Estimating Optimal Dynamic Treatment Regimes from Clustered Data

Bibhas Chakraborty
Department of Biostatistics, Columbia University
bc2425@columbia.edu

Society for Clinical Trials Annual Meetings
Boston, May 22, 2013

Acknowledgements

- Joint work with:
  - Guqian Du (Student @ Columbia Biostatistics)
  - Ken Cheung (Columbia Biostatistics)
  - Myunghee Cho Paik (Seoul National University)
- Supported by:
  - NIH grant R01 NS072127-01A1
  - Calderone Research Prize for Junior faculty (Mailman School of Public Health, Columbia University)

The Context

- Stroke prevention through optimal dynamic behavioral intervention
- Many behavioral interventions are delivered in group settings (e.g. family, clinic, classroom, ...)
  - This leads to clustered (outcome) data
- Overarching Methodological Questions:
  - How do we best modify the framework for SMART design to adapt to the group setting?
  - How do we best modify the existing estimation techniques to account for clustering?

A Motivating Study

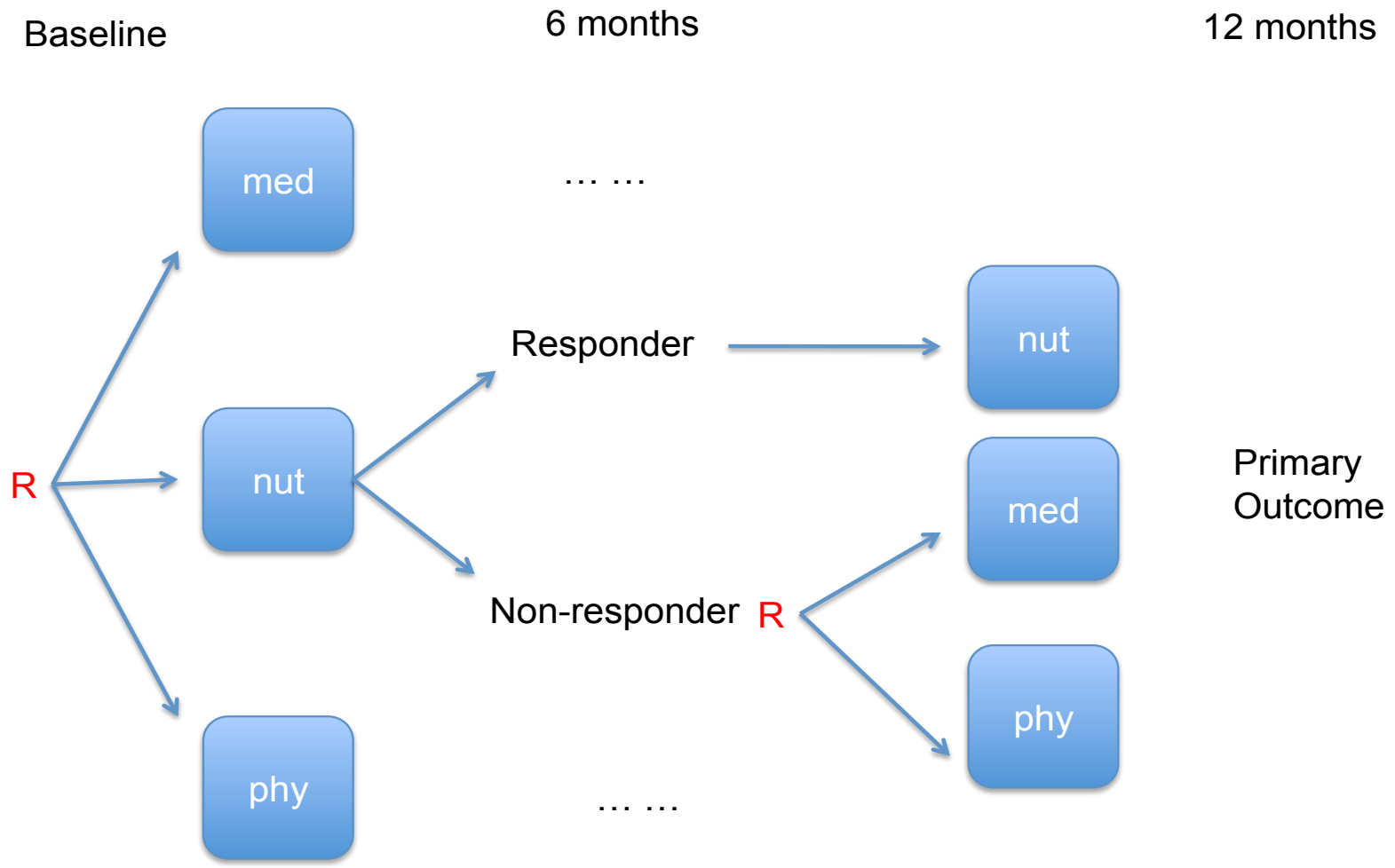
- Family-based behavioral intervention study in the Northern Manhattan area of NYC
  - Still in planning phase; no data available
- Ultimate goal: Prevent 2nd stroke
- Operative objective: Reduce blood pressure (over 12 months)
- Interventions: Behavioral modifications through 3 educational components over 12 months
  - Medication adherence (med)
  - Nutrition modifications (nut)
  - Physical activity reinforcement (phy)

