GENERATING CANDIDATE OPTIMAL INDIVIDUALIZED DOSING STRATEGIES

Erica E M Moodie

Department of Epidemiology, Biostatistics, & Occ. Health
McGill University
erica.moodie@mcgill.ca

Joint work with Benjamin Rich & David A Stephens

McGill

Optimal dose-finding

- In a pharmacological treatment setting, we often wish to consider adapting treatment dosage to individual patient profiles when
  - there is heterogeneity in patient response;
  - over-treating can lead to side-effects, treatment fatigue (poor compliance), and higher costs;
  - under-treating can lead to poorer patient outcomes.
- We aim to balance *efficacy* and *tolerability*.

Warfarin

- Warfarin is a highly effective and frequently-prescribed anti-coagulant that works to decrease the risk of thrombosis by depleting the body's active vitamin K.
- The impact of warfarin varies considerably between and within individuals, as:
  - dietary choices can replenish the vitamin K,
  - warfarin interacts with a variety of other medications,
  - there are known genetic variants that increase the risk of thrombosis.
- The appropriate dose to achieve a clotting time in the target therapeutic range can vary by $>$5-fold between individuals.

Warfarin

- Following an initial dose, warfarin takes about one day to show anti-coagulant effects; the duration of the effect from a single dose lasts 2-5 days.
- Clotting time must be monitored regularly; it is measured using the international normalized ratio (INR), which should generally lie between 2 and 3.
  - High INR: increased risk of bleeding.
  - Low INR: insufficient protection against thrombosis.
- Inappropriate dosing is a major cause of emergency hospitalizations resulting from adverse drug events.

Finding candidate ATSs

- Generating candidate optimal dosing strategies could be done using observational data or sequential multiple assignment randomized trials (SMARTs).
- The problem?
  - Observational data may be subject to unmeasured confounding.
  - SMARTs can be expensive, in terms of (i) participants, (ii) money, and (iii) time.
- What alternatives exist? Could we use the well-understood biological actions and effects of a drug such as warfarin to create realistic simulations that would suggest a small number of candidate rules?

Aim

- Our aim is to use methods proposed for finding adaptive treatment strategies (ATS) to suggest an optimal dosing strategy for a continuous-valued treatment.

- Challenges:
  - The true model for outcome as a function of covariates is complex and unknown.
  - Efficacy and side-effects must be balanced.
  - Simulation protocols for ATSs have focused on overly-simple settings.

- We draw on the pharmacokinetics (PK) and pharmacodynamics (PD) literature to determine a realistic data-generating algorithm in which the true form of the optimal dosing regime is not known.

Notation

- We consider K treatment intervals.
- Pre-treatment covariates are denoted L_j, for j=0 (baseline) up to $K-1$.
- Treatments (doses) are denoted A_j, $j = 1, ..., K$.
- Treatment and covariate history at the start of the j-th interval is denoted $H_j = \{L_0, A_1, L_1, ..., L_j\}$.
- Parentheses are used to denote potential outcomes, e.g. $Y(0_1, ..., 0_j, d^{\text{opt}}_{j+1}, ..., d^{\text{opt}}_K)$ is the outcome that would be observed under no treatment up to the j-th interval, followed by optimal treatment.

Q-learning: background

- Q-learning is a popular, regression-based approach for estimating ATS.
- Define the Quality of Treatment, Q-functions:

$$Q_2(h_2, a_2) = E[Y | H_2 = h_2, A_2 = a_2],$$

$$Q_1(h_1) = E\left[\max_{a_2} Q_2(H_2, a_2) \middle| H_1 = h_1, A_1 = a_1\right].$$

- The optimal ATS is then

$$d_j(h_j) = \arg\max_{a_j} Q_j(h_j, a_j), \quad j = 1, 2.$$

Q-learning: typical implementation

Model for Q-functions:

$$Q_j(h_j, a_j) = \beta_j^T S_j' + (\psi_j^T S_j) A_j$$

where S_j' and S_j are two vector summaries of H_j. Set $j = K$.

