An Introduction to Dynamic Treatment Regimes (DTRs)

Silvia Metelli
CRESS, Université de Paris

ENS Paris Saclay Lecture

9 March 2021

Topics covered

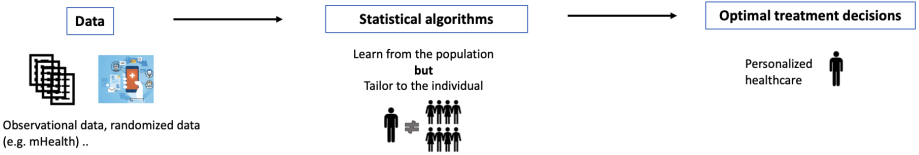
1. Overview on Dynamic Treatment Regimes (DTRs)
   - context - personalised medicine -
   - examples
   - the formal set-up
2. DTRs under randomised studies
3. DTRs under observational studies
4. Special case: mHealth DTRs (aka just-in-time adaptive interventions, JITAIs)
5. Discussion: opportunities and challenges

Personalised medicine

The right treatment/s for the right patient

The individual is 'at the center':

From a statistical perspective:



Time-varying personalised medicine

Precision medicine

- deals with heterogeneity between individuals
- treatment decision is tailored to patient's characteristics/medical history
- can increase treatment effectiveness
- can reduce costs and treatment burdens

Time-varying Precision medicine

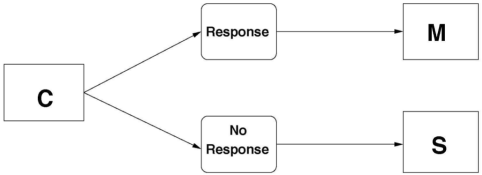
- same as above
- plus, deals with heterogeneity between individuals over time
- move away from *"once and for all"* treatment paradigm

↓

....This precisely fits a Dynamic Treatment Regimes framework!

Motivating example

Acute leukemia



Here, a clinician confronts two decisions:

- Decision 1: Induction chemotherapy (C, options c1, c2) [induce positive response]
- Decision 2: [sustain the response]
  - Maintenance treatment for patients who respond (M, options m1, m2)
  - Salvage chemotherapy for patients who do not respond (S, options s1, s2)

"Give induction therapy C followed by maintenance M if response else give salvage therapy S" is a (dynamic) treatment regime

The clinician makes these decisions so to maximise the expected benefit w.r.t an outcome (overall or disease-free survival)

How? he uses his judgement to evaluate accrued information at each decision point, including demographics, medical history, genetics, biomarkers, adverse events...

Similarly, statistics aims to operationalise this decision-making process in a quantitative way

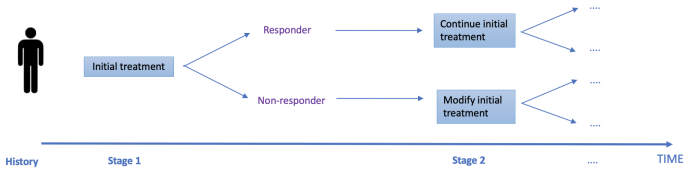
How? building 'algorithms' to make these decisions, maximising some predefined criteria

Part I: Dynamic treatment regimes (DTRs)

A **dynamic** treatment regime

DTRs/Adaptive Interventions/Adaptive treatment strategies...

- A sequence of decision rules (algorithm) that dictates how to treat a patient over time



- An intervention may not work for all people at all times
  - heterogeneity across people → decisions should be *personalized*
  - heterogeneity over time → personalized decision can change over time
- DTRs are needed because
  - chronic nature of many health outcomes (adjust, change, add, discontinue treatment)
  - can improve disease management and balance effectiveness and risk

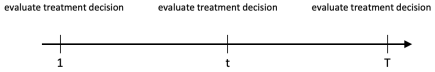
Typical questions a DTR aims to answer

- What are the options at each step?
- What should be the timing of the steps? (will see this later with JITAIs)
- How do we best personalise the sequence of treatments? i.e. What individual information (variables) should we use to make these decisions?
- What is the best sequence of treatment steps? i.e. optimal regime

Can statistics help?

