Disease Degradation Models For Chronic Digestive Conditions

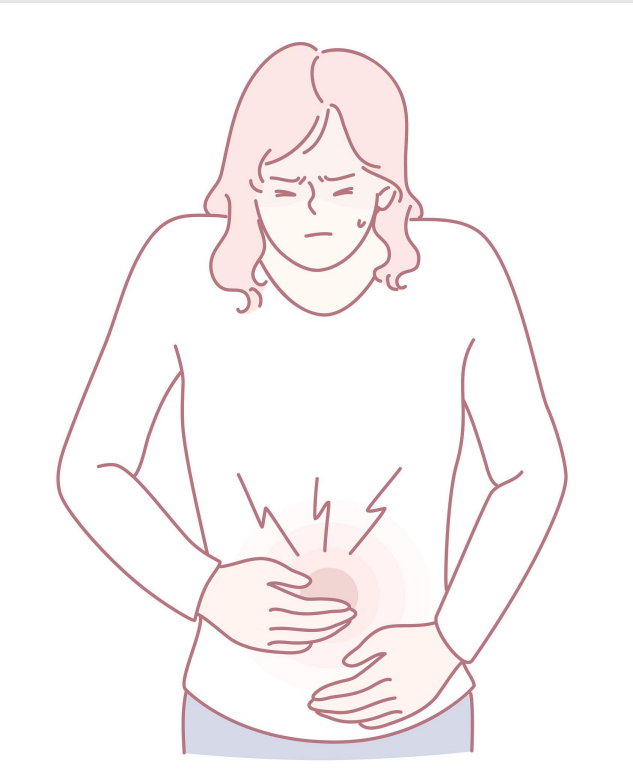
Christopher Agesen - NJIT
Talla Babou - MIT
Julianna Capece - FSU
Norman Chen - UCLA
Solari Elee - UD
Uzochi Gideon - RU
Anshi Gupta - UH
Tajkera Khatun - IU Indianapolis
Walter Kiefer - Vironix
Diana Y. Lee - SDSU
Martyna Lukaszewicz - UIdaho
Kees McGahan - BU
B Sagar - UCF
Kriti Sehgal - UChicago
Tamanna Dhankhar - NJIT
W. Hayden Taylor - UD

Math Problems in Industry Workshop
June 14-18, 2016

Overview

Intro - Background

- Chronic digestive diseases (IBD, IBS, GERD, Celiac disease) affect millions of people worldwide.
- Disease progression is nonlinear, with long periods of remission interrupted by unpredictable flare-ups.
- Flare: a sudden onset of a subset of the following symptoms - intense abdominal bowel pain, diarrhea / bloody poop, cramps, weight loss, fatigue, fever, vomiting.
- Repeated flares can lead to irreversible tissue damage, hospitalization, and reduced quality of life.



Intro - Problem Statement

"Our goal here is to develop a hybrid 'Disease Degradation Model' that integrates mechanistic foundations with statistical learning methods and clinical data"
Model Objectives:

- Short term prediction: active flares and flare forecast risk
- Mapping the long term progression of disease

Intro - Goals

1. Literature study: Identify and compare existing mechanistic models of the GI tract and ML-based diagnostic/prognostic models for IBD.Compare assumptions and what mechanistic details of the disease are available.
2. Formulate a set of models that describe a specific degradation process and its response to the given treatment protocol
3. Propose methods to assess model input parameters and couple model outputs to predict flares and long-term progression using patient longitudinal data
4. Identify which model parameters are most sensitive to patient-specific data. How can we validate the mechanistic plausibility of your hybrid model?

Overview

Data

Vironix could not release specific data for privacy reasons.

Therefore:

- Synthesize data as placeholder
- Develop useful workflows

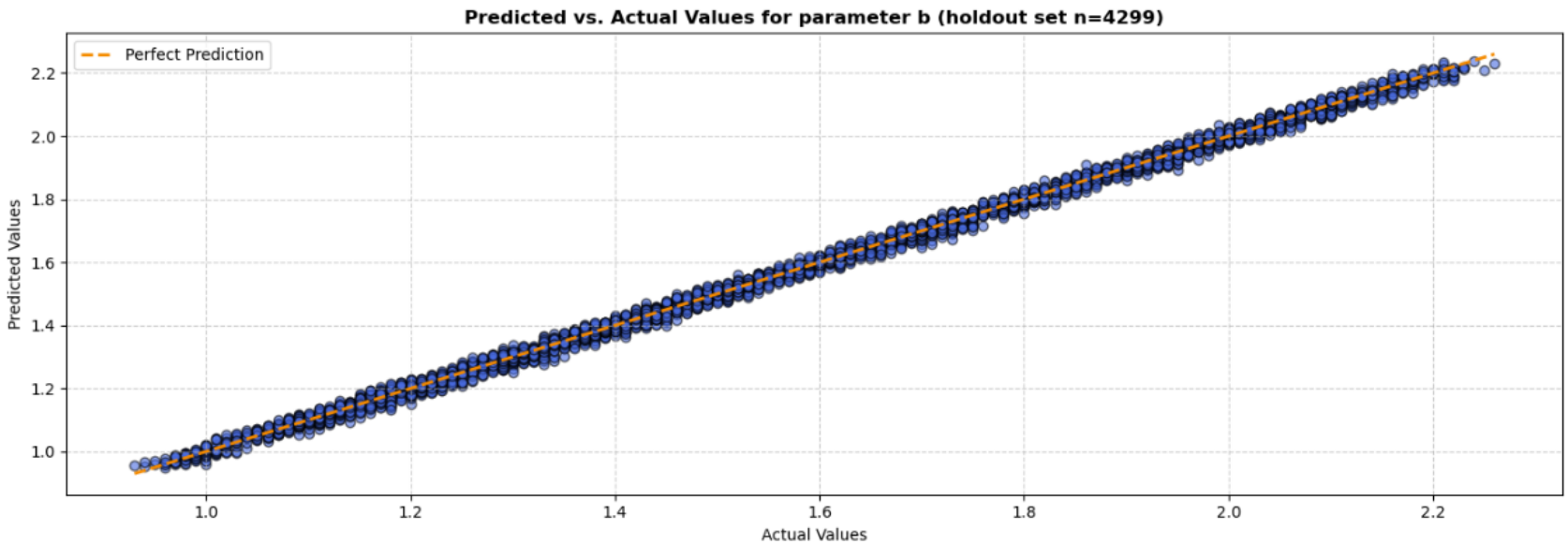
Data - Static

- Generated reasonable patient data
- We want to use ensemble learning models to estimate patient-specific parameters (for differential equation models later)

Parameter	**Definition**
b	Natural recovery rate
c_s	Treatment effect on active disease (short term)
c_l	Treatment effect on active disease (long term)
q_s	Treatment effect on symptoms (short term)
k	Symptom production rate from disease burden
r	Symptom recovery rate
m	Flare effect on long-term degradation
η_l	Long-term treatment effect on degradation