1. Stage j: regress Y on S_j' and $S_j A_j$ to obtain $(\hat{\beta}_j, \hat{\psi}_j)$.
2. Set the stage-(j-1) pseudo-outcome to $\max_{a_j} Q_j(h_j, a_j)$, estimated by
$$\tilde{Y}_{j-1} = \hat{\beta}_j^T s_j' + |\hat{\psi}_j^T s_j|.$$
3. Stage j-1: regress $\tilde{Y}_{j-1}$ on S_{j-1}' and $S_{j-1} A_{j-1}$ to obtain $(\hat{\beta}_{j-1}, \hat{\psi}_{j-1})$.
4. Repeat steps 1-3 until $(\hat{\beta}_1, \hat{\psi}_1)$ is found.

Estimated optimal ATS: $\hat{d}_j(h_j) = \arg\max_{a_j} Q_j(h_j, a_j; \hat{\beta}_j, \hat{\psi}_j)$.

Q-learning

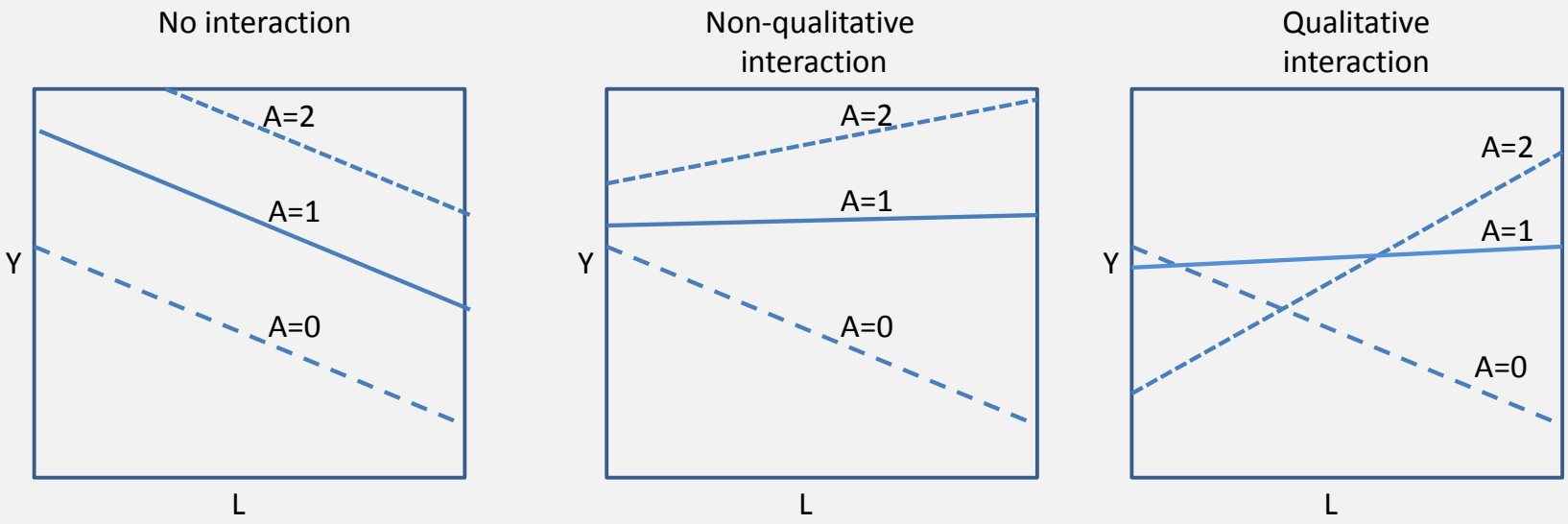
- In Q-learning, we may distinguish between two components of the model, $E[Y(a_1, ..., a_j, d_{j+1}^{\text{opt}}, ..., d_K^{\text{opt}})|H_j]$:
  - the contrast (or blip, or SNM) model, describing how the receipt of A_j affects outcome (both its main effect and any interactions with other covariates), denoted $\gamma_j(a, h_j)$; and
  - the main effects of all other covariates on outcome, which is in effect a model for

$$E[Y(a_1, ..., a_{j-1}, 0_j, d_{j+1}^{\text{opt}}, ..., d_K^{\text{opt}})|H_j],$$

sometimes called the expected counterfactual model.