Statistical framework

Stages: $1, \ldots, T$



On each individual: $o_1, a_1, y_2, \ldots, o_T, a_T, y_{T+1}$

- t: decision point
- o_t: observation (pre-treatment) between stage t-1 and stage t
- a_t: treatment options at the t-th stage: $a_t \in \mathcal{A}_t$ (e.g. $a_t \in \{0, 1\}$)
- history $h_t = (o_1, a_1, \ldots, o_{t-1}, a_{t-1}, o_t)$: information accrued prior to t
- y: primary outcome

A DTR is a sequence of decision rules $d_1(h_1), d_2(h_2), \ldots, d_T(h_T)$ s.t.:

$$d \equiv (d_1, \ldots, d_T) \quad \text{with} \quad d_t : \mathcal{H}_t \to \mathcal{A}_t$$

- $d \in \mathcal{D}$, $\mathcal{D}$ set of all possible regimes

Example: Acute leukemia again

Decision 1, d_1:

If age $<$ 50 years and White Blood Cell (WBC) $<$ 10.0 $\times$ 103/μl, give chemotherapy c_2, otherwise give c_1
rule $d_1(h_1) : \mathcal{H}_1 \rightarrow \{c_1, c_2\}$

Decision 2, d_2:

If patient responded and baseline WBC $<$ 11.2, current WBC $<$ 10.5, no grade 3+ hematologic adverse event, current ECOG Performance Status $\leq$ 2, give maintenance m_1, otherwise, give m_2; otherwise
If patient did not respond and age $>$ 60, current WBC $<$ 11.0, ECOG $\geq$ 2 give s_1, otherwise, give s2
rule $d_2(h_2) : \mathcal{H}_2 \rightarrow \{m_1, m_2\}$ if responder, $d_2(h_2) : \mathcal{H}_2 \rightarrow \{s_1, s_2\}$ if nonresponder

Regime $d = (d_1, d_2)$

Optimal DTR

A DTR is called optimal if it optimises the long-term mean outcome, formally:

An optimal regime $d^{opt} \in \mathcal{D}$ satisfies:

- If all patients in the population were to receive all T treatments according to d^{opt}, the expected (average) outcome for the population would be as large as possible

Formalising this, take ideas from causal inference...

Potential Outcomes (counterfactuals) for T-stage decisions

Potential outcomes in a static regime for randomly chosen individual with history H_1:

- $Y^*(0)$ and $Y^*(1)$, under options 0 and 1

Potential outcomes, suppose individual with $H_1 = X_1$ were to receive $d = (d_1, \ldots, d_T) \in \mathcal{D}$

- $\{X_2^*(d), X_3^*(d), \ldots, X_T^*(d), Y^*(d)\}$

  with X_t^* potential intermediate outcome that would arise between $t-1$ and t

Optimal DTR: formal definition

Goal: given data either from randomised trial or observational study, estimate an optimal regime d^{opt}, based on some maximisation criterion

For regime $d \in \mathcal{D}$: $E\{Y^*(d)\}$ is the expected (average) outcome if all individuals in the population were under $d \in \mathcal{D}$

Definition of an optimal regime: d^{opt} is a regime in $\mathcal{D}$ such that

$$d^{opt} = \arg\max_{d \in \mathcal{D}} E\{Y^*(d)\}$$

$$\text{i.e. } E\{Y^*(d^{opt})\} \geq E\{Y^*(d)\} \quad \text{for all} \quad d \in \mathcal{D}$$

Assumptions for identifiability

Use causal framework to assess DTRs under:

1. Consistency: if a patient's observed treatment history is compatible with a DTR, his/her clinical outcomes are the same as the counterfactual ones under the DTR
2. Sequential randomisation (no unmeasured confounding) Treatment decision at a given time is independent of future observations and counterfactual outcomes, conditional on all covariates and treatment history
   - automatically satisfied in a randomised trial
   - unverifiable but necessary in an observational study
3. Positivity: With probability 1, each subject follows the DTR with non-negative probability

Data sources

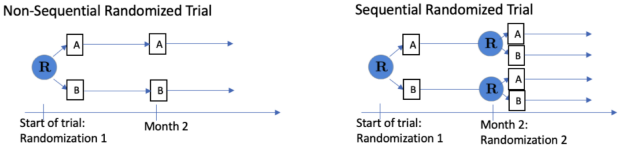
Studies:

1. Randomized trials
2. Longitudinal observational data

DTRs from Randomized Trials

Trials for testing the time-varying effects of interventions → standard trials are inadequate

Need for **Sequential Multiple Assignment Randomized Trial (SMART)**

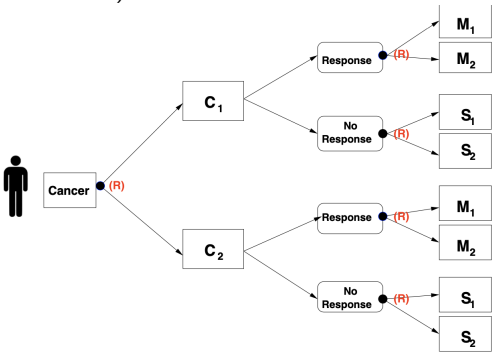


- Multi-stage trial: each stage is a treatment decision
- At each stage the patient is randomised
- After observing o_t, the clinical trialist takes action a_t with randomisation prob. $p_t(a_t|h_t)$

SMART Example

Sequential Multiple Assignment Randomized Trials

Leukemia example (Davidian *et al.*)

