SMART STUDIES AND THE PERSONALIZATION OF MEDICAL CARE

Erica E M Moodie

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

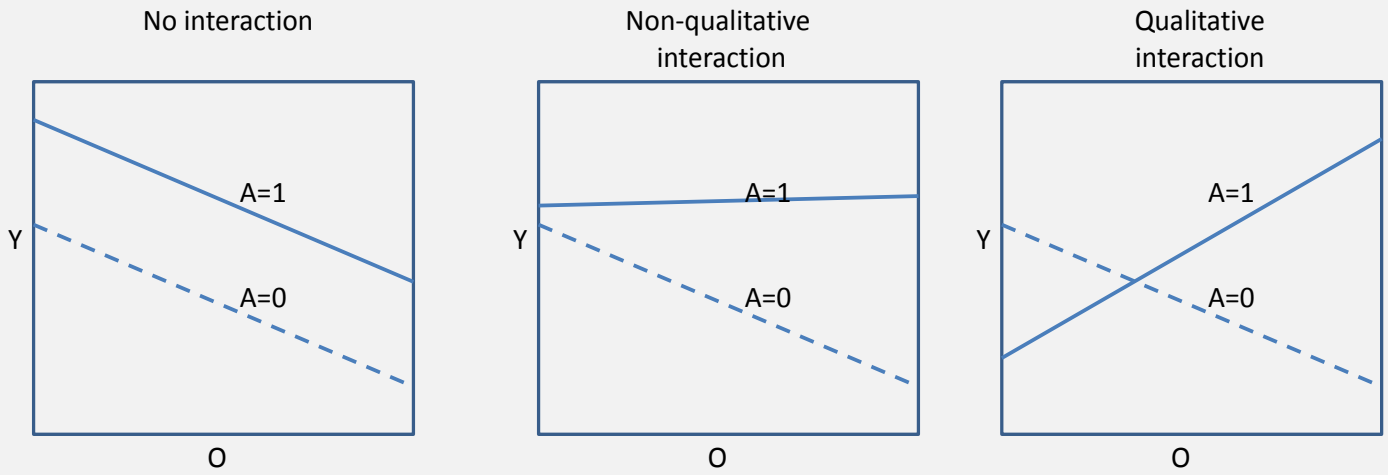
Joint work with Benjamin Rich and David A. Stephens

McGill

Personalized medicine

- A bit of a catch-all expression for tailoring treatment to the individual.
- Can refer to tailoring by genetic profile, but common in the statistics literature to use this for tailoring on more "macro" level characteristics, some of which may change over time (symptoms scores, blood pressure, side effects, etc.)
- Key idea – tailoring treatment to the patient rather than the diagnosis is better because:
  - There is heterogeneity in patient response
  - Chronic conditions where individual patient response may change over time
  - Over-treating can lead to side-effects, treatment fatigue (poor compliance), and higher costs
  - Under-treating can lead to poorer patient outcomes

When would we want treatment to be dynamic?



- We distinguish between two types of covariates: those that are **prescriptive** (tailoring variables) and those that are **predictive**; of course a variable may be neither, in which case it is irrelevant.

A dynamic treatment regimen (DTR)

- A dynamic treatment regimen is a treatment protocol that can change over time based on a subject's observed characteristics.
  - Provides a list of decision rules for how treatment should be allocated over time.
  - A function that takes covariates and treatment/response history *to the current time* as arguments and outputs an action to be taken.
  - A subject's interval-specific treatment is not known at the start of a dynamic regimen, since treatment depends on time-varying variables.
- An optimal dynamic regimen is defined to be the "best possible" regimen in the sense that it maximizes some good outcome.
- Note: DTRs are also called *adaptive treatment strategies* or *policies*.

How can we learn about treatment tailoring?

SMARTs

Erica E M Moodie

SMARTs

Our work

Simulations

Q-learning

Results

Final thoughts

- Standard randomized controlled trials (RCT) assess the effectiveness of a single dose-level of a single treatment, as compared to another.
- Estimating the sequence of actions that optimizes response in a longitudinal setting requires studying the elements in the sequence.

- Can this be accomplished by a series of single-action comparisons?

How can we learn about treatment tailoring?

- Standard randomized controlled trials (RCT) assess the effectiveness of a single dose-level of a single treatment, as compared to another.
- Estimating the sequence of actions that optimizes response in a longitudinal setting requires studying the elements in the sequence.

- Can this be accomplished by a series of single-action comparisons?

Example: Treating MDD

- Suppose we wish to compare both front-line and second-line treatment of major depressive disorder:
  - Front-line options: citalopram (Cit) or cognitive behavioural therapy (CBT)
  - Second-line options: treatment switch to Cit, CBT, or Lithium (Li)
  - All responders to first-line therapy will continue with maintenance and follow-up