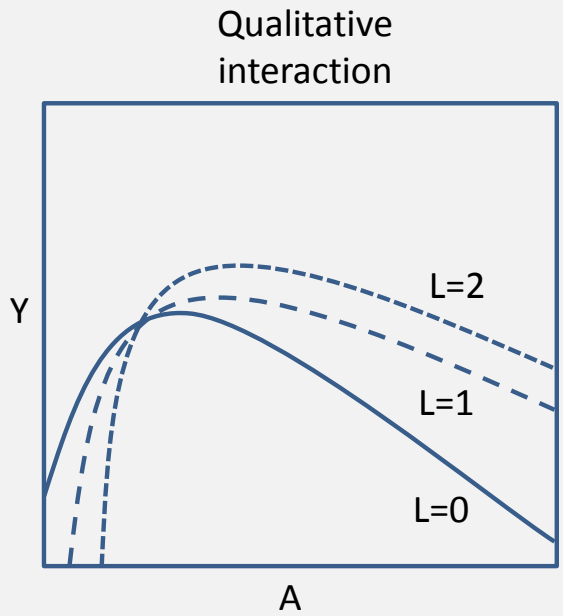
Q-learning

- The contrast function captures how doses should be tailored, which depends on qualitative interactions:



Q-learning

- The contrast function captures how doses should be tailored, which depends on qualitative interactions:



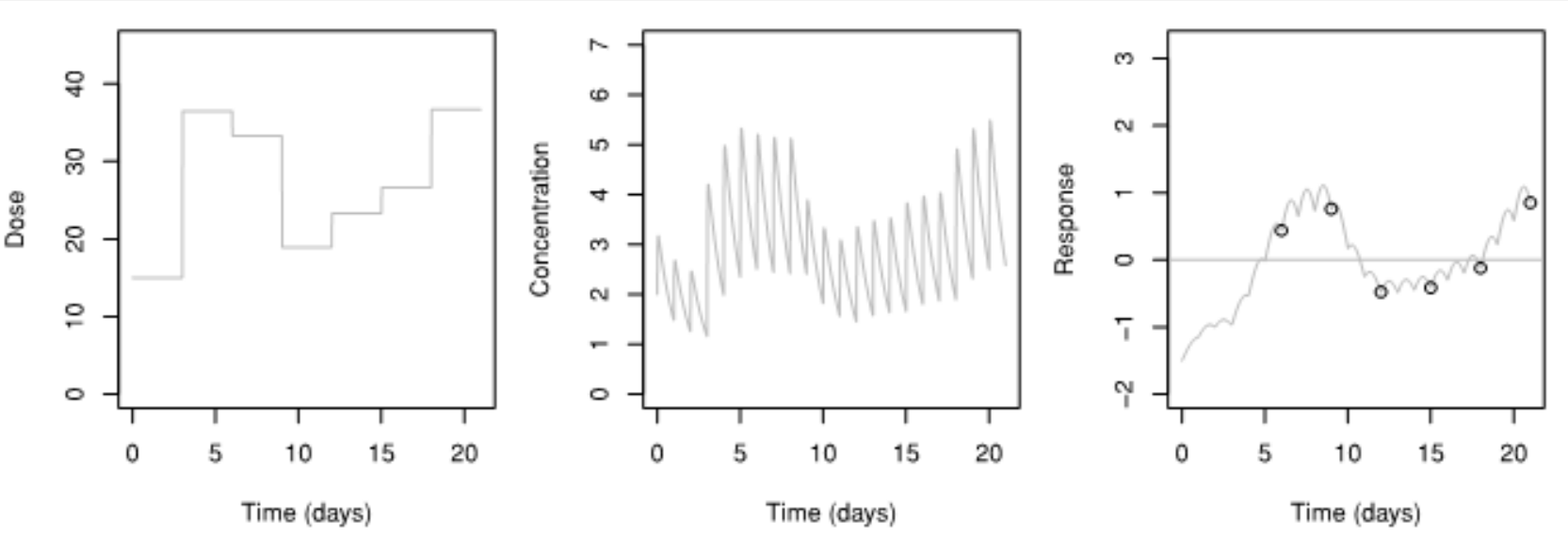
The simulation: Big picture

- The drug is taken orally once daily for 21 days.
- Doses are modified every 3 days (days 1, 4, 7, ..., 19).
- The first 6 days consist of a loading phase which serves to establish steady state conditions: this is considered part of the baseline period.
- Five 3-day treatment intervals were considered, starting on day 7.
- Thus, L_0 is the response on day 7 and A_1 to be the dose assigned on the same day. L_{-1} and A_0 denote the response and dose assigned on day 4 respectively.
- Doses in our siimulated trial are adjusted according the following rule