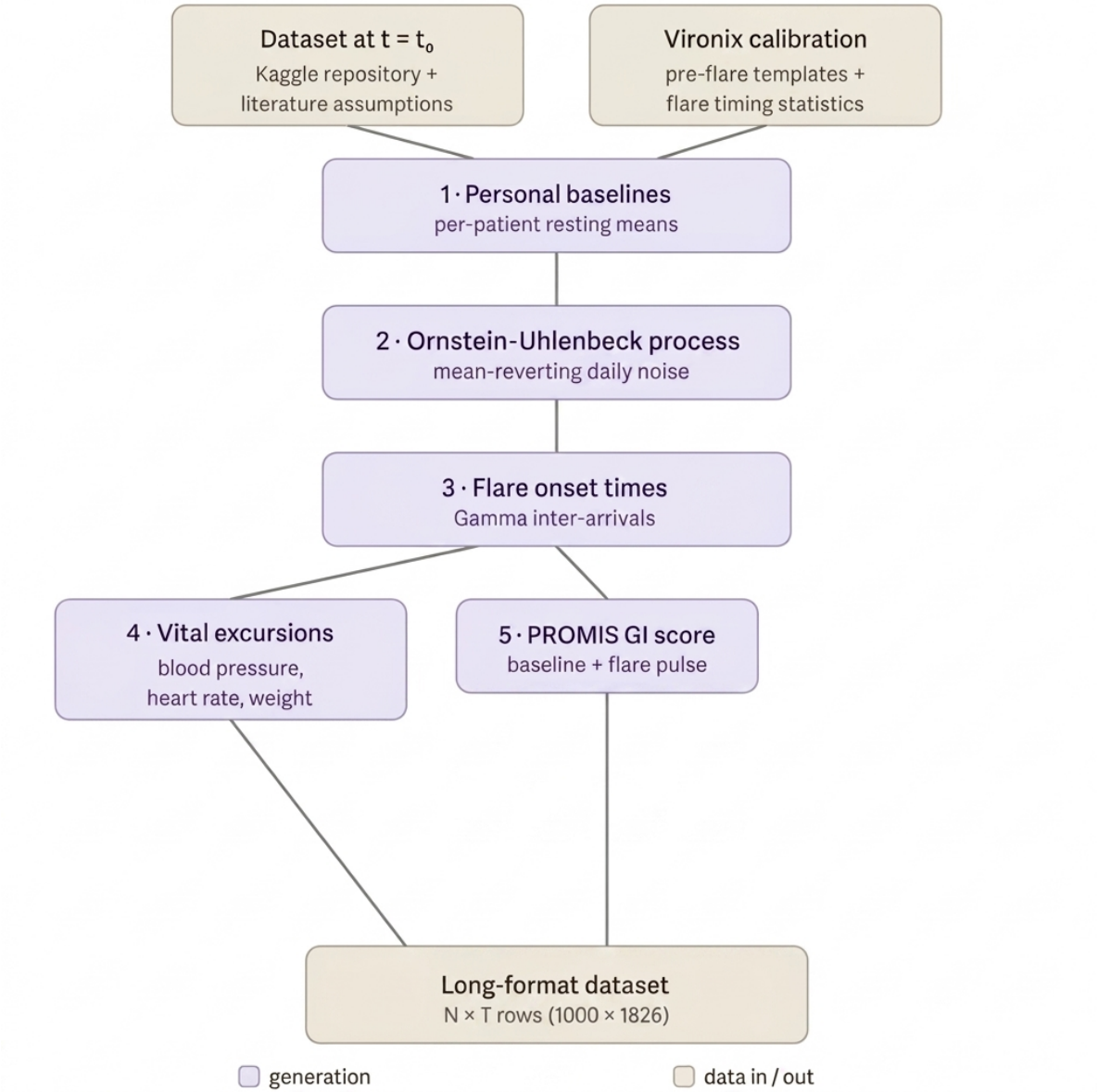
Random Forest Models

- Patient data $\longrightarrow$ random forest $\longrightarrow$ parameter values
- Each model is parameter-specific (starting with b).
- These predicted parameters are then used in the mechanistic disease model



Time Series Data 1 Algorithm

- Expands a cross-sectional snapshot into 5-year daily trajectories (vitals, symptoms, flare labels)
- Per-patient mean-reverting (Ornstein–Uhlenbeck) baselines, with Gamma-timed flares and pre-flare vital-sign changes taken from real Vironix data



Time Series Data 1 Characteristics

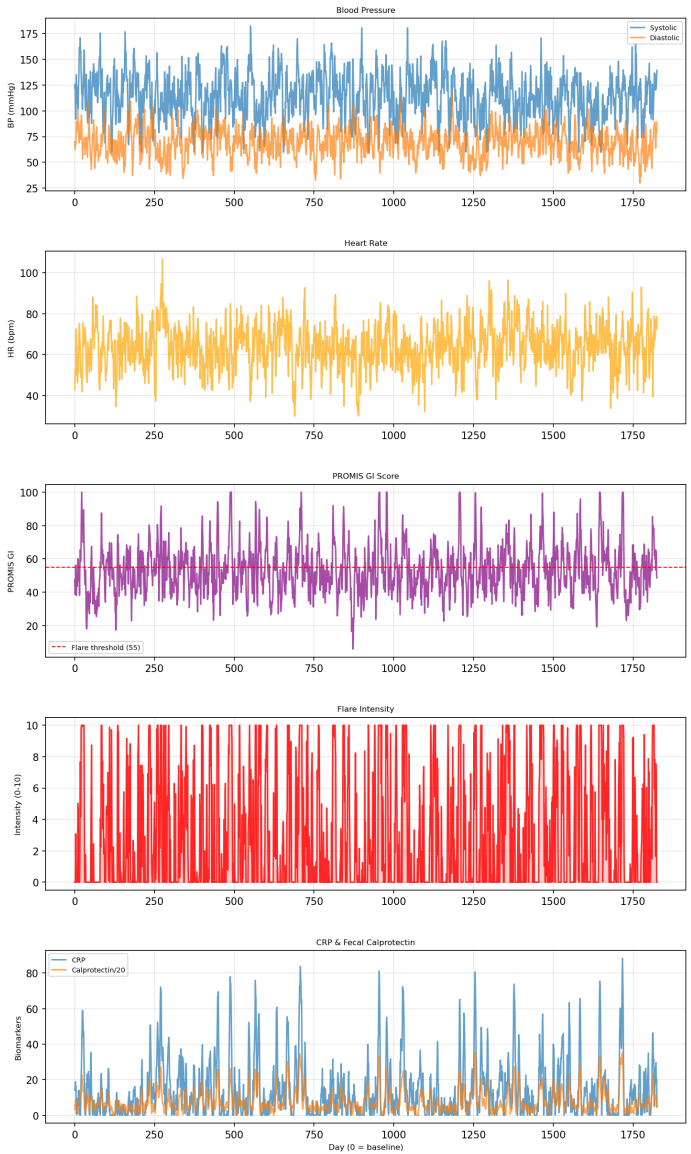
Patient 0 — 5-year trajectory | Gender: Female, Age: 9.0, Disease: Unexplained weight loss

CHECK 1: Shape | CHECK 2: Physiological Ranges

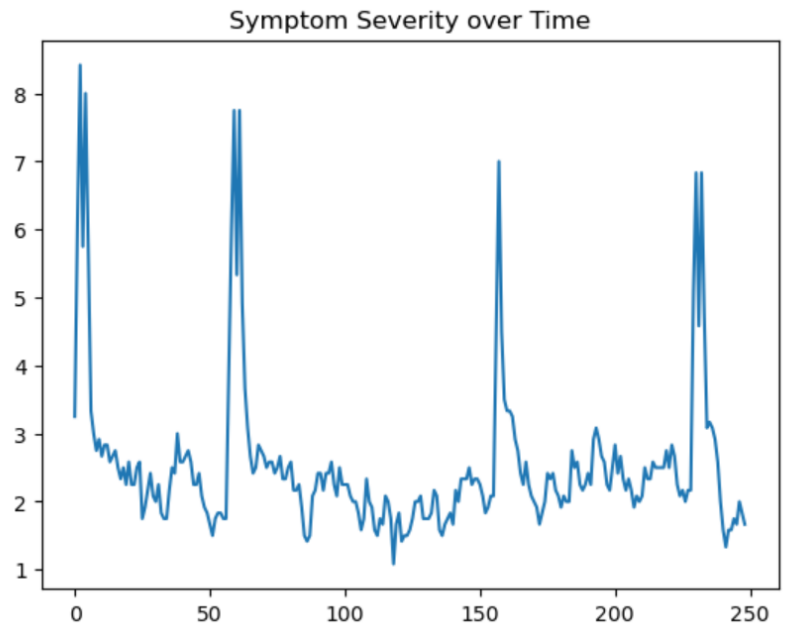
```
Shape: (1826000, 38)
Unique patients: 1000

--- Range Check ---
  systolic_bp              [  60,  220]  ->  OK
  diastolic_bp             [  30,  140]  ->  OK
  heart_rate               [  30,  200]  ->  OK
  spo2                     [  70,  100]  ->  OK
  weight_lbs               [  50,  600]  ->  OK
  bmi                      [  10,   80]  ->  WARNING: 2 out-of-range
  glucose                  [  50,  400]  ->  OK
  crp_esr                  [   0,  200]  ->  OK
  fecal_calprotectin       [   0, 2000]  ->  OK
  promis_gi_score          [   0,  100]  ->  OK
  flare_intensity          [   0,   10]  ->  OK
  upper_gi_score           [   0,   10]  ->  OK
  lower_gi_score           [   0,   10]  ->  OK
```