Example: Treating MDD

SMARTs

Erica E M Moodie

SMARTs

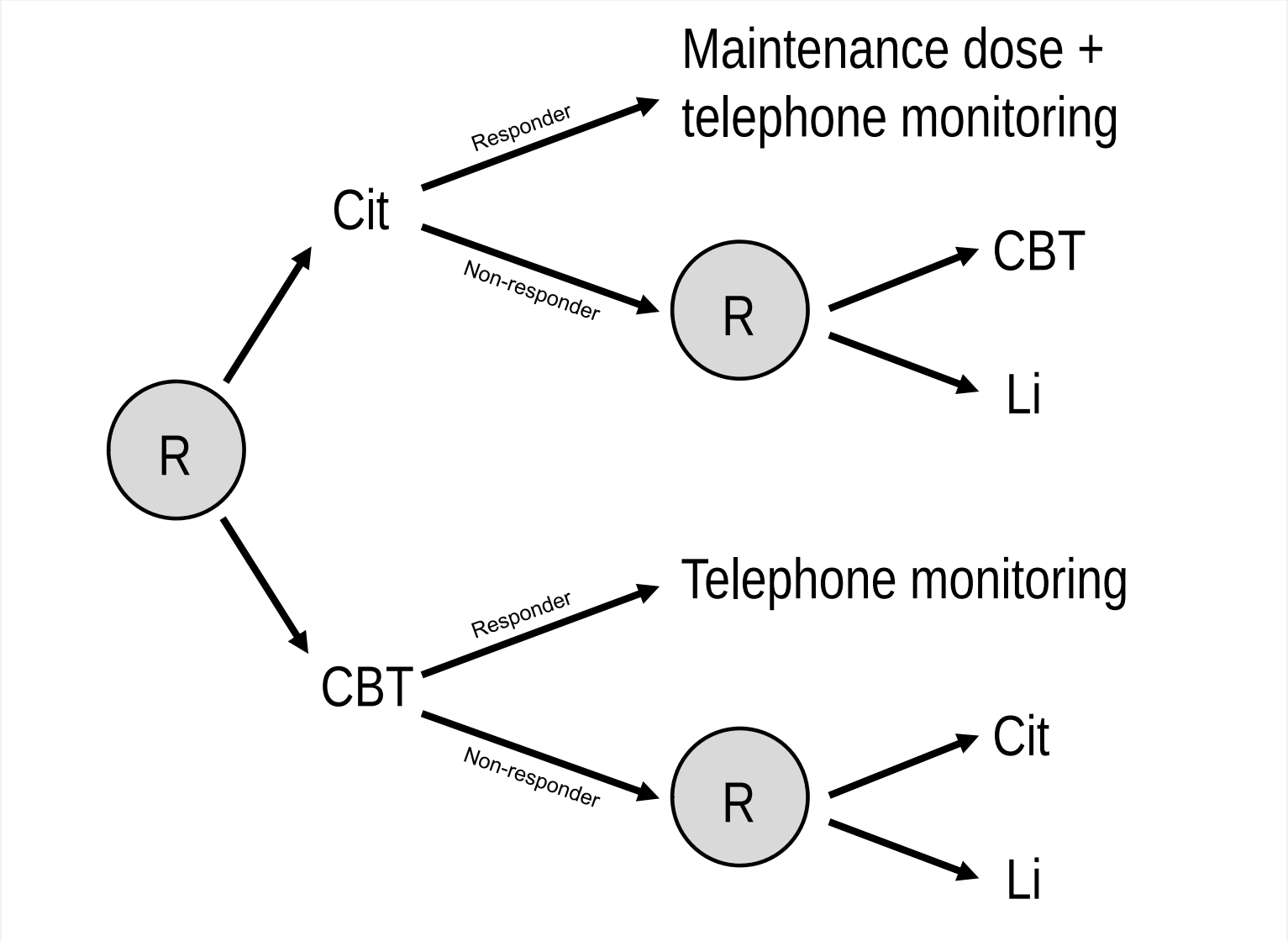
Our work

Simulations

Q-learning

Results

Final thoughts



Example: Treating MDD

SMARTs

Erica E M Moodie

SMARTs

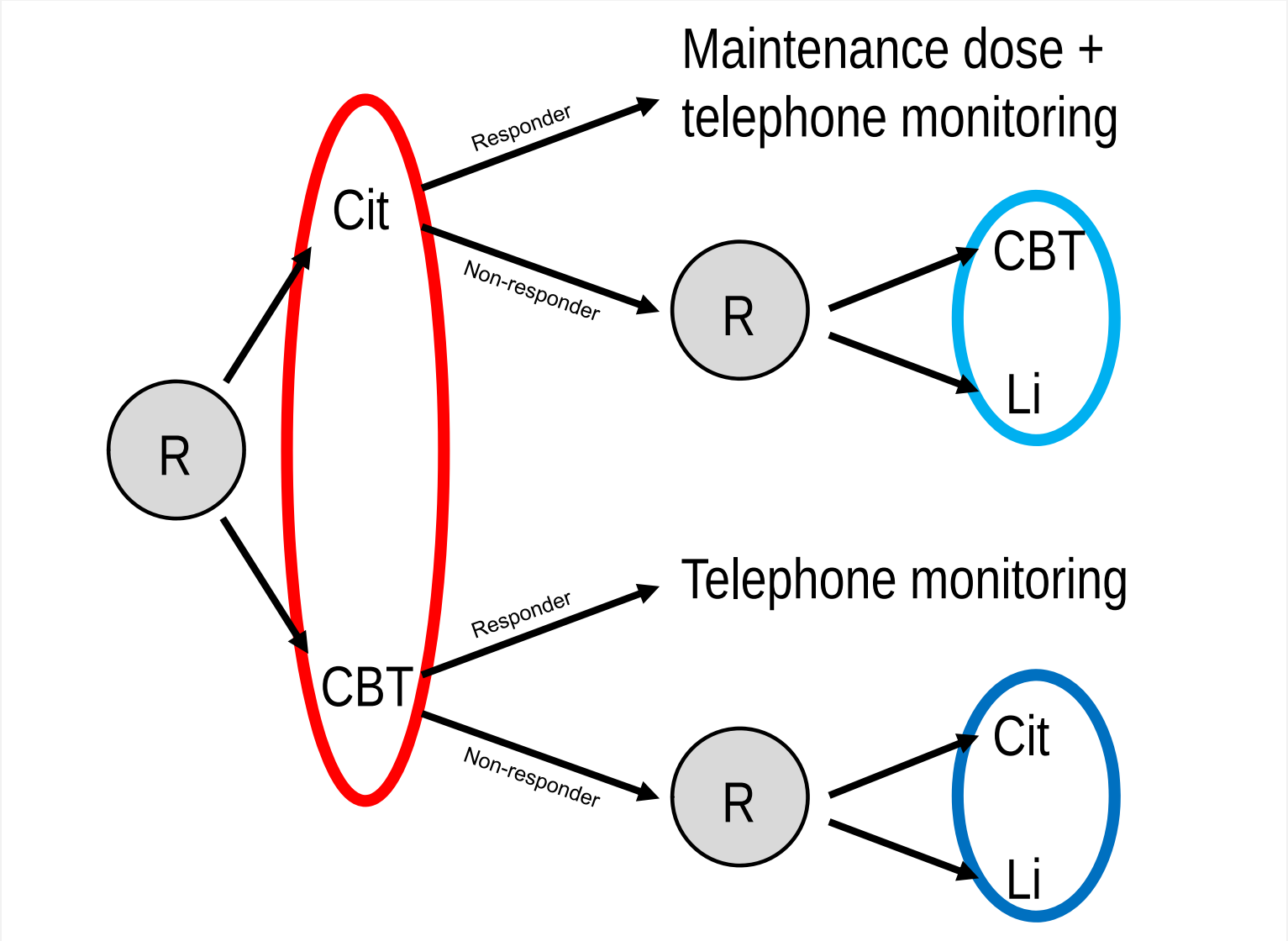
Our work

Simulations

Q-learning

Results

Final thoughts



Front-line treatment of MDD