$$A_{ij} = -0.6L_{i(j-1)} + 0.8A_{i(j-1)} + \varepsilon_{ij},$$

$$\varepsilon_{ij} \sim \mathcal{N}(0, 0.2)$$

A sample profile



Evaluation

- We generated a single data set of size 2,000 to estimate the contrast parameters using Q-learning with
  - contrast:

$$a_j(\psi_{j0} + l_{j-1}\psi_{j1} + a_{j-1}\psi_{j2} + l_{j-2}\psi_{j3}) + a_j^2(\psi_{j4})$$

  - splines for main effects on l_{j-1}, a_{j-1}, and l_{j-2} chosen by generalized CV.
- Optimal rules were then implemented in a new population of 1,000 individuals to see how these individuals fared under the estimated optimal rule as compared to the trial protocol for allocating doses.

First findings

- Under the “trial protocol,” the quartiles of the outcome are -2.050, -1.620, and -1.230.
- Under the new regime estimated by Q-learning, the quartiles are notably higher: -1.370, -1.040, and -0.816.
- We are thus seeing a substantial improvement in outcomes, tailoring on last treatment and last INR.

- Can we do better?

Doubly-robust Q-learning

- While the PK/PD characteristics of warfarin are quite well understood, the correct model for the outcome is unknown and likely to be very complex.
- We can make use of the fact that we have designed and conducted a (simulated) trial, because the treatment allocation model is known and can be correctly specified.

- The usual Q-learning EE is $E[\tilde{Y}_j - Q_j | H_j, A_j]$, which is mean 0 when Q_j is correctly specified.
- A doubly-robust EE is

$$E\left[(\tilde{Y}_j - Q_j)(A_j - E[A_j|H_j])\lambda(H_j)|H_j, A_j\right],$$

which is mean 0 when *either* Q_j or $E[A_j|H_j]$ is correctly specified, for $\lambda(H_j)$ an analyst-specified function of the data.