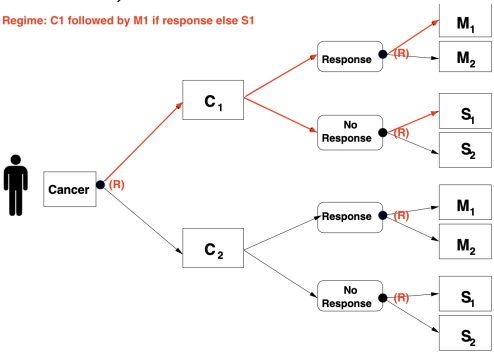


note: Sequential Randomisation Assumption automatically holds

SMART Example

Sequential Multiple Assignment Randomized Trials

Leukemia example (Davidian *et al.*)



note: Sequential Randomisation Assumption automatically holds

Estimating d^{opt}: Q-learning

Q-learning (Watkins, 1989)

- popular method from Reinforcement (Machine) Learning
- a generalisation of regression to multistage decision problems
- implemented with many variations
- intuitions from dynamic programming (Bellman, 1957)

Reinforcement learning (RL) algorithm used to learn an optimal policy for each user: at each decision time, select treatment action $A_t = a_t$ to maximize a sum of rewards

Estimating d^{opt}: Q-learning

For simplicity $T = 2$ and $A_k = \{0, 1\}$, $t = 1, 2$

Accrued information $h_1 = x_1, h_2 = (x_1, a_1, x_2)$

Optimal regime d^{opt}: move backward in time

Define the "Quality of Treatment", Q-functions (total expected future reward starting from t=2):

$$Q_2(h_2, a_2) = E\{Y|H_2 = h2, A_2 = a_2\}$$

$$Q_1(h_1, a_1) = E\{\max_{a_2} Q_2(H_2, a_2)|H_2 = h2, A_1 = a_1\}$$

Optimal DTR: $d_t^{opt} = \arg\max_{a_t} Q(h_t, a_t)$, $t = 1, 2$

When the true Q-functions are unknown, need to estimate from data,
using e.g. regression models ...

Estimating d^{opt}: Q-learning

Regression models for Q-functions for $t = 1, 2$ (posit linear for simplicity)

$$Q_t(H_t, A_t, \beta_t, \psi_t) = \beta_t^T H_t + (\psi_t^T H_t) A_j$$

At decision stage 2: regress y on $(H_2, H_2 A_2)$ to obtain $(\hat{\beta}_2, \hat{\psi}_2)$

For each i, construct stage 1 'pseudo-outcome':

$$\tilde{y}_{it} = \max_{a_2} Q_2(H_{2i}, a_2, \hat{\beta}_2, \hat{\psi}_2)$$

At decision stage 1: regress $\tilde{y}$ on $(H_1, H_1 A_1)$ to obtain $(\hat{\beta}_1, \hat{\psi}_1)$

Estimated optimal regime: $\hat{d}_t^{opt} = \arg\max_{a_t} Q(h_t, a_t, \hat{\beta}_t, \hat{\psi}_t) = sign(\hat{\psi}_j^T h_t)$

Other methods (non comprehensive list)

Model the conditional mean

- Q-learning
- Likelihood-based methods
  (first estimate full multivariate distribution of the data, then maximise)
  - Bayesian methods

Model the conditional mean contrasts

- A-learning
- Structural Nested Mean Models
- G-estimation (GMM)
- ...

DTRs from observational studies

Most sources of data are observational in nature
Examples: databases from prospective studies, registries..

Main challenge is time-varying confounding

- subjects receiving one treatment or another may not be similar
- in DTRs confounding can mix over stages
- standard bias-adjustment methods cannot handle it well

Observational studies vs. Randomised studies

Under the Sequential Randomisation Assumption:

- in SMART: automatic through randomisation
- in observational study: only if all information used to make decisions is collected (assumption is unverifiable)
- main differences:
  - SMART: weighting based on known randomisation probabilities
  - Observ: weighting based on propensities of receiving treatment as a function of history (e.g. Q-learning extended via propensity scores)
  - SMART: randomised but short follow-ups, highly selected patients
  - Observ: non-randomised but long follow-ups, frequent in practise

Part II: mHealth dynamic treatment regimes (JITAIs)

The special case of JITAIs

Timing in mobile health

mHealth:

- use of mobile/wearable devices for healthcare delivery
- routinely collect information about an individual in real time
- use information to decide whether or not to intervene on the individual
- potential for providing low-cost supportive behavioural interventions

Timing:

- due to mhealth nature, interventions can be timingly
- avoid delays, maximize benefits of interventions
- "Just-in-time" + "Adaptive" interventions

What is a JITAI?