Outline

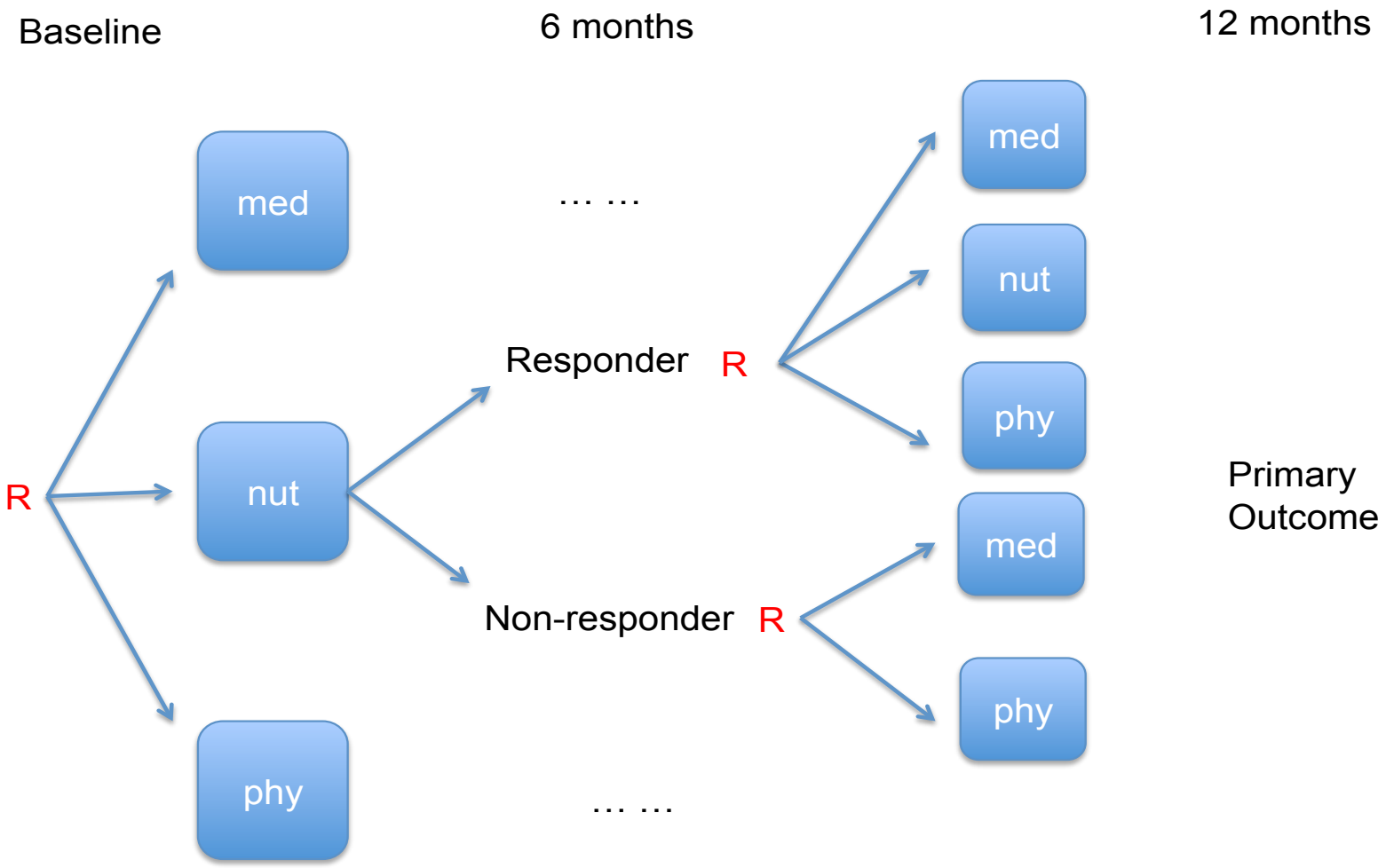
Design Considerations

- 3 possible behavioral interventions, delivered at family level:
  - Medication adherence (med)
  - Nutrition modifications (nut)
  - Physical activity reinforcement (phy)