CHECK 3: Flare Rates | CHECK 4: Covariates | CHECK 5: FVC

```
--- Flare Rates (expect 0.10-0.30) ---
  flare_active rate:    0.4814  WARNING
  red_flag rate:        0.3788  OK (< flare rate)
  on_medication rate:   0.6367
  mean flare_intensity: 2.4222  OK

--- Covariate Direction (CHECK 4) ---
  Female PROMIS > Male  (expected)
  Smoker PROMIS < Non-smoker  (expected)
  Female: 51.002  vs  Male: 48.828  ->  OK
  Smoker: 49.129  vs  Non-smoker: 50.946  ->  WRONG DIRECTION

--- FVC/FEV1 Coverage (CHECK 5) ---
  COPD/ASTHMA rows:  270248
  Non-NaN FVC rows:  270248  OK
  FVC declining:     0.689  WARNING (<0.70)
```



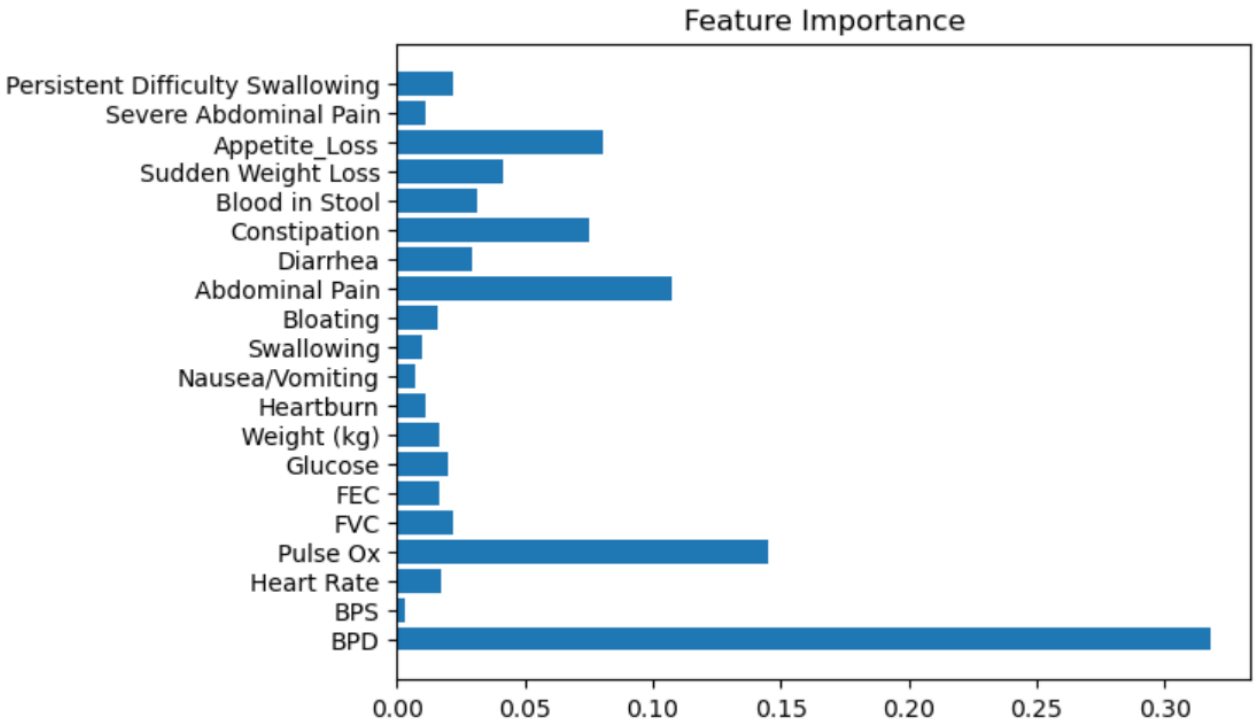
Time Series 2 Algorithm + Characteristics



- Patient measurements vary stochastically
- When certain measurements (e.g blood pressure) abnormal, symptoms increase, causing flare

Not realistic- but useful for testing workflows involving flare detection

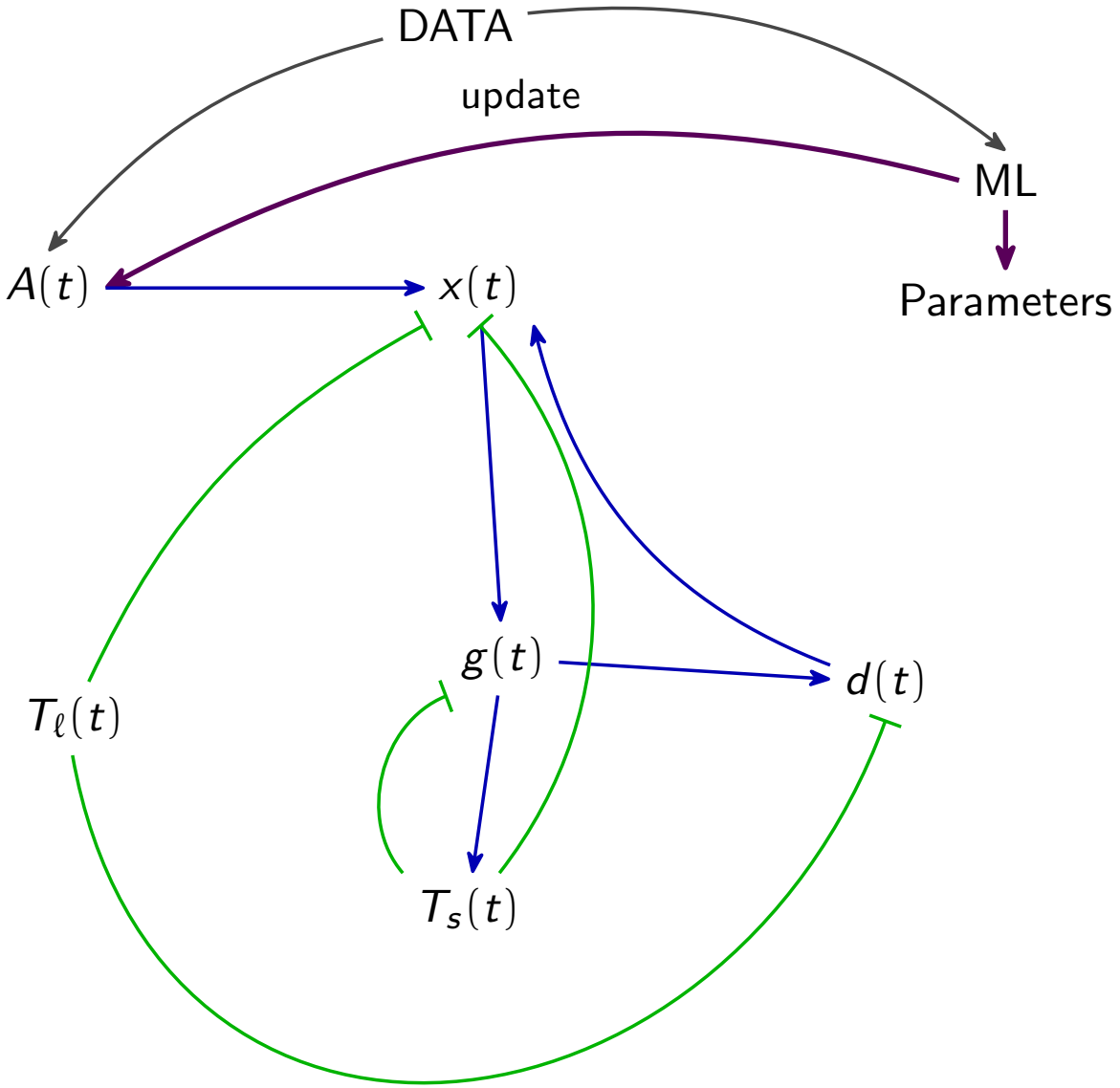
Machine Learning - Feature Importance



Use random forests to determine which features are most relevant for flares

Overview

Models - General Model Outline



Models - Deterministic Extensions

Active damage d_a and permanent scarring d_s, with feedback to disease activity.