SMARTs

Erica E M Moodie

SMARTs

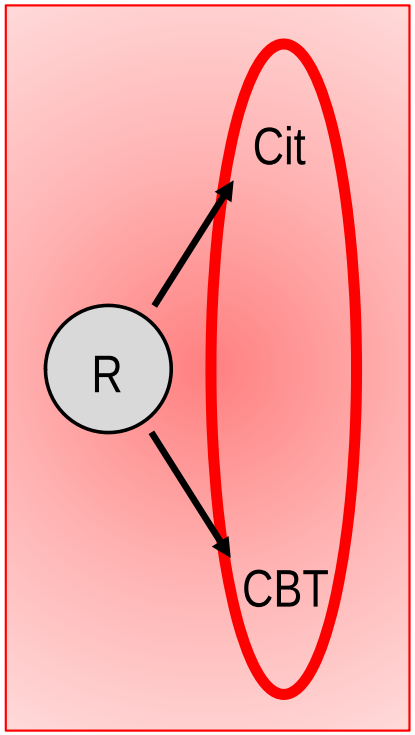
Our work

Simulations

Q-learning

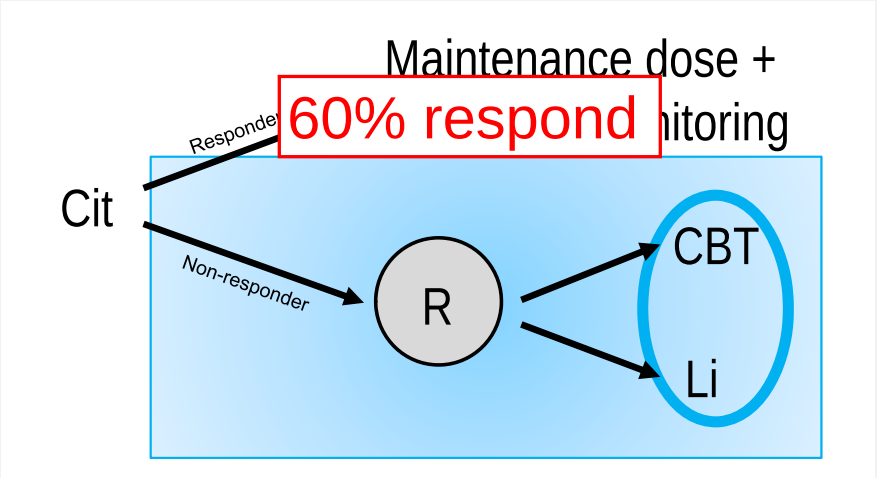
Results

Final thoughts



- Suppose we observe 60% response with Cit, and only 50% with CBT.
- Conclude: Cit is the best front-line therapy.
- Now run another one-stage trial amongst Cit non-responders.

