalling more than 1000 overall measurements. The GynCycle time evolution of E2, P4, LH, and FSH of the reference VP $\lambda^{(0)}$ is shown in Figure 2.

SM2.2.1 Representative portfolio of input functions

As discussed in the main paper (Section 3.2), in order to derive VPs for GynCycle whose behaviour is meaningful also when drugs are administered, we defined a *representative portfolio* $\mathbf{U}$ of 5 different input functions.

Beyond the no-drug input (under which the GynCycle observation function must show a behaviour similar to that in Figure 2, hence representing a healthy natural menstrual cycle), we considered the following standard treatment strategies, consisting of daily administrations of different doses for each of the two pharmaceutical compounds supported by the model: Triptorelin and Norethisterone, whose effect is the inhibition of the regular menstrual cycle. These strategies are common in fertility treatments.

Our portfolio $\mathbf{U}$ of input functions is defined as follows:

1. No drug;
2. A standard dose of Triptorelin (0.1 mg) for 28 days starting from cycle day 31;
3. 50% of a standard dose of Triptorelin (0.05 mg) for 28 days starting from cycle day 31;
4. A standard dose of Norethisterone (10 mg) for 10 days starting from cycle day 25;
5. 50% of a standard dose of Norethisterone (5 mg) for 10 days starting from day 25.

Figure 3 and Figure 4 show the expected effects of, respectively, Triptorelin and Norethisterone injections on the observable hormones, namely LH, FSH, E2 and P4.

SM2.2.2 Physiological meaningfulness as qualitative similarity

As discussed in the main paper (Section 3.2), in order to define our physiological meaningfulness criteria, we exploited qualitative similarity metrics standard in (discrete-time) *signal processing* (see, *e.g.*, 3). Here, we formally detail such metrics.

Let $\mathcal{S} = (\mathcal{T}, \Lambda, \mathcal{U}, \mathcal{Y}, \mathbf{y})$ (Definition 1) be a VPH model, whose output space $\mathcal{Y}$ is $\mathbb{R}^n_{0+}$ where $n \in \mathbb{N}_+$ is the number of (real-valued) observables.

Given a time function $\mathbf{f} : \mathcal{T} \to \mathcal{W}$ (with $\mathcal{W}$ being any non-empty set of values) and $\alpha \in \mathbb{R}_+$ and $\tau \in \mathbb{R}_{0+}$, we denote with $\mathbf{f}^{\alpha,\tau} : \mathcal{T} \to \mathcal{W}$ the time function defined, for all time points $t \in \mathcal{T}$, as:

$$\mathbf{f}^{\alpha,\tau}(t) = \mathbf{f}(\alpha t + \tau).$$

Thus, $\mathbf{f}^{\alpha,\tau}$ is function $\mathbf{f}$ time-shifted by τ and time-stretched by factor α.

Let also $\lambda^{(0)} \in \Lambda$ be a distinguished VP of $\mathcal{S}$(the reference VP as for GynCycle) and $\mathbf{U} \subset \mathcal{U}^{\mathcal{T}}$ be our portfolio of representative input functions for $\mathcal{S}$ (for GynCycle, the portfolio defined in Section SM2.2.1).

Given a parameter assignment $\lambda \in \Lambda$, values for $\alpha \in \mathbb{R}_+$ (time stretch) and $\tau \in \mathbb{R}_{0+}$ (time shift), a model observable i and an input function $\mathbf{u} \in \mathbf{U}$, we define the following two qualitative similarity metrics between the $\mathcal{S}$ evolution of observable i under parameter λ (time-shifted by τ and time-stretched by α) and reference parameter $\lambda^{(0)}$, when subject to the same input function $\mathbf{u}$:

- Normalised Zero-Lag Cross-Correlation (NZC):

$$\rho_{\mathbf{u},\lambda^{(0)},\lambda,i}(\alpha,\tau) = \frac{\sum_{t\in\mathcal{T}} \mathbf{y}_i^{\alpha,\tau}(t;\mathbf{u},\lambda) \times \mathbf{y}_i(t;\mathbf{u},\lambda^{(0)})}{||\mathbf{y}_i^{\alpha,\tau}(\mathbf{u},\lambda)|| \times ||\mathbf{y}_i(\mathbf{u},\lambda^{(0)})||} \quad (1)$$