$$d = d_a + d_s$$

$$d_a \longrightarrow d_s \longrightarrow x$$

Short term "rescue" T_s activated in response to flares, and long term maintenance T_l.

$$g \longrightarrow T_s \dashv g$$

$$T_s \dashv x$$

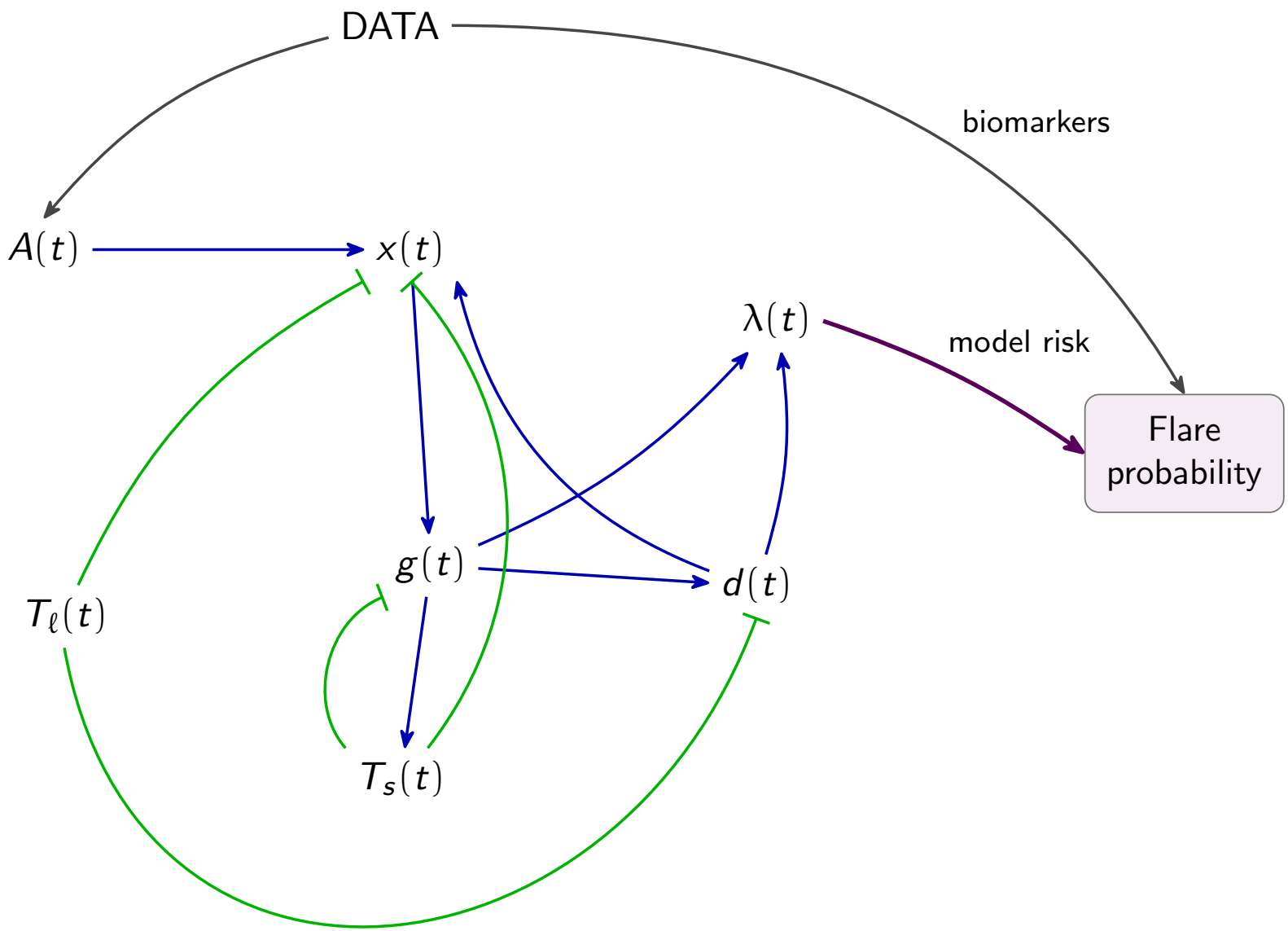
$$T_l \dashv x$$

$$T_l \dashv d_a$$

Additional Assumptions

- d_a converts to d_s at a given rate ϕ if left untreated.
- Permanent scarring contributes endogenous adverse load ρ to disease activity.
- T_l is a constant treatment adherence, and reduces active damage and disease activity.
- T_s is a proportional response to g with a logistic switch as symptom burden g crosses flare threshold g_c.

Stochastic Model Outline



Models — Stochastic Extensions (I)

Stochastic adverse load (Ornstein–Uhlenbeck in logit space)

$$dY = \theta\,(\bar{Y} - Y)\,dt + \sigma_A\,dW_t, \qquad A(t) = \sigma\big(Y(t)\big) \in [0,1].$$

Flares as a stochastic point process (state-dependent hazard)

$$\lambda(t) = \lambda_0 + \lambda_{\max}\,\sigma\big(\beta(g - g_c) + \beta_d\,d\big) + h(t)\big), \qquad \Pr(\text{flare in } dt) = 1 - e^{-\lambda\,dt}.$$

Treatment (logit-OU)

$$T_s = \sigma(Z_s), \qquad dZ_s = \kappa_s\big(\operatorname{logit} u_s - Z_s\big)dt + \sigma_{T_s}\,dW_t,$$

rescue T_s responds to symptoms; maintenance T_ℓ is analogous.

Biological motivation: disease is relapsing–remitting and bursty; inflammatory load drifts day to day; flares are *probabilistic*; drug levels vary with adherence.

Models — Stochastic Extensions (II): Flare Jumps

At a flare event the state jumps discontinuously:

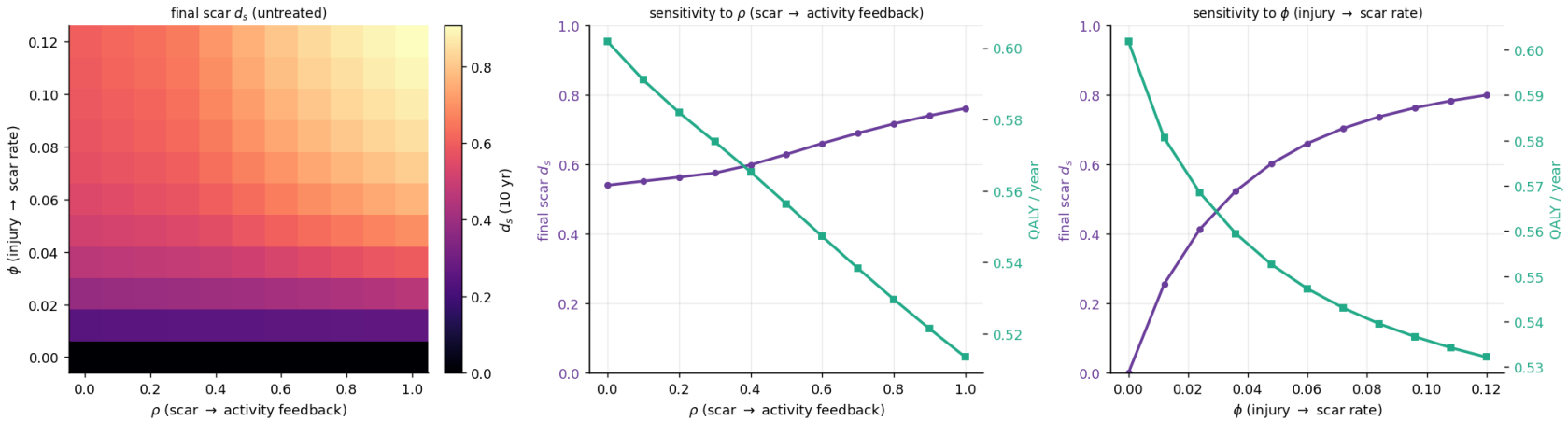
$$
\begin{aligned}
g &\mapsto g + J_g(1-g) && \text{(symptom spike)},\\
d_a &\mapsto d_a + J_d(1-d_a-d_s) && \text{(reversible damage)},\\
d_s &\mapsto d_s + J_s\,(g^+ - g_c)_+ && \text{(severe-flare scarring)},\\
h &\mapsto h + \gamma && \text{(self-excitation kick)},\\
T_s &\mapsto T_s + J_{T_s}(1-T_s) && \text{(rescue)}.
\end{aligned}
$$