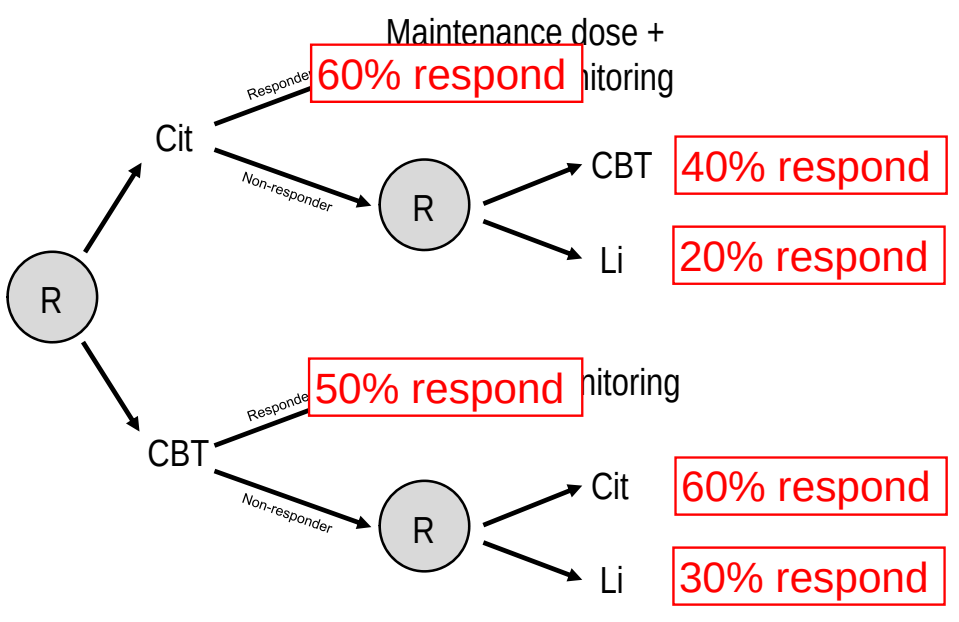
Second-line treatment of MDD



- We now observe 40% response to CBT and 20% to Li.
- Conclude: CBT is the best second-line therapy.
- Final treatment sequence: Cit followed by CBT for non-responders. Under this regimen, we expect to see 76% of patients respond.

The big 'what if'?

- What if initial treatment with CBT increases treatment adherence $\Rightarrow$ subsequent therapies more successful?



- Optimal DTR: CBT followed by Cit for non-responders; 80% response expected.

SMART!

- Two single interval trials would not have detected the best overall strategy for treatment.
- Instead, we should have used to two interval trial. These are known as SMARTs, i.e.

  sequential multiple assignment randomized trials.

SMARTs: the gains

- Ability to detect
  - delayed effects/treatment interactions
  - diagnostic effects
- More generalizable?
- Better retention?