G-estimation

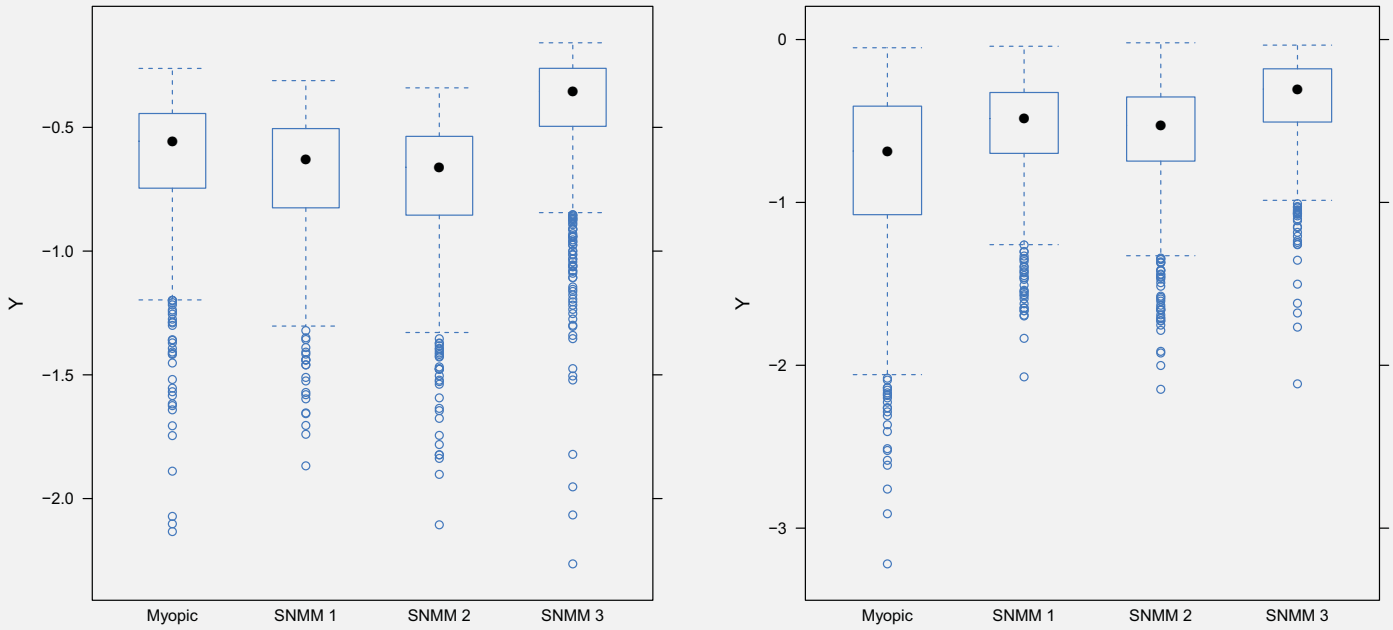
- The EE of the previous slide, it turns out, is just a G-estimation equation for binary treatment.
- G-estimation can be accomplished as a recursive series of weighted regressions – or indeed as a weighted form a Q-learning.
  - Weighting is by a function of $A_j - E[A_j|H_j]$ in the case of a binary exposure.
  - In the case of a continuous dose, weighting is by a non-linear function of dose involving, for example, $\begin{pmatrix} A_j - E[A_j|H_j] \\ A_j^2 - E[A_j^2|H_j] \end{pmatrix}$.

Contrast models evaluated

Model	$\gamma_j(a, h_j)$ for $j = 2, \ldots, 5$
SNMM 1	$a(\psi_{j0} + l_{j-1}\psi_{j1} + a_{j-1}\psi_{j2}) + a^2(\psi_{j3})$
SNMM 2	$a(\psi_{j0} + l_{j-1}\psi_{j1} + a_{j-1}\psi_{j2} + l_{j-1}a_{j-1}\psi_{j3}) + a^2(\psi_{j4})$
SNMM 3	$a(\psi_{j0} + l_{j-1}\psi_{j1} + a_{j-1}\psi_{j2} + l_{j-2}\psi_{j3}) + a^2(\psi_{j4})$

- We used a spline model with 3df on each of the two most recent INRs and the previous treatment for the expected counterfactual model.
- We also compared performance with a myopic regime, that seeks only to ensure INR remains in the therapeutic range in the *next* three day interval based only on current INR and previous dose.