Biological motivation

- A flare is a discrete hit: symptoms spike, damage accrues, rescue fires.
- Only *severe* flares ($g^+ > g_c$) scar — d_s is *irreversible* (never heals).
- Each flare transiently raises the hazard ($+\gamma$), so flares cluster.

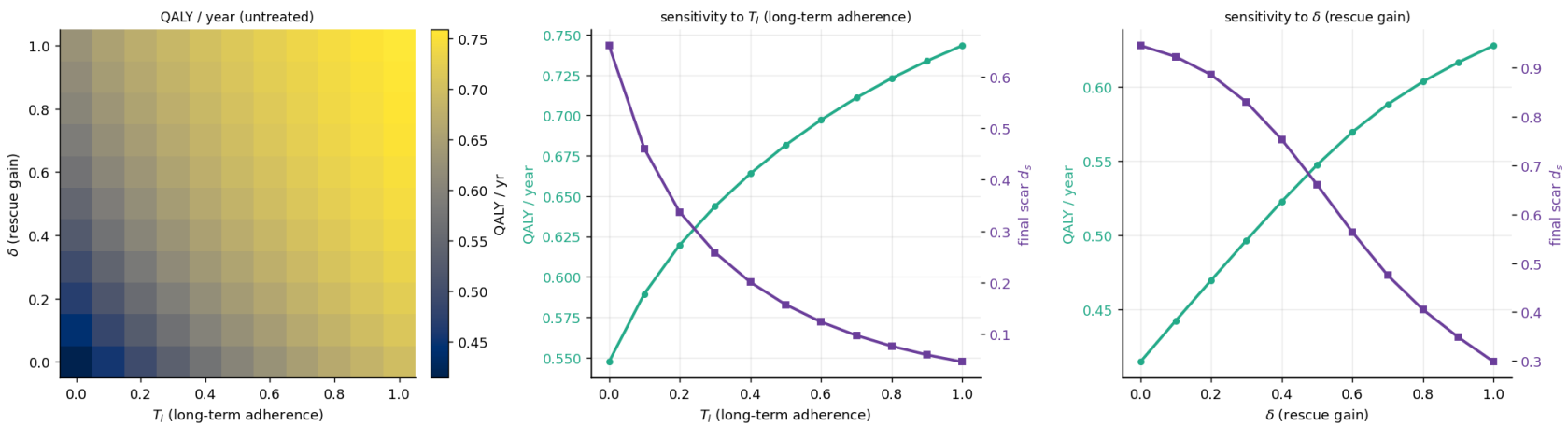
Sensitivity - Damage and Treatment

Fig. 16a final scar d_s over (ρ (scar → activity feedback), ϕ (injury → scar rate))



(a) Sensitivity to feedback ρ and irreversibility rate ϕ

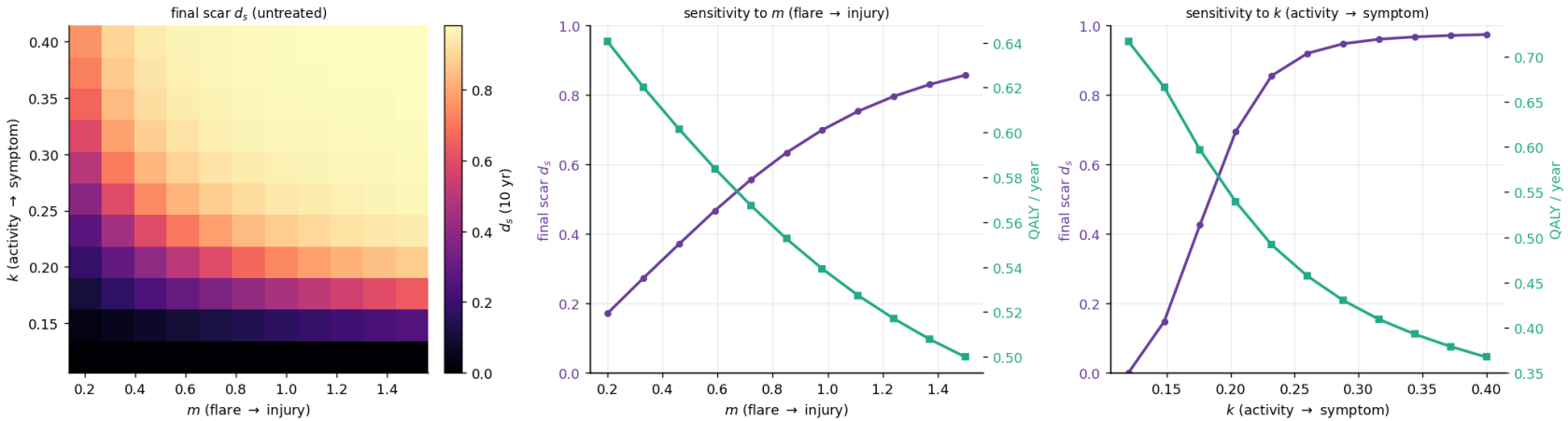
Fig. 16c QALY / year over (T_l (long-term adherence), δ (rescue gain))



(b) Sensitivity to treatment adherence T_l and rescue dose response δ

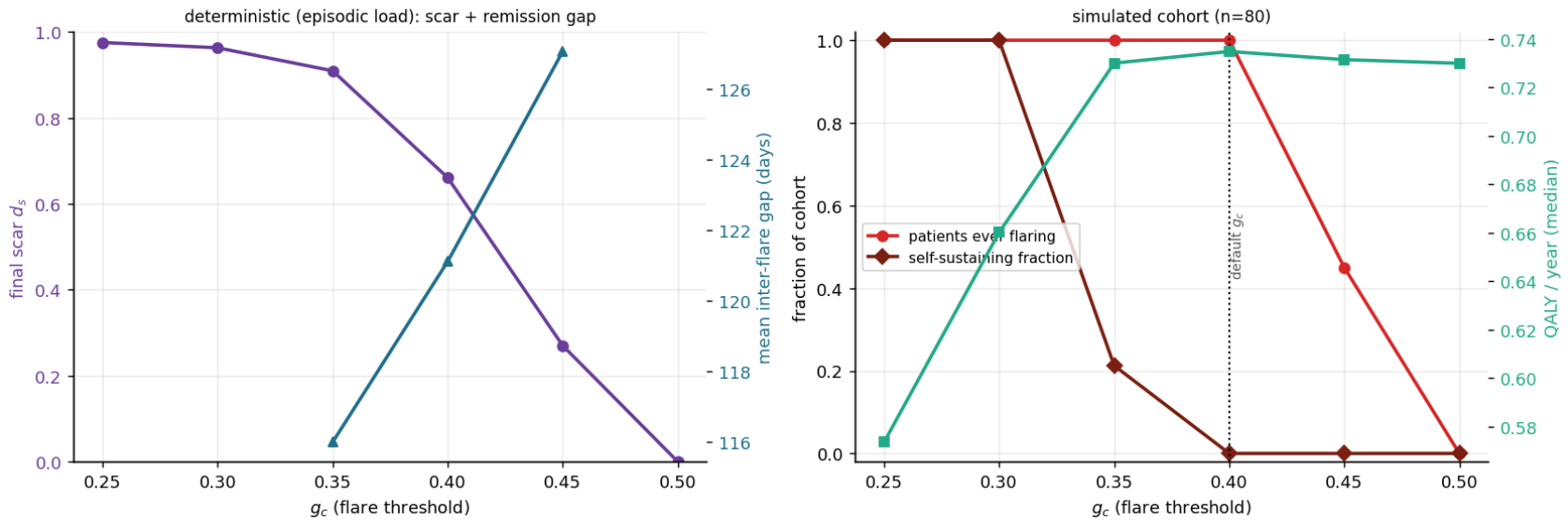
Sensitivity - Disease Characteristics and Flare Threshold