SMARTs: the costs

SMARTs

Erica E M Moodie

SMARTs

Our work

Simulations

Q-learning

Results

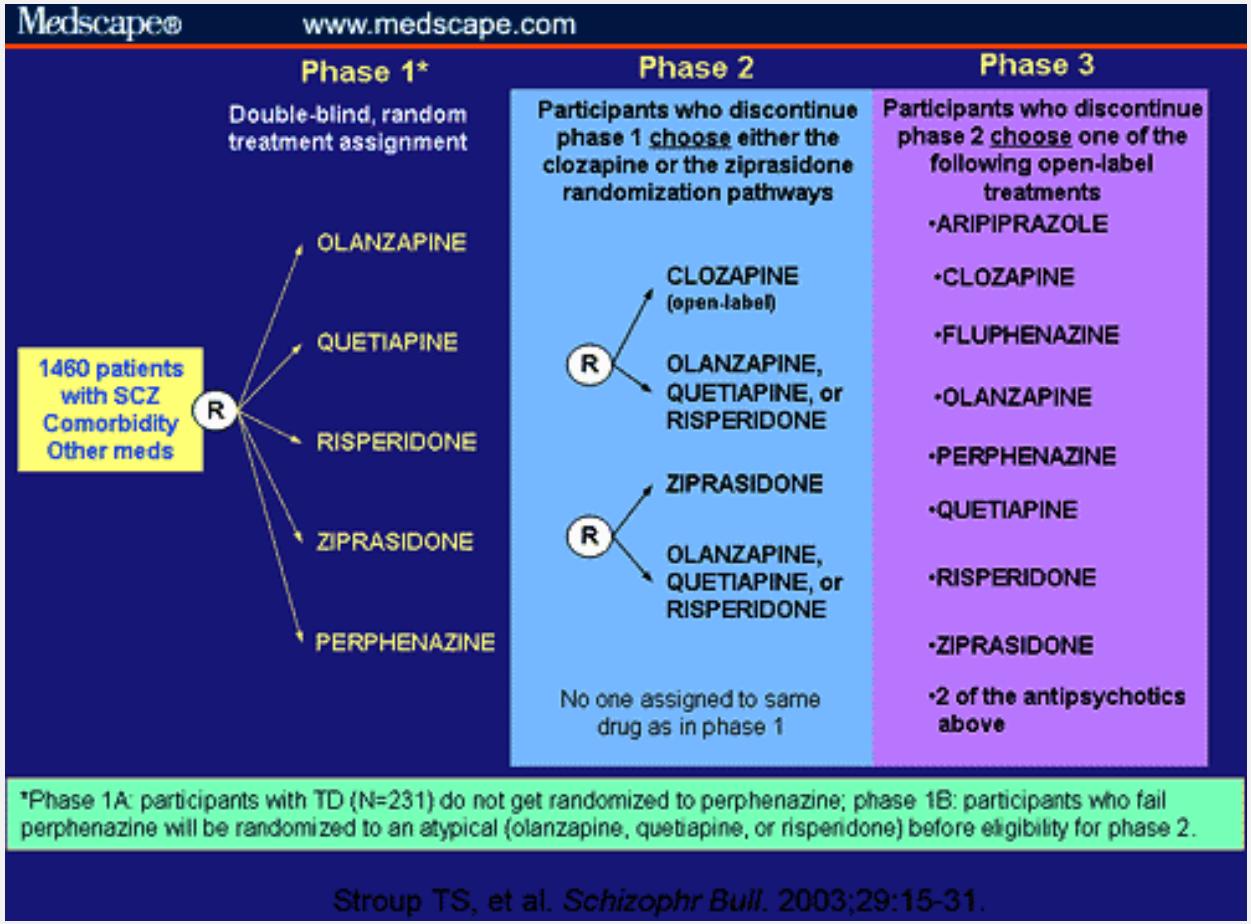
Final thoughts

- More expensive than a single-interval RCT
  - Typically need more participants
  - Longer follow-up
- More complex methods for planning and analysis: requires an experienced statistician, or one with time to learn new methods
- May require additional work to fund: still relatively new, unfamiliar

Does anyone actually use SMARTs?

SMARTs
Erica E M Moodie
SMARTs
Our work
Simulations
Q-learning
Results
Final thoughts

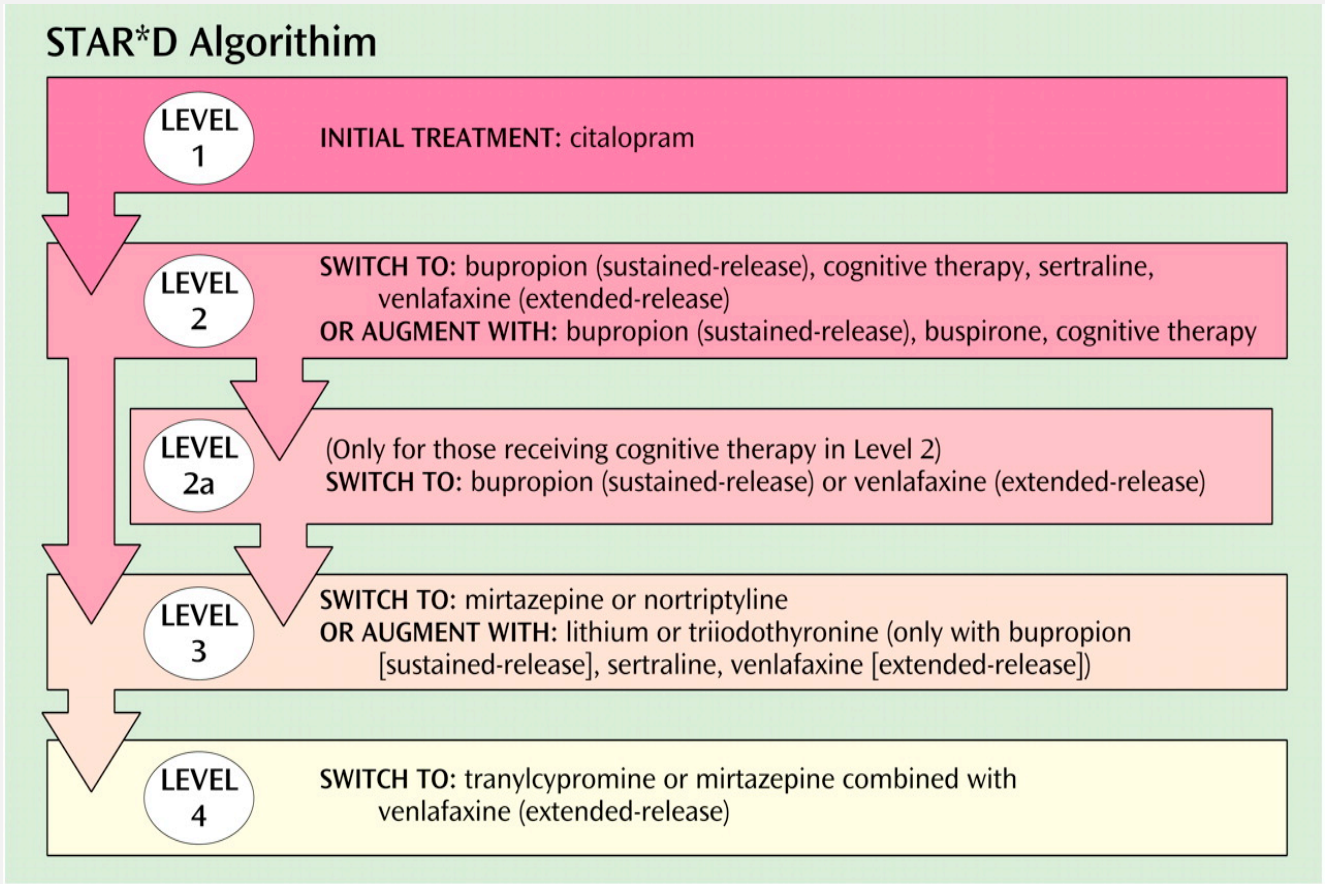
NIMH Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE) - Schizophrenia



Does anyone actually use SMARTs?