- Scientific Question: What is the optimal sequence of these interventions?
  - Need to conduct a SMART
- Statistical Question: What are some of the SMART options?

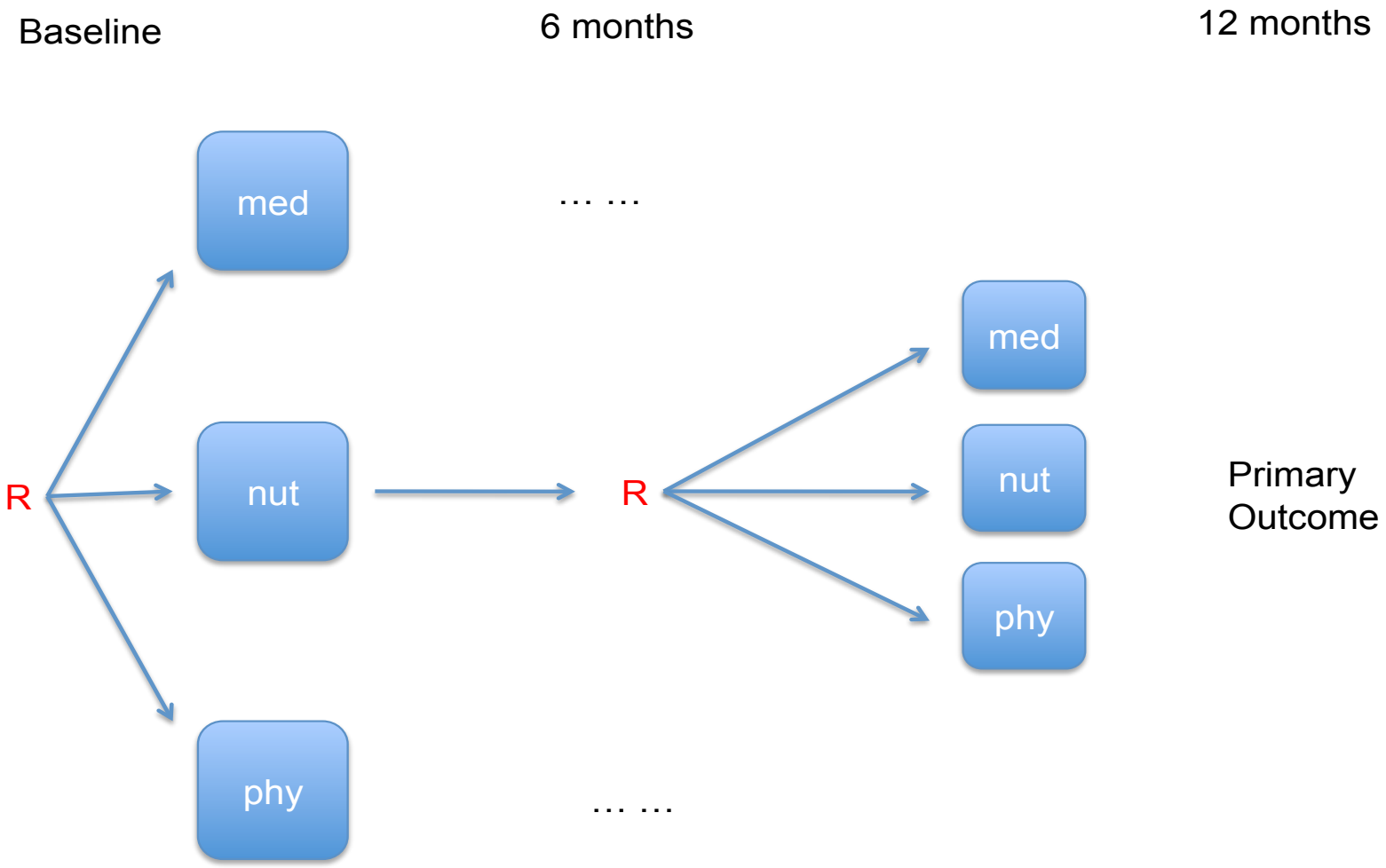
"Play the Winner, Switch from the Loser" SMART