Distribution of outcomes

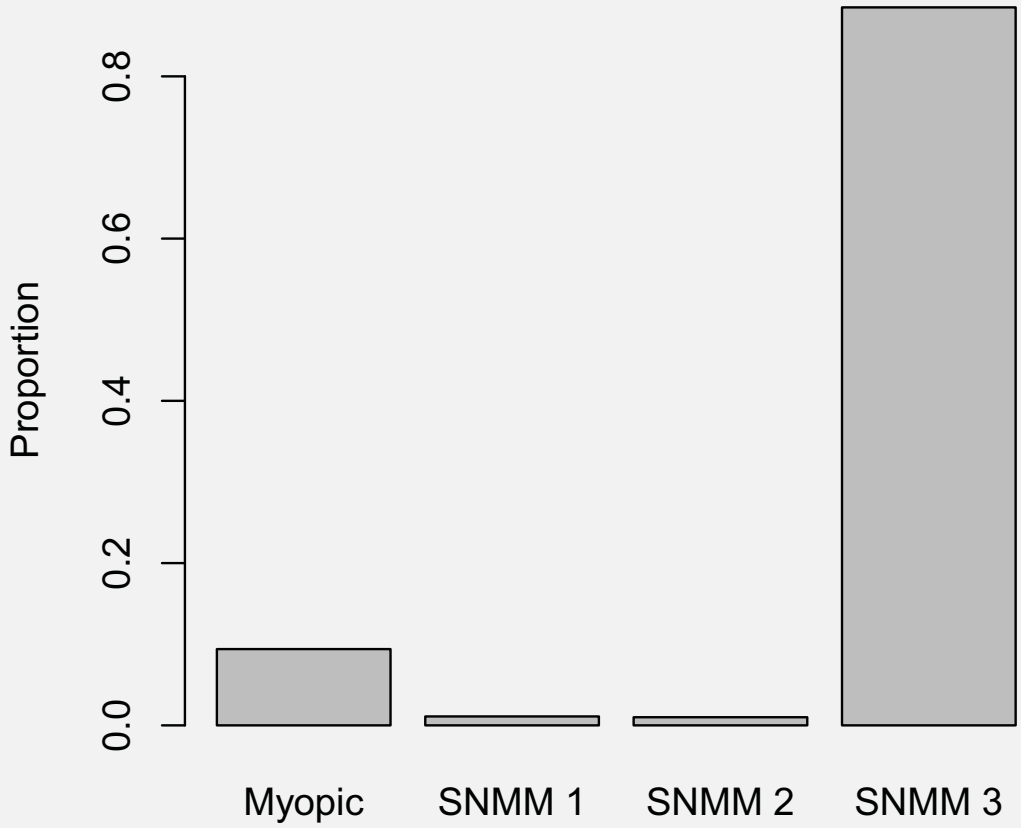


(a) 3-days between dose adjustments

(b) 1-day between dose adjustments

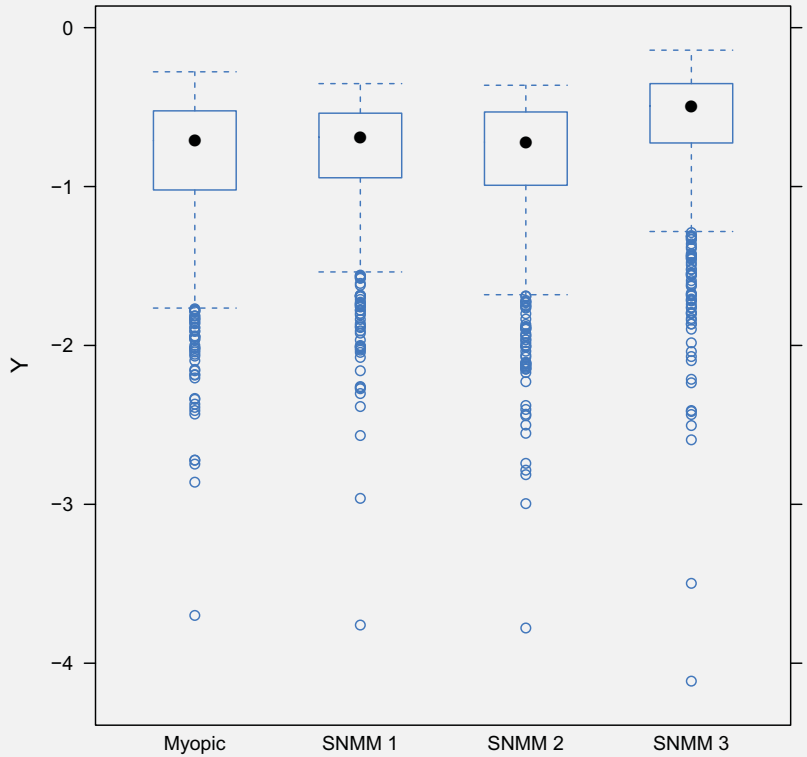
Q1 of the outcome distribution under the G-estimation ATS is larger than Q3 under the Q-learning ATS.

The best strategy by individual



Generalizing

Distribution of outcomes when the estimated regimes are applied to a population of slow metabolizers.



Concluding remarks

- We have demonstrated that we can take therapies whose PK/PD characteristics are well-understood, and simulate trials in order to suggest good ATSs.
- This approach may be particularly valuable in the setting in which treatments are measured on a continuous scale, and the true model for the outcome as a function of dose and patient history is complex.
- Following the implementation of realistic simulations as a means of suggesting candidate ATSs, these strategies could then be assessed in a confirmatory RCT.