SMARTs
Erica E M Moodie
SMARTs
Our work
Simulations
Q-learning
Results
Final thoughts

Sequenced Treatment Alternatives to Relieve Depression (STAR*D)



Does anyone actually use SMARTs?

SMARTs

Erica E M Moodie

SMARTs

Our work

Simulations

Q-learning

Results

Final thoughts

Yes, but...

Does anyone actually use SMARTs?

Several newer, less ambitious trials (2 interval):

- Smoking cessation
- ADHD
- cancer (several)

Does anyone actually use SMARTs?

Yes, but primarily for:

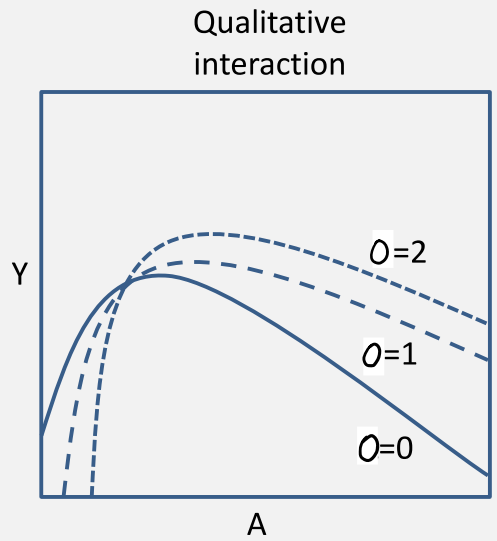
- *exploring* possible strategies (Phase II), or
- head to head comparison of a small number of 'treatment packages'

Optimal dose-finding

- In a pharmacological treatment setting, we wish to consider adapting treatment *dosage* to individual patient profiles, balancing efficacy and tolerability.
- As noted, we need to keep the possible dosing strategies small to conduct a SMART.

- The difficulty?
  - Most of the theory (and simulations) for DTR methodology has focused on binary treatments.
  - The estimating functions are necessarily more complex.

Optimal dose-finding



- **Idea:** 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 to consider in a SMART?

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.

Aim

SMARTs

Erica E M Moodie

SMARTs

Our work

Simulations

Q-learning

Results

Final thoughts

- Our aim is to use methods proposed for finding DTRs 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 tolerability must be balanced.
  - Simulation protocols for DTRs 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 regimen is not known.

Data/notation

Suppose now that we have K distinct treatment intervals. Then:

- Treatments are $A_1, A_2, ..., A_K$
- Outcome/utility Y is measured at some point after the last treatment interval, i.e. after A_K has been given
- Pre-treatment covariates O_{k-1} are measured at each interval
- Treatment and covariate history at the start of the j-th interval is denoted $H_j = \{O_0, A_1, O_1, ..., O_{j-1}\}$.
- 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.