Fig. 17 final scar d_s over (m (flare → injury), k (activity → symptom))

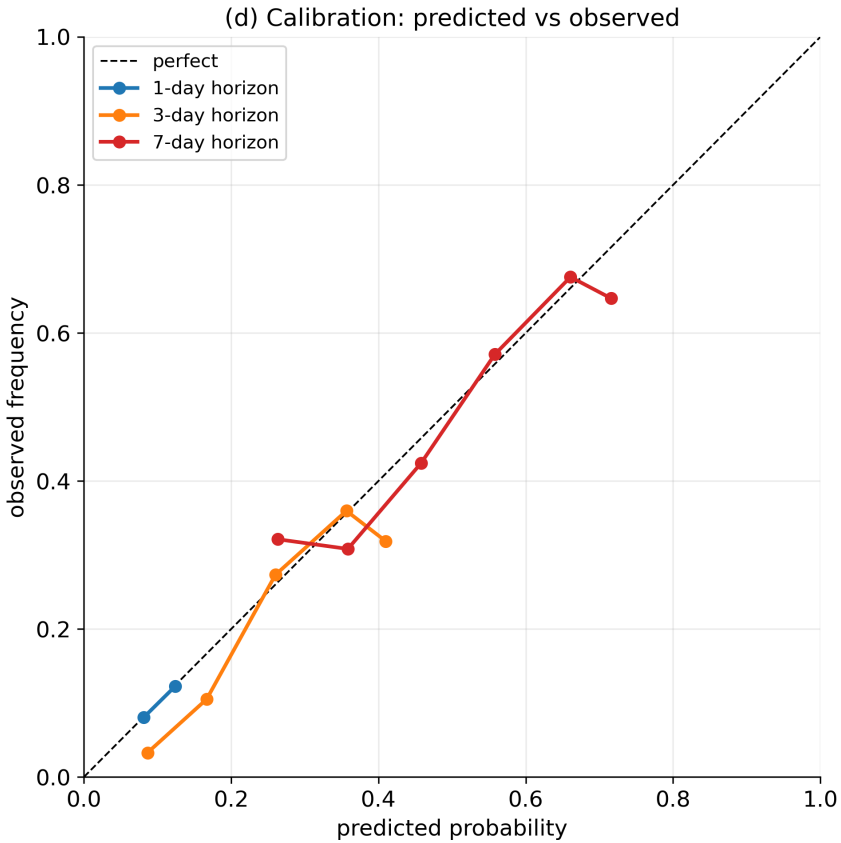


(a) Sensitivity to flare to injury rate m and activity to symptom rate k

Fig. 18 Flare-threshold sweep $g_c \in [0.25, 0.5]$



Results — Calibrated Flare-Probability Prediction



Predicted vs. observed flare frequency.

Held-out patients; data are synthetic, so calibration is the operative metric.

Horizon (d)	Base rate	Mean pred.	Calib. error
1	0.112	0.114	0.001
3	0.326	0.329	0.013
7	0.629	0.630	0.024

- Probabilities are **well-calibrated**: a predicted $X\%$ flares $\approx X\%$.
- **Unbiased** across horizons (mean prediction $\approx$ base rate).

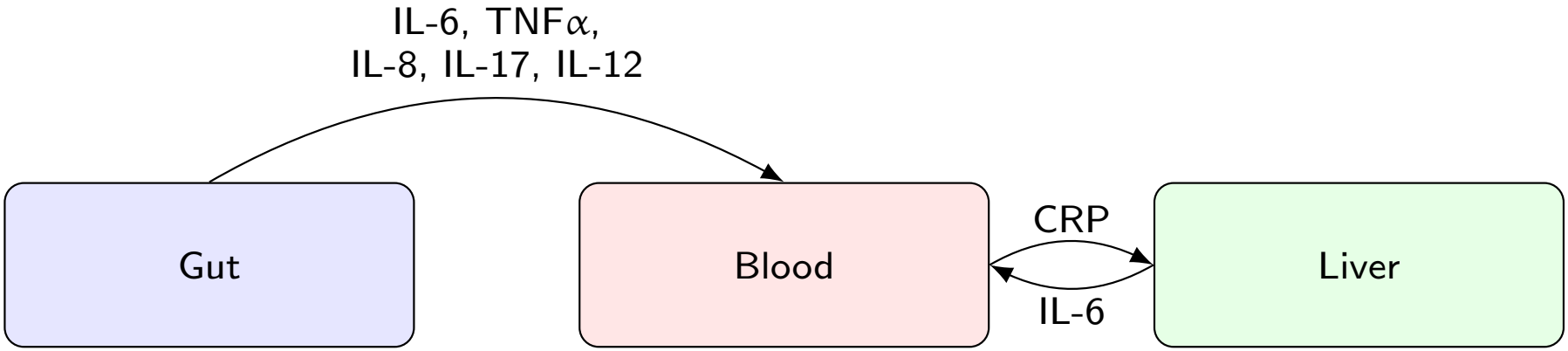
Model 2: A Hybrid Mechanistic–Stochastic Model of Gut Inflammation and Flare Risk

This work is a **systems-biology simulation** of gut inflammation (an IBD-like disease model), built in two layers:

1. **Mechanistic core :** a 47-state ordinary differential equation (ODE) model of immune cells and cytokines, plus a pharmacokinetic (PK) drug layer for four biologic therapies.

Three Body Compartments

The model tracks concentrations across three physical compartments, each with its own volume:



- **Gut**: where the local immune reaction (T-cells, macrophages, dendritic cells) unfolds.
- **Blood**: systemic spillover of cytokines and neutrophils.
- **Liver**: acute-phase response—converts circulating IL-6 into CRP, a clinical inflammation marker.