"Randomize the Winner, Switch from the Loser" SMART



"Randomize All" (Factorial) SMART



Design Considerations

- In case of a family-based SMART:
  - The label "Responder" means the "patient in the family is a responder, irrespective of the response status of other family members"
  - Sample-size requirements will be higher due to intra-class correlation
- The 3 design options vary in terms of "exploration vs. exploitation dilemma" – a phrase from Reinforcement Learning (*Sutton and Barto, 1998*)
  - Which one is the best? What is the right metric to judge that?

Outline

Data Structure (w/o Clustering)

$$O_1, A_1, O_2, A_2, O_3$$

O_t	:	Observation (pre-treatment) at the tth stage
A_t	:	Treatment (action) at the tth stage, $A_t \in \mathcal{A}_t$
H_t	:	History at the tth stage, $H_1 = O_1, H_2 = (O_1, A_1, O_2)$
Y	:	Primary Outcome (larger is better)

A DTR is a sequence of decision rules:

$$d \equiv (d_1, d_2) \text{ with } d_t(h_t) \in \mathcal{A}_t$$

Q-learning

- The intuition comes from Dynamic Programming (*Bellman, 1957*) in case the multivariate distribution of the data is known
- Move backward in time to take care of the delayed effects
- Define the "Quality of treatment", Q-functions:

$$
\begin{aligned}
Q_2(H_2, A_2) &= \mathbb{E}\Big[Y_2 \Big| H_2, A_2\Big] \\
Q_1(H_1, A_1) &= \mathbb{E}\Big[Y_1 + \underbrace{\max_{a_2} Q_2(H_2, a_2)}_{\text{delayed effect}} \Big| H_1, A_1\Big]
\end{aligned}
$$

- Optimal DTR: $d_t(h_t) = \arg\max_{a_t} Q_t(h_t, a_t),\ t = 1, 2$

When the true Q-functions are unknown, need to first estimate them from a data set: $\{O_{1i}, A_{1i}, O_{2i}, A_{2i}, O_{3i}\},\ i = 1, \ldots, n$, and then apply dynamic programming