- Normalised Energy Difference (NED):

$$E_{\mathbf{u},\lambda^{(0)},\lambda,i}(\alpha,\tau) = \frac{\left| ||\mathbf{y}_i^{\alpha,\tau}(\mathbf{u},\lambda)||^2 - ||\mathbf{y}_i(\mathbf{u},\lambda^{(0)})||^2 \right|}{||\mathbf{y}_i(\mathbf{u},\lambda^{(0)})||^2} \quad (2)$$

Intuitively, a high-enough NZC between two functions shows that they share peaks and valleys (this metrics takes its maximum value 1 when the two functions differ only by a multiplicative factor), while a low-enough NED shows that the two functions have a similar energy.

Definition 3 (Similarity-based physiological admissibility function). *Let $\mathcal{S} = (\mathcal{T}, \Lambda, \mathcal{U}, \mathcal{Y}, \mathbf{y})$ (Definition 1) be a VPH model, whose output space $\mathcal{Y}$ is $\mathbb{R}^n_{0+}$ where $n \in \mathbb{N}_+$ is the number of (real-valued) observables.*

Let also $\theta_\rho, \theta_E \in \mathbb{R}_+$ be two thresholds*, $\mathbb{A}$, $\mathbb{T}$ and $\mathbb{B}_1, \ldots \mathbb{B}_n$ be bounded ranges of $\mathbb{R}_{0+}$ (*time-stretch*, time-shift and* observable bounds *ranges) such that $1 \in \mathbb{A}$ and $0 \in \mathbb{T}$.*

We define our similarity-based physiological admissibility function $\varphi : \Lambda \to \{0, 1\}$ as follows. For any $\lambda \in \Lambda$, $\varphi(\lambda) = 1$ if there exists $\alpha \in \mathbb{A}$,

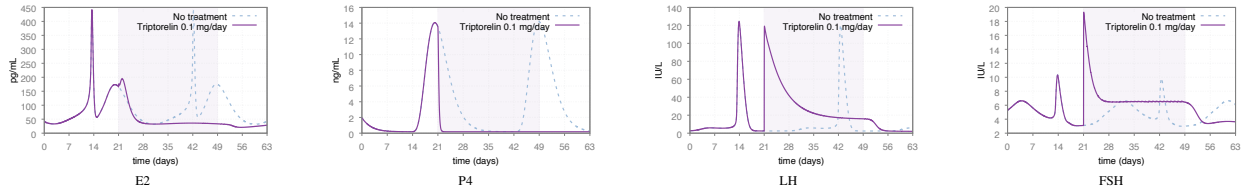

Fig. 3. Expected effects Triptorelin on the menstrual cycle. The purple area defines the time window during which the drug is administered.

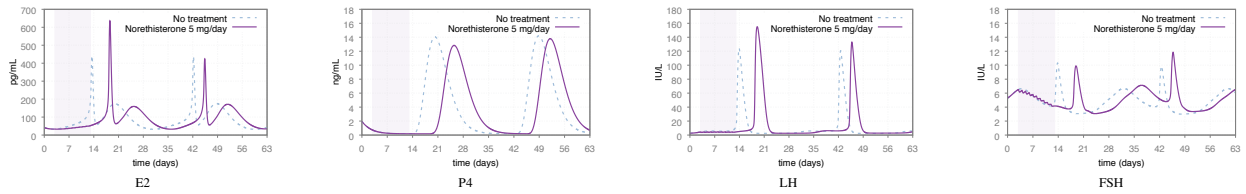

Fig. 4. Expected effects Norethisterone on the menstrual cycle. The purple area defines the time window during which the drug is administered.

$\tau \in \mathbb{T}$ *such that, for each model observable* $i \in [1, n]$ *and* $\mathbf{u} \in \mathbf{U}$, all *conditions below hold:*

- $\rho_{\mathbf{u},\lambda^{(0)},\lambda,i}(\alpha, \tau) \geq \theta_\rho$
- $E_{\mathbf{u},\lambda^{(0)},\lambda,i}(\alpha, \tau) \leq \theta_E$
- $\left[\min_{t\in\mathcal{T}}(\mathbf{y}_i^{\alpha,\tau}(t; \mathbf{u}, \lambda)), \max_{t\in\mathcal{T}}(\mathbf{y}_i^{\alpha,\tau}(t; \mathbf{u}, \lambda))\right] \subseteq \mathbb{B}_i$.