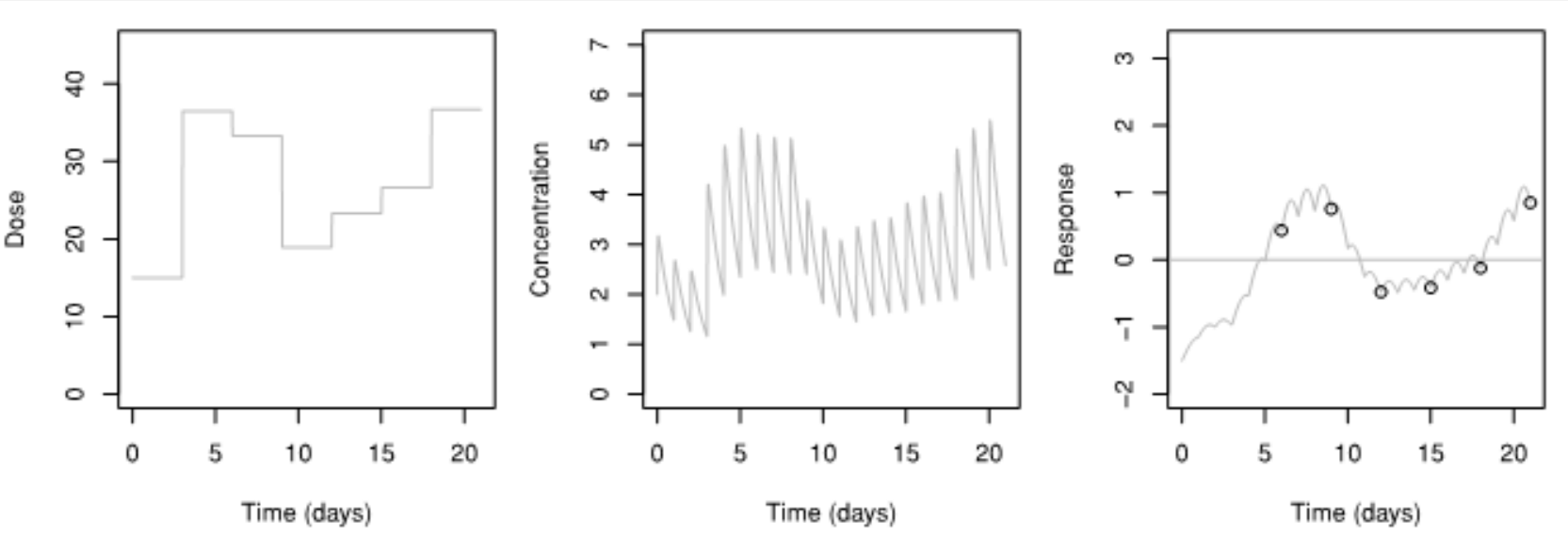
The simulation: Big picture

- 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, O_0 is the response on day 7 and A_1 to be the dose assigned on the same day. O_{-1} and A_0 denote the response and dose assigned on day 4 respectively.
- Doses in our simulated trial are adjusted according the following rule

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

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

A sample profile



Q-learning – two intervals

- Q-learning is a popular, regression-based approach for estimating DTRs.
- 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 DTR 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:

$$Q_j(h_j, a_j) = \beta_j^T H_j^M + (\psi_j^T H_j^C) A_j$$

where H_j^M and H_j^C are two vector summaries of H_j.

1. Stage 2: use OLS to regress Y on H_2^M and $H_2^C A_2$, obtaining $(\hat{\beta}_2, \hat{\psi}_2)$.
2. Set the stage-1 pseudo-outcome to $\max_{a_2} Q_2(h_2, a_2)$.
3. Stage 1: use OLS to regress $\tilde{Y}_1$ on H_1^M and $H_1^C A_1$, obtaining $(\hat{\beta}_1, \hat{\psi}_1)$.

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

Q-learning: editorial comments

- Q-learning is appealing because it is easy to implement in standard software, and easy to explain to collaborators who may be non-quantitative provided they understand the basics of regression.
- Q-learning also works nicely with continuous treatments (doses).
- The approach has several limitations, e.g.:
  - Q-learning is not robust to model mis-specification.
  - Only limited results are available for discrete outcomes.
- More sophisticated approaches exist, at least for binary treatment options.

Evaluation

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

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

  - splines for main effects on o_{j-1}, a_{j-1}, and o_{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 regimen 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

SMARTs

Erica E M Moodie

SMARTs

Our work

Simulations

Q-learning

Results

Final thoughts

- 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.
- Can make use of the fact that we have designed and conducted a (simulated) trial: treatment allocation model is known.

- 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:
  - 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,

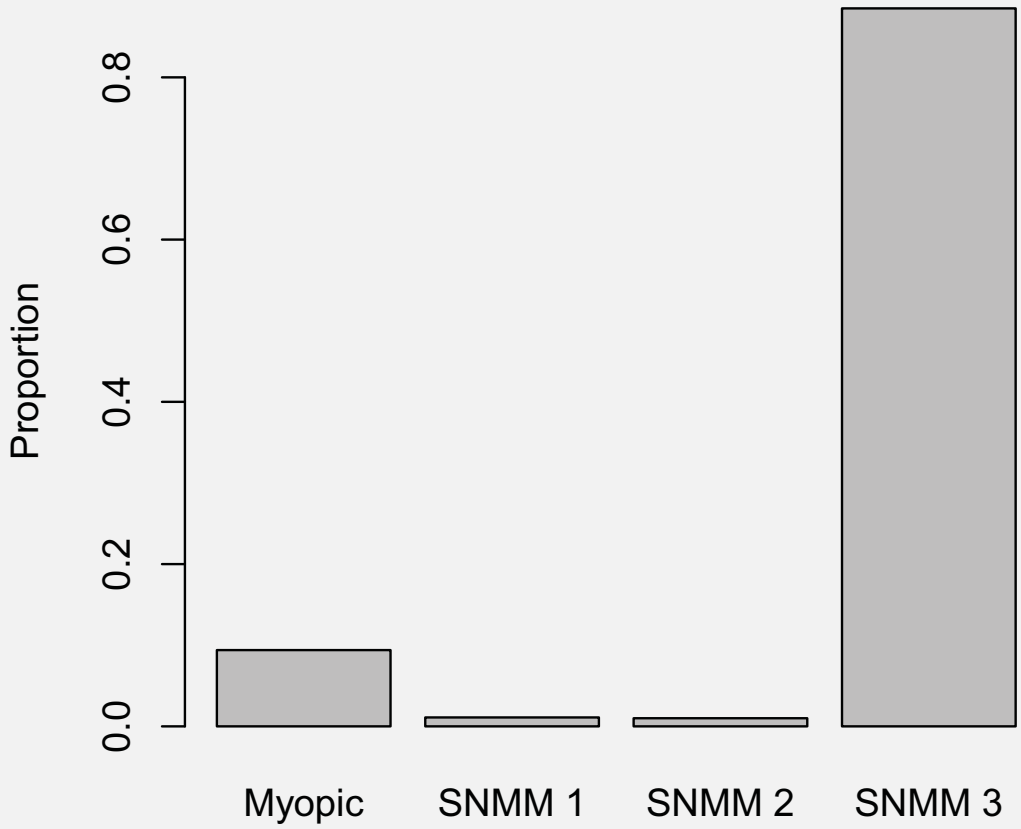
$$\begin{pmatrix} A_j - E[A_j|H_j] \\ A_j^2 - E[A_j^2|H_j] \end{pmatrix}.$$