Q-learning with Linear Regression

- Linear regression models for Q-functions: $Q_t(H_t, A_t; \beta_t)$
- At stage 2, obtain $\hat{\beta}_2$ by least squares
- Construct stage-1 dependent variable:

$$\widehat{Y}_{1i} = Y_{1i} + \max_{a_2} Q_2(H_{2i}, a_2; \hat{\beta}_2), \ i = 1, \ldots, n$$

- At stage 1, regress $\widehat{Y}_1$ on suitable covariates to obtain $\hat{\beta}_1$
- Estimated Optimal DTR:

$$\hat{d}_t(h_t) = \arg\max_{a_t} \ Q_t(h_t, a_t; \hat{\beta}_t), \ t = 1, 2$$

Q-learning with Clustered Data

- Data trajectory for the jth member in the ith family, $1 \leq i \leq N$, $1 \leq j \leq n$ (WLOG, $j = 1$ indicates patient, and $j > 1$ family member): $(O_{1ij}, A_{1i}, O_{2ij}, A_{2i}, O_{3ij})$

- H_{tij} denote the subject history, and $H_{ti} = (H_{ti1}, \ldots, H_{tin})$ the corresponding family history

- Define the "clustered Q-functions" for subject j in family i:

$$\begin{aligned} Q_{2j}(H_{2i}, A_{2i}) &= E\left[Y_{ij} \middle| H_{2i}, A_{2i}\right] \\ Q_{1j}(H_{1i}, A_{1i}) &= E\left[\max_{a_2} Q_{2j}(H_{2i}, a_2) \middle| H_{1i}, A_{1i}\right] \end{aligned}$$

- Q-functions are defined by conditioning on the family history
  - As a result, the maximization accounts for information available from family members; and Q_{tj} and $Q_{tj'}$ are correlated within family

- Estimated Optimal DTR: $\hat{d}_{tj}(h_{ti}) = \arg\max_{a_t} Q_{tj}(h_{ti}, a_t; \hat{\beta}_t)$, $t = 1, 2$

Q-learning with GEE

- Define $Y_i = (Y_{i1}, \ldots, Y_{in})^T$, with working covariance $\Sigma_{2i} = \sigma^2 R, \forall i$, where σ^2 is the common variance of Y_{ij}'s, and R is the working correlation matrix (e.g. exchangeable)

- For the ith family, let $\Big(Q_{t1}(\beta_t), Q_{t2}(\beta_t) \ldots, Q_{tn}(\beta_t)\Big)^T = X_{ti}\beta_t$ where X_{ti} is the design matrix consisting of H_{ti} and A_{ti} and their interactions

- Next, define $Y = (Y_1^T, \ldots, Y_N^T)^T$, $X_2 = (X_{21}, \ldots, X_{2N})^T$ and $\Sigma_2 = diag(\Sigma_{21}, \ldots, \Sigma_{2N})$ is block-diagonal

- Using GEE, estimate β_2 by $\hat{\beta}_2 = (X_2^T \Sigma_2^{-1} X_2)^{-1} X_2^T \Sigma_2^{-1} Y$

- Construct the stage-1 dependent variables: $\widehat{Y}_{1ij} = \max_{a_2} Q_{2j}(h_{2i}, a_2; \hat{\beta}_2)$, and combine them to get $\widehat{Y}_1$

- Specify a working covariance and define X_1 and Σ_1 as above; hence estimate β_1 by $\hat{\beta}_1 = (X_1^T \Sigma_1^{-1} X_1)^{-1} X_1^T \Sigma_1^{-1} \widehat{Y}_1$

Q-learning with GEE

- The stage-2 estimator $\hat{\beta}_2$ is consistent for β_2 even for incorrect specification of Σ_2, and its variance can be consistently estimated by the well-known sandwich formula
- Theory unavailable at this point for stage-1 estimator (due to the infamous non-regularity)
- Practical Question: Does incorporating GEE into the Q-learning framework improve the quality of estimation in terms of the Value of the estimated DTR (i.e. mean outcome if the estimated DTR is used to assign treatment to the entire population)?