We have $\varphi(\lambda) = 0$ *otherwise.*

In other words, Definition 3 declares a parameter value λ as physiologically meaningful (*i.e.*, a VP) if the entailed model evolution can be time-scaled and/or -shifted so that all observables are within the given physiological bounds and the values of all similarity metrics are within the given thresholds.

In order to compute $\varphi(\lambda)$ for any given λ, we need to find suitable values for α and τ, which, however, are *real* values. To cope with this problem, we deployed the following (greedy) approach, starting from the model evolutions computed by the simulator under parameter λ and $\lambda^{(0)}$ and all input functions in $\mathbf{U}$: we considered only the (α, τ) pairs for which $\mathbf{y}_i^{\alpha,\tau}(t; \mathbf{u}, \lambda)$ and $\mathbf{y}_i(t; \mathbf{u}, \lambda^{(0)})$ share, in the time domain, at least a *local maximum* or a *local minimum* for at least one $\mathbf{u} \in \mathbf{U}$.

Since we defined the time set of GynCycle as a bounded interval of $\mathbb{N}$ (see, Section 3.1 in the main paper), we can efficiently enumerate all possible values for (α, τ) (which represent all possible peak-alignments) by defining and computing all solutions of a Constraint Satisfaction Problem (CSP) (by means of an off-the-shelf solver). Namely, for each candidate assignment to α and τ (a solution of our CSP), we compute our qualitative similarity metrics, and return $\varphi(\lambda) = 1$ as soon as we find one assignment for which the qualitative similarity metrics satisfy the given thresholds (and 0 if we do not find any).

SM2.3 Stratified phenotypes

Forthcoming Definition 4 formalises the equivalence relation $\sim_\psi$ we used to define behavioural indistinguishability among GynCycle VPs. Equivalence relation $\sim_\psi$ is parametric in $\psi \in \mathbb{R}_+$, the *quantisation factor*.

Definition 4 (Equivalence relation $\sim_\psi$). *Let* $\mathcal{S} = (\mathcal{T}, \Lambda, \mathcal{U}, \mathcal{Y}, \mathbf{y})$ *be a dynamical system with* $\mathcal{T} = \mathbb{N}_+$*, parameter space* Λ*, and* $\mathcal{Y} = \mathbb{R}_{0+}^n$*, where* $n \in \mathbb{N}_+$ *is the number of model observables. Let also* $\lambda^{(0)} \in \Lambda$ *be a distinguished parameter assignment,* $\psi \in \mathbb{R}_+$ *and* $\mathbf{U} \subseteq \mathcal{U}^\mathcal{T}$ *be a set of input functions for* $\mathcal{S}$*.*

Given $\lambda^{(1)}, \lambda^{(2)} \in \Lambda$*, let* $Y_{c,i,\mathbf{u}}^{(k)}$ *be the c-th coefficient (*$c \in [0, h-1]$*) of the Discrete Fourier Transform (DFT) of the time evolution of the* i*-th model observable (*$i \in [1, n]$*) when the model is fed with input* $\mathbf{u} \in \mathbf{U}$ *under parameter* $\lambda^{(k)}$ *(*$k \in [0, 2]$*, which refers to the distinguished* $\lambda^{(0)}$ *when* $k = 0$*).*

The quantum *of* $Y_{c,i,\mathbf{u}}^{(k)}$*,* $Q_\psi(Y_{c,i,\mathbf{u}}^{(k)})$*, is:* $\left\lfloor \frac{\left|Y_{c,i,\mathbf{u}}^{(k)}\right|^2}{\psi \left|\left|\mathbf{y}_i(\mathbf{u}, \lambda^{(0)})\right|\right|^2} \right\rfloor$.

We stipulate that $\lambda^{(1)} \sim_\psi \lambda^{(2)}$ *if, for each* $\mathbf{u} \in \mathbf{U}$*,* $c \in [0, h-1]$*, and* $i \in [1, n]$*, it holds* $Q_\psi(Y_{c,i,\mathbf{u}}^{(1)}) = Q_\psi(Y_{c,i,\mathbf{u}}^{(2)})$.