Contrast models evaluated

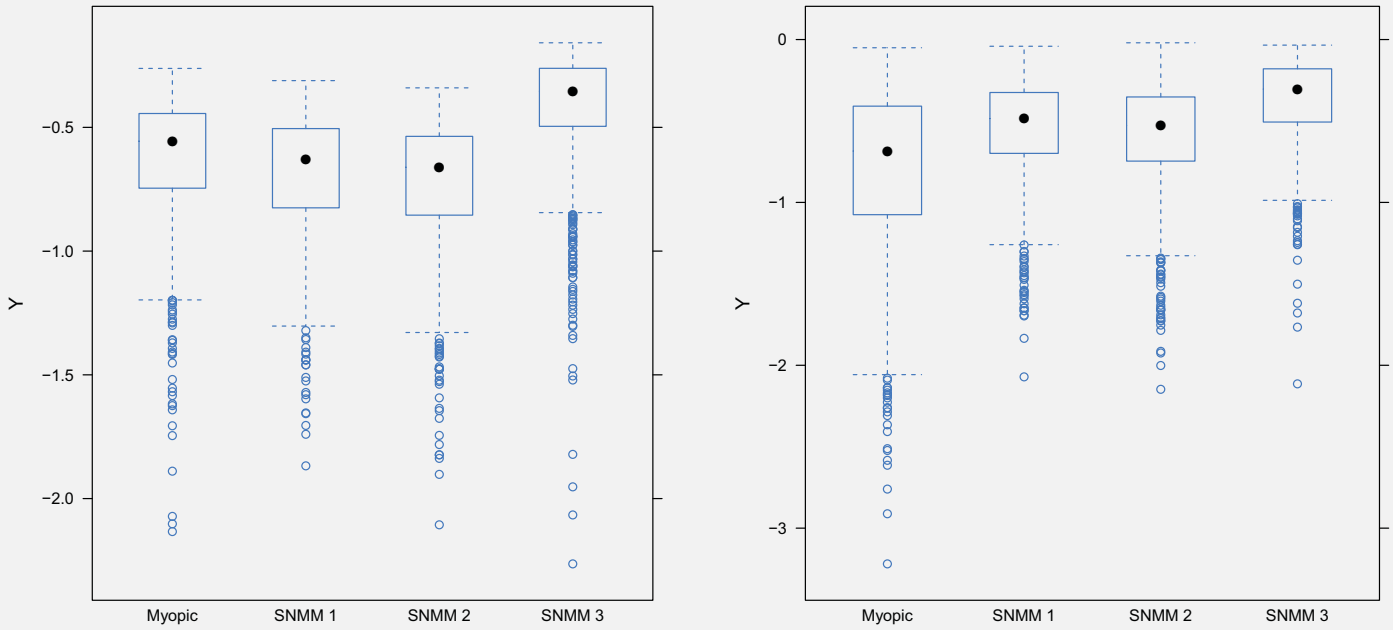
Model	$\gamma_j(a, h_j)$ for $j = 2, \ldots, 5$
SNMM 1	$a(\psi_{j0} + o_{j-1}\psi_{j1} + a_{j-1}\psi_{j2}) + a^2(\psi_{j3})$
SNMM 2	$a(\psi_{j0} + o_{j-1}\psi_{j1} + a_{j-1}\psi_{j2} + o_{j-1}a_{j-1}\psi_{j3}) + a^2(\psi_{j4})$
SNMM 3	$a(\psi_{j0} + o_{j-1}\psi_{j1} + a_{j-1}\psi_{j2} + o_{j-2}\psi_{j3})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

- We also compared performance:
  - with a *myopic* regimen (seeks only to ensure INR remains in the therapeutic range in the *next* interval based only on current INR and previous dose);
  - in a *different* population where adjustments were made daily rather than every 3 days; and
  - in a population with different PK characteristics.

The best strategy by individual



Distribution of outcomes



(a) 3-days between dose adjustments

(b) 1-day between dose adjustments

Q1 of the outcome distribution under the G-estimation DTR is larger than Q3 under the Q-learning DTR ($\approx$-0.8).

Next steps

Current work involves two significant steps to increasing the realism of the simulations to better mimic the sort of data we might actually be able to collect in a non-experimental setting:

- parameter sharing
  - computational issues: average interval-specific parameters or estimate in a single equation?
- irregular visits
  - how to define proportion of time out of therapeutic range if patients skip visits where they feel well?

Concluding remarks

- SMARTs are essential when evaluating treatment sequences.
- SMARTs are expensive, and should therefore be planned with great care.
- For therapies whose PK/PD characteristics are well-understood, we can simulate trials in order to suggest good DTRs prior to conducting a SMART.
- We hope to further develop this to test the adequacy of modelling approaches under different realistic assumptions about non-experimental data to learn about DTRs from electronic medical records.