General Form of the ODEs

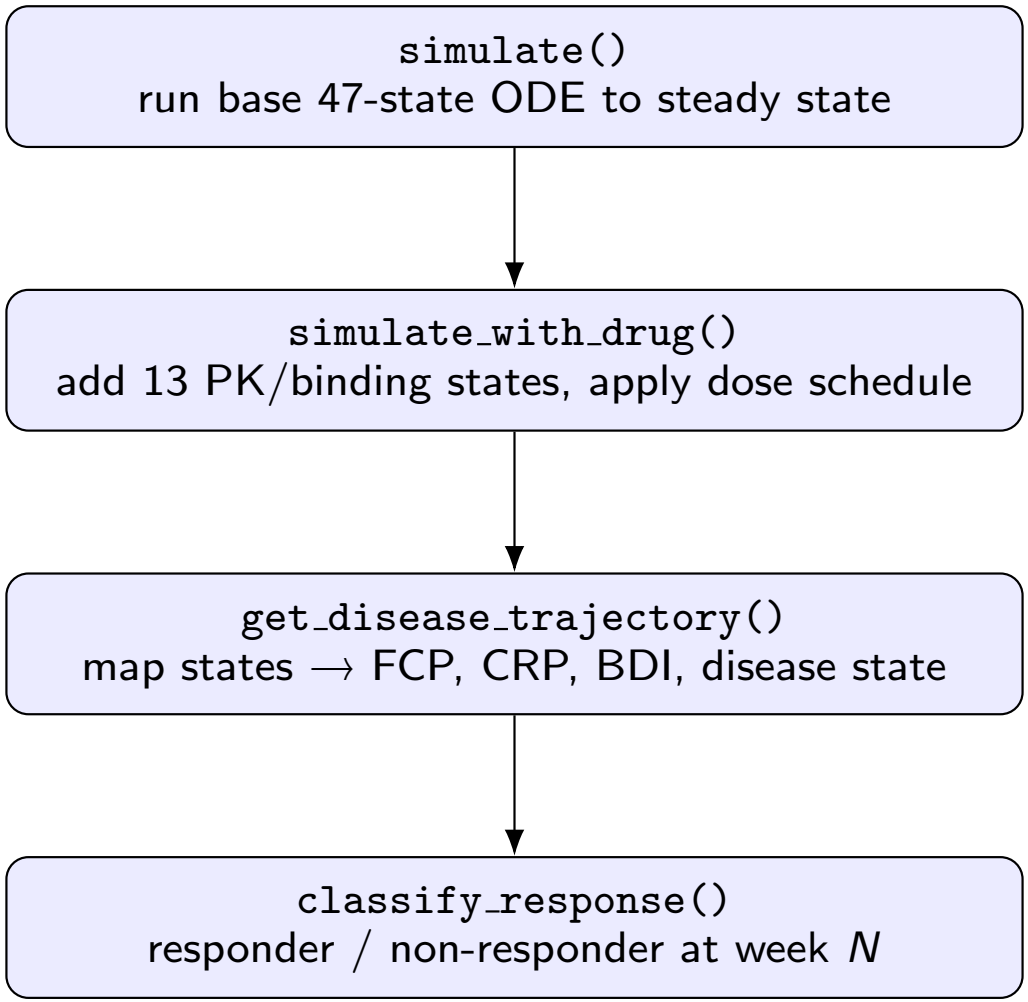
The mathematics — general form

For each species X_i, the model integrates

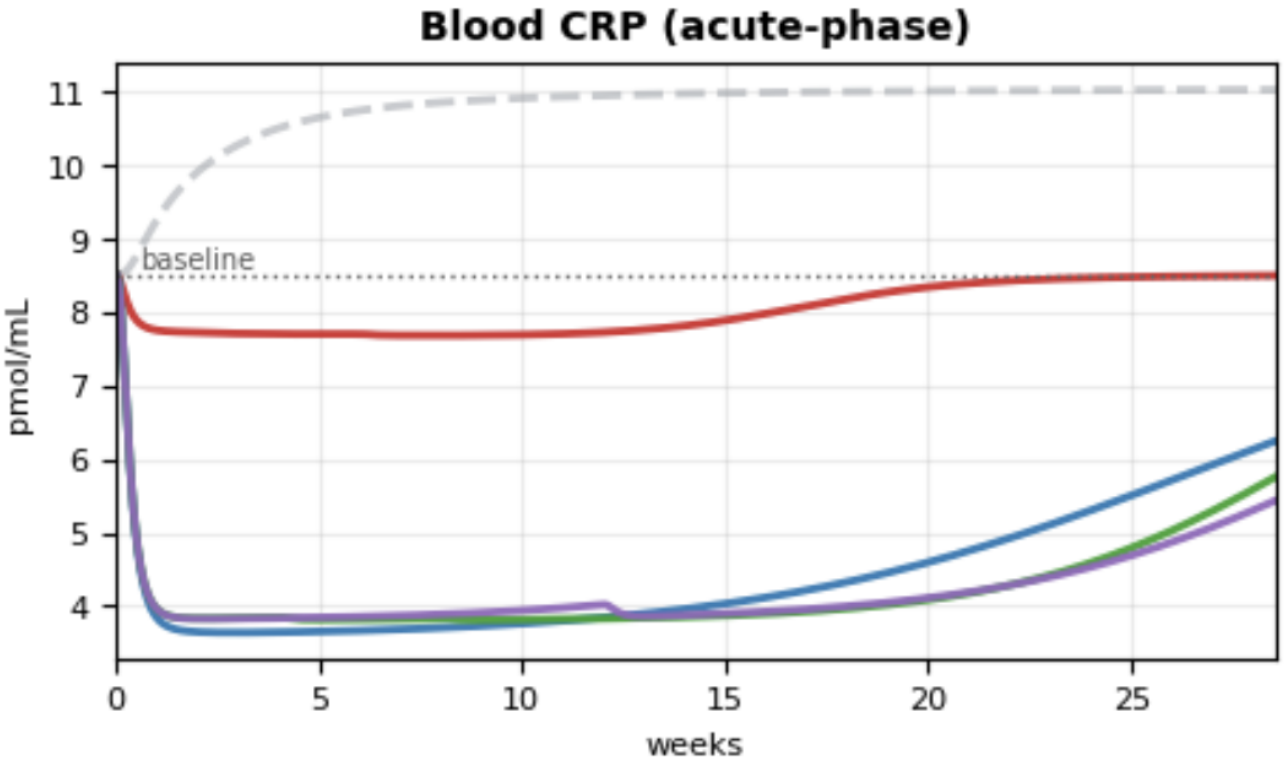
$$\frac{dX_i}{dt} = S_i(\mathbf{y}) - k_{\deg,i} X_i - \sum_j T_{ij}(\mathbf{y})$$

where $\mathbf{y} = (X_1, \ldots, X_{47})$. The term S_i represents production, $k_{\deg,i}$ is the first-order decay rate (day^{-1}), and T_{ij} describes transport between compartments.