Remark 1 formalises the observation we made in Section 3.3 of the main paper, namely that our definition of $\sim_\psi$ (Definition 4 in the main paper) immediately gives us a theoretical upper bound to the NED shown by the observation functions of any two VPs belonging to the same equivalence class of $\sim_\psi$, for any model observable $i \in [1, n]$ and input function $\mathbf{u} \in \mathbf{U}$.

Remark 1. *Let* $\mathcal{S} = (\mathcal{T}, \Lambda, \mathcal{U}, \mathcal{Y}, \mathbf{y})$*,* $\lambda^{(0)}$*,* n*,* ψ*,* $\mathbf{U}$*, and* $\sim_\psi$ *be as in Definition 4. For each* $i \in [1, n]$ $\mathbf{u} \in \mathbf{U}$*, and* $\lambda \in \Lambda$*, let* $\mathbf{y}_i(\mathbf{u}, \lambda)$ *be the (real-valued) observation function of the* i*-th model observable under parameter* λ *when* $\mathcal{S}$ *is fed with input function* $\mathbf{u}$*.*

For every $\lambda^{(1)}, \lambda^{(2)} \in \Lambda$ *such that* $\lambda^{(1)} \sim_\psi \lambda^{(2)}$ *we have that, for each* $i \in [1, n]$ *and* $\mathbf{u} \in \mathbf{U}$*, the Normalised Energy Difference (NED) between* $\mathbf{y}_i(\mathbf{u}, \lambda^{(1)})$ *and* $\mathbf{y}_i(\mathbf{u}, \lambda^{(2)})$ *is upper-bounded by* ψ*, that is:*

$$\frac{\left|\left|\left|\mathbf{y}_i(\mathbf{u}, \lambda^{(1)})\right|\right|^2 - \left|\left|\mathbf{y}_i(\mathbf{u}, \lambda^{(2)})\right|\right|^2\right|}{\left|\left|\mathbf{y}_i(\mathbf{u}, \lambda^{(0)})\right|\right|^2} \leq \psi.$$

Proof. Time evolution $\mathbf{y}_i(\mathbf{u}, \lambda^{(k)})$ ($k \in [0, 2]$) is a vector of real-valued samples $\mathbf{y}_i(\mathbf{u}, \lambda^{(k)}) = (y_{t,i,\mathbf{u}}^{(k)})_{t=0}^{h-1} \in \mathbb{R}_{0+}^h$.

Let $Y_{i,\mathbf{u}}^{(k)} = (Y_{c,i,\mathbf{u}}^{(k)})_{c=0}^{h-1} \in \mathbb{C}^h$ be the vector of coefficients of the DFT of $\mathbf{y}_i(\mathbf{u}, \lambda^{(k)})$, each one being a complex number. Namely: $Y_{c,i,\mathbf{u}}^{(k)} = \sum_{t=0}^{h-1} y_{t,i,\mathbf{u}}^{(k)} \exp(-\frac{2\pi \jmath t c}{h})$, where $\jmath = \sqrt{-1}$.

We recall that, given vector $\mathbf{X} = (X_c)_{c=0}^{h-1} \in \mathbb{C}^h$ (or, as a special case, $\mathbf{X} \in \mathbb{R}_{0+}^h$), the squared L2-norm of $\mathbf{X}$ is $||\mathbf{X}||^2 = \sum_{c=0}^{h-1} |X_c|^2$.

From the Parserval's theorem we know (see, *e.g.*, 3) that

$$\left|\left|\mathbf{y}_i(\mathbf{u}, \lambda^{(k)})\right|\right|^2 = \frac{1}{h}\left|\left|Y_{i,\mathbf{u}}^{(k)}\right|\right|^2.$$

Hence, $\left|\left|\left|\mathbf{y}_i(\mathbf{u}, \lambda^{(1)})\right|\right|^2 - \left|\left|\mathbf{y}_i(\mathbf{u}, \lambda^{(2)})\right|\right|^2\right|$ rewrites as:

$$\frac{1}{h}\sum_{c=0}^{h-1}\left|\left|Y_{c,i,\mathbf{u}}^{(1)}\right|^2 - \left|Y_{c,i,\mathbf{u}}^{(2)}\right|^2\right|.$$

Since $\lambda^{(1)} \sim_\psi \lambda^{(2)}$, we have, for each $c \in [0, h-1]$, $Q_\psi(Y^{(1)}_{c,i,\mathbf{u}}) = Q_\psi(Y^{(2)}_{c,i,\mathbf{u}})$, which implies (Definition 4) that

$$\left| \left| Y^{(1)}_{c,i,\mathbf{u}} \right|^2 - \left| Y^{(2)}_{c,i,\mathbf{u}} \right|^2 \right| \leq \psi \left| \left| \mathbf{y}_i(\mathbf{u}, \lambda^{(0)}) \right| \right|^2$$

for all $c \in [0, h-1]$.