Outline

Simulation Design

- Time-varying variable of interest: (negative) systolic blood pressure (BP) – measured at baseline, 6 months and 12 months

- The generative model follows the effects observed in the Northern Manhattan Study (NOMAS) – a longitudinal cohort study of stroke risk factors among Black, Hispanic, and White populations in the Northern Manhattan area: http://columbianomas.org/study.html

- BP is generated via a linear model involving
  - subject-level characteristics: baseline physical activity (ACT), education, diabetes [following NOMAS]
  - family-level characteristic: ethnicity (represented by two indicators BLACK and HISP) [following NOMAS]
  - family-level random effect: $\sim N(0, \sigma_F^2)$ [simulated SMART]
  - treatment variables: NUT, PHY, HISP $\times$ NUT, ACT $\times$ PHY [simulated SMART]
  - random error: $\sim N(0, \sigma_e^2)$ [following NOMAS]

Simulation Design

- The size of treatment variables are varied through a parameter b, set according to Cohen's (1988) small and medium effect sizes
- Three types of SMART designs are employed to assign treatments
- Methods compared: Q-learning with patient only data (QL-Patient), Q-learning with all subjects (QL-All), Q-learning with GEE (QL-GEE)
- Metric to judge performance: Value
- 1000 simulated data sets, each consisting of 300 families, where each family has 3 members (1 patient)
- Given a training data (and hence an estimated DTR), its Value is computed using a separate Monte Carlo evaluation data set of size 2000 patients

Results

"Play the Winner, Switch from the Loser" SMART

Example	Effect Size, b	σ_F/σ_e	QL-Patient	QL-All	QL-GEE
1	Small	1	27.24	27.45	27.47
2	Small	2	26.68	26.79	26.82
3	Small	4	25.91	25.98	26.02
4	Medium	1	27.96	27.97	28.00
5	Medium	2	27.60	27.68	27.79
6	Medium	4	25.80	26.06	26.16

Results

"Randomize the Winner, Switch from the Loser" SMART

Example	Effect Size, b	σ_F/σ_e	QL-Patient	QL-All	QL-GEE
1	Small	1	27.26	27.30	27.42
2	Small	2	26.55	26.68	26.73
3	Small	4	25.93	25.96	26.04
4	Medium	1	27.99	27.96	28.02
5	Medium	2	27.72	27.80	27.82
6	Medium	4	25.76	25.99	26.21

Results

"Randomize All" (Factorial) SMART

Example	Effect Size, b	σ_F/σ_e	QL-Patient	QL-All	QL-GEE
1	Small	1	27.27	27.40	27.42
2	Small	2	26.61	26.74	26.78
3	Small	4	26.09	26.13	26.15
4	Medium	1	27.98	27.97	28.00
5	Medium	2	27.70	27.74	27.81
6	Medium	4	25.81	26.00	26.18

Outline

Summary and End Notes

- All 3 versions of SMART perform comparably in terms of Value – one may be preferred over the others by ethical and/or logistical considerations

- We propose to use an extension of Q-learning incorporating GEE (QL-GEE) for estimating optimal DTRs from family-based or community-based studies where clustered data arise naturally
  - In the simulations tried out so far, QL-GEE performs only marginally better than standard Q-learning
  - The hope is to see bigger differences in performance in yet-unexplored settings, and perhaps smaller sample sizes

- Formally constructing CIs for the Value of the estimated DTRs is the natural next step – it is challenging, but early inroads have been made into that problem
  (talk @ UPenn Conference in April – will be published in a special issue of 'Clinical Trials' containing the conference proceedings)

It’s all work-in-progress... Don’t take me too seriously!!