Part 1 Summary: Pipeline

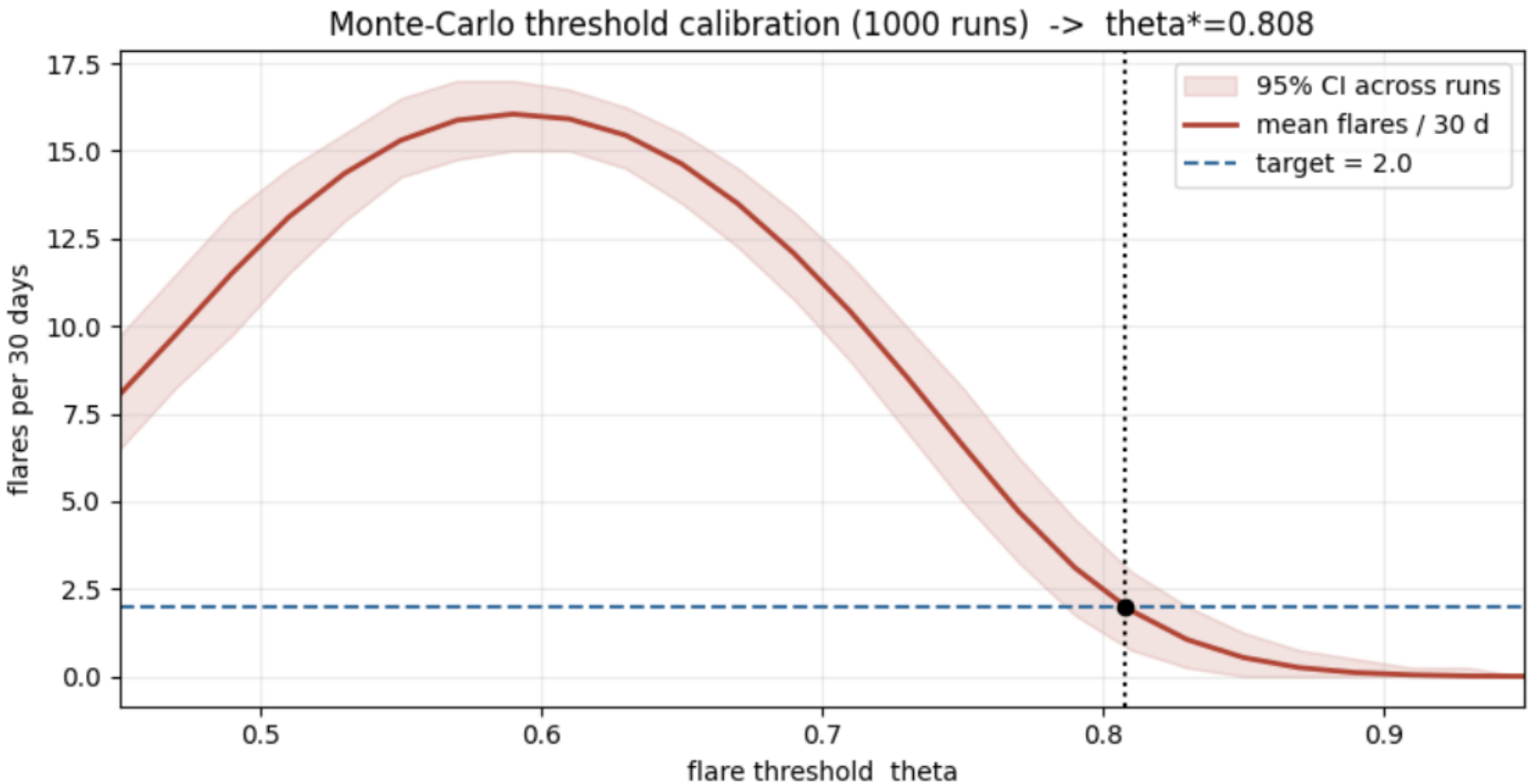


Biomarkers Vs Medication

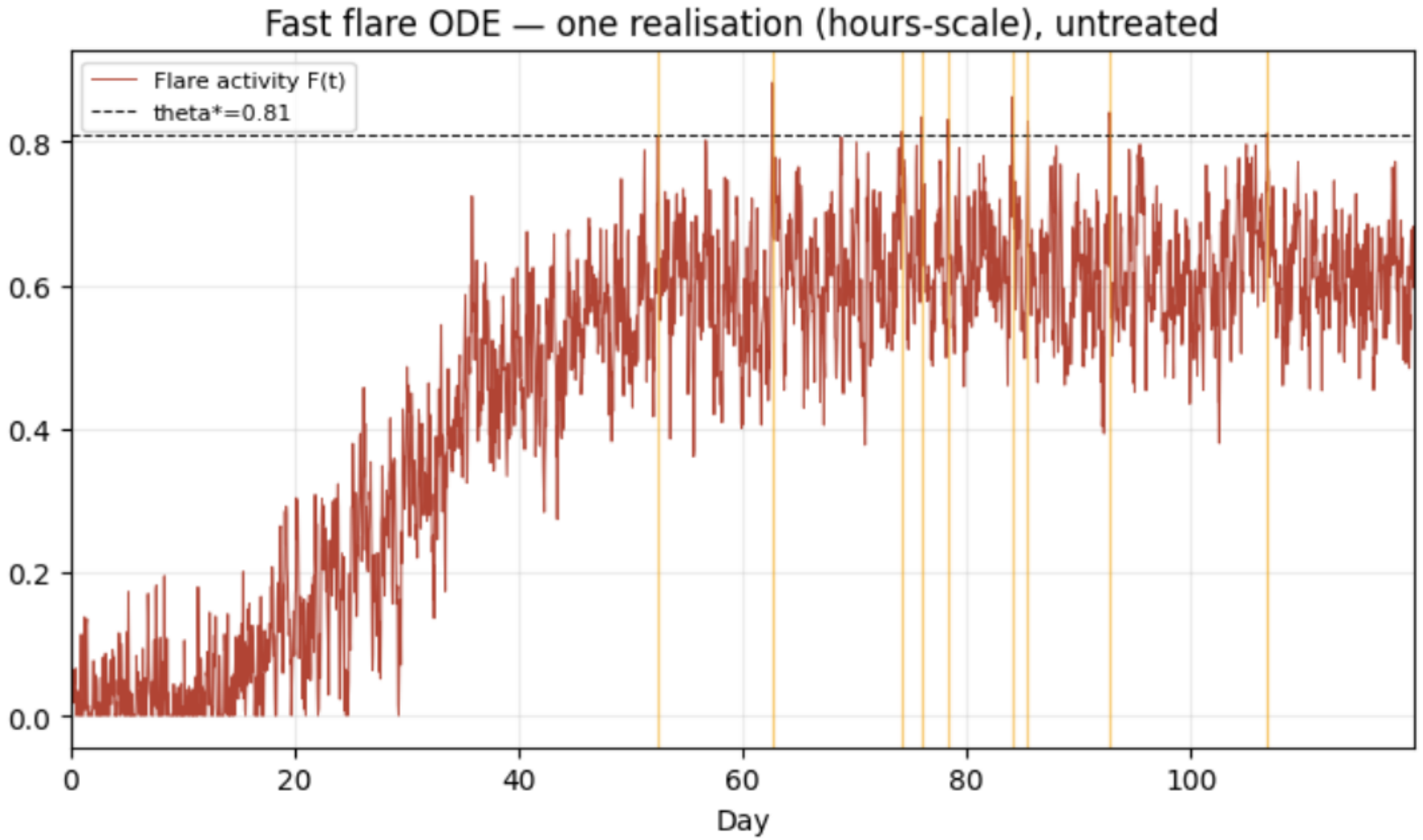


Hybrid Flare Model: Threshold for flares

Hybrid flare model (Script 2): Sits on top of the mechanistic core and converts slow biological drift into a fast, stochastic "flare" process with a self-calibrated risk threshold.



Flares


Fast flare ODE — one realisation (hours-scale), untreated
Flare activity F(t)
theta*=0.81
0.0
0.2
0.4
0.6
0.8
0
20
40
60
80
100
Day

Overview

Conclusion - Goals Revisited

1. **Literature study**: We were able to identify and compare existing models in the literature and use them as a starting point for our work.

2. **Formulate and propose models**: We created mechanistic models of increasing complexity to predict disease load and symptom flares, a mechanistic model for describing disease degradation, and ML models to feed patient-specific parameters into those mechanistic models.

3. **Assess input parameters and Identify sensitive model parameters and evaluate plausibility**: We have created the mechanisms for evaluating the model sensitivity to particular inputs.
4. **Real results hinge on real data** ...but we did make a sturdy scaffold to pin them on.

Conclusion - Future Directions

Simulate models with Vironix patient data:

1. Test and validate parameter estimation methods for the hybrid approach
2. Create update process for incorporating new data as it is collected
3. Personalized treatment functions based on the individuals
4. Convolutional Neural Net (CNN/GCN) models to predict flares based on time series data

Incorporate more mechanisms at multiple scales.

Thank you!

Acknowledgments

We thank the SIAM Mathematical Problems in Industry Workshop organizing committee for facilitating this modeling experience. We are especially grateful to our mentor, **Walter Kiefer** for his guidance, feedback, and support throughout the development of this work.

Appendix - Disease Degradation Model with Stochastic Load

Stochastic adverse exposure

$$A(t) = \mu_A + \sigma_A Z(t), \qquad Z(t) = \rho_A Z(t-1) + \sqrt{1-\rho_A^2}\, w_t, \quad w_t \sim \mathcal{N}(0,1)$$

Disease dynamics

$$\frac{dx}{dt} = a\,(A + \rho d_s)\,(1-x) - (b + c_s T_s + c_\ell T_\ell)\,x,$$

$$\frac{dg}{dt} = kx(1-g) - (r + \gamma_s T_s)\,g,$$

$$\frac{dd_a}{dt} = \varepsilon\,[m(g-g_c)_+(1-d) - \eta T_\ell d_a - \phi d_a],$$

$$\frac{dd_s}{dt} = \varepsilon\phi d_a.$$

Treatment

$$T_s = \delta g\,\sigma\left(\frac{g-g_c}{s_w}\right), \qquad T_\ell = T_{\text{constant}}$$

$$d = d_a + d_s, \qquad (g-g_c)_+ = \max(g-g_c, 0).$$