Thus:

$$\left| \left| Y^{(1)}_{i,\mathbf{u}} \right|^2 - \left| Y^{(2)}_{i,\mathbf{u}} \right|^2 \right| \leq h\psi \left| \left| \mathbf{y}_i(\mathbf{u}, \lambda^{(0)}) \right| \right|^2$$

and the thesis follows. □

SM3 Validation against clinical data: quantitative evaluation

We formalise each health record r in our datasets as a pair (o, u), where o (observations) consists of a time series $o(s)$ for each $s \in \{\text{E2, P4, LH, FSH}\}$ (observable species) and u (inputs) consists of a time series $u(d)$ of each pharmaceutical compound d (drug) administered in r. In turn, each time series q (defining $|q|$ elements, either measurements or inputs) is a set of pairs $\{\text{time}(q,j), \text{val}(q,j)\}_1^{|q|}$, where, for each j, $\text{time}(q,j)$ and $\text{val}(q,j)$ are, respectively, the time instant and the value (observation or input) of the j-th available element in q.

Let r be a health record reporting a time series for the measurements of each observable s of our VPH model (*i.e.*, LH, FSH, E2, P4 in our experiments), and let $\mathbf{u}_r$ be the input function defining the sequence of drug administrations which the patient was subject to.

Given a VP λ, the Average Normalised Mean Absolute Error $\text{aNMAE}(\lambda, r)$ of λ with respect to r is the average (among the n_s observables s, 4 in our experiments) of the average (among the m_s clinical measurements of s in r) of the absolute errors between the model output (under parameter λ and input function $\mathbf{u}_r$) and the j-th measurement of s in r ($j \in [1, m_s]$), normalised with respect to the maximum measured value $\text{maxval}(s)$ of s in r. Namely:

$$\text{aNMAE}(\lambda, r) = \frac{1}{n_s} \sum_{s=1}^{n_s} \frac{1}{m_s} \sum_{j=1}^{m_s} \frac{|\mathbf{y}(\text{time}(s,j); \mathbf{u}_r, \lambda) - \text{val}(s,j)|}{\text{maxval}(s)}.$$

In the formula above, model trajectories $\mathbf{y}(\mathbf{u}_r, \lambda)$ have been aligned with respect to the LH peak (typically used to estimate the ovulation day) shown in each health record in order to take into account for their transient periods.

SM4 List of Acronyms

aNMAE	Average Normalised Mean Absolute Error
APP	All-Different Phenotype Population
CAPP	Complete APP
CSP	Constraint Satisfaction Problem
DFT	Discrete Fourier Transform
HPC	High Performance Computing
HPG	Hypothalamic-Pituitary-Gonadal
iid	independent and identically distributed
ISCT	*In Silico* Clinical Trial
NED	Normalised Energy Difference
NZC	Normalised Zero-Lag Cross-Correlation
ODE	Ordinary Differential Equation
PK	Pharmacokinetics
PBPK	Physiologically-based Pharmacokinetics
PK/PD	Pharmacokinetics/Pharmacodynamics
SMC	Statistical Model Checking
VP	Virtual Patient
VPH	Virtual Physiological Human

References

[1]S. Röblitz, C. Stötzel, P. Deuflhard, H. Jones, D.-O. Azulay, P. van der Graaf, and S. Martin. A mathematical model of the human menstrual cycle for the administration of GnRH analogues. *Journal of Theoretical Biology*, 321:8–27, 2013.

[2]E. Sontag. *Mathematical Control Theory: Deterministic Finite Dimensional Systems (2nd Ed.)*. Springer, 1998.

[3]S. Vaseghi. *Advanced Digital Signal Processing and Noise Reduction*. John Wiley & Sons, 2009.