Special adaptive intervention delivered when & where needed (mHealth DTR):

- support is provided just at the right time (send message to propose intervention)
- finer time scales: minutes, hours, instead of weeks/months

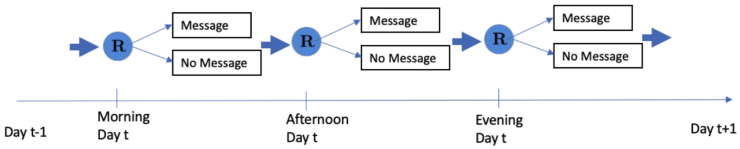
JITAI: the right type/amount of support, **at the right time**, by adapting to an individual's changing internal and contextual state, i.e. when he needs it most and is most likely to be receptive (Nahum-Shani *et al.*, 2018)

$\Downarrow$

Great potential for promoting health behaviour change

Typical domains: physical inactivity, alcohol use, smoking, mental illness,..
but also Intensive Care Unit frameworks...

Micro Randomized Trials



- MRT specific for delivering quick/frequent interventions (mostly in mHealth)
- MRTs typically used to estimate short-term effect of mHealth interventions

Example: HeartSteps app

Mobile app to encourage physical activity for individuals at risk of coronary artery disease (110 participants monitored for 42 days)

Outcome: sustained physical activity

1: at a given decision point t (5 times per day) **do**
2: mobile phone collects observations O_t (weather, location, step count..)
3: a decision rule maps observations into an intervention action a_t:
a binary risk factor R_t set 1 if step count < 150
if $R_t = 1$, (R): *then* a_t={Send message} (suggestion to walk)
else if $R_t = 0$, *then* a_t={Send nothing}
4: mobile phone records the proximal outcome Y_{t+1} (step count in next 30 min)
5: **done**

- HeartSteps is an MRT with 5×42=210 decision points
- Randomization probabilities chosen to avoid overburden:
  - average of 2.5 message per day in high risk situations, $p_t(A_t) = 0.6$

Conceptual models

Need to develop statistical methods to model and estimate the time-varying proximal treatment effect (work mostly developed by S. Murphy and collegues)

Data: $O_1, A_1, Y_2, \ldots, O_T, A_T, Y_{T+1}$
History $H_t = (O_1, A_1, Y_2, \ldots, O_t)$: information up to t

$$Y_{t+1}\text{`} \sim' g(H_t)^T \alpha + \beta_0 A_t$$
$$Y_{t+1}\text{`} \sim' g(H_t)^T \alpha + \beta_0 a_t + \beta_1 A_t S_t$$

- β_0, β_1: quantities of interest
- S_t: potential moderators (e.g., current weather good or day of study)
- '$\sim$' : e.g., logit or log for binary Y

Estimate time-varying effects

In a standard setting, we would have
β_0: effect of activity suggestion on subsequent activity
$\beta_0 + \beta_1$: effect of activity suggestion on subsequent activity when good whether

..but here treatment is time-varying $(A_t, t = 1, \ldots, T)$ and O_t and S_t may be affected by prior treatment action

$\downarrow$

need to develop methods consistent with standard meaning of β
need to use control variables $g(H_t)$ for noise reduction in a robust way

- Models for continuous proximal treatment effect (Boruvka et. al., 2018)
- Models for binary proximal treatment effect (Qian et al., 2020)

A case study: results from HeartSteps data

Research question: is there a causal effect of delivering an activity suggestion vs. not delivering on the proximal outcome (subsequent 30 min step count)?

Causal effect term	estimate	95%CI	p-value
$\beta_0 A_t$	β_0 =0.13	(-0.01,0.27)	0.6
$\beta_0 A_t + \beta_1 A_t d_t$	β_0 =0.51	(0.20,0.81)	< 0.1
	β_1 =-0.21	(-0.30,-0.10)	< 0.1

$\beta_0 = 0.13 \rightarrow$ 14% increase in step count over no treatment
$\beta_0 = 0.51 \rightarrow$ 67% increase in step count over no treatment
β_1 =-0.21 means the effect of intervention deteriorates with time

- data indicates there is an effect of activity suggestion on step count over next 30 min, but users disengage with time (d_t days in study)
- more sophisticated designs, e.g. a one-week break from activity suggestion

"Online" algorithms

Online: continuously learn and optimize the treatment policy in the JITAI as the user experiences the intervention

Online reinforcement learning: goal is to learn as fast as possible which actions maximise the T-horizon total reward

Some challenges to RL:

- need to balance complexity with learning speed
- algorithm must adjust for longer term effects of current actions: optimal treatment can only be identified by taking into account the impact of current action on the future rewards

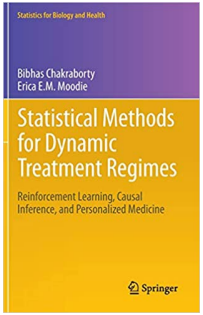
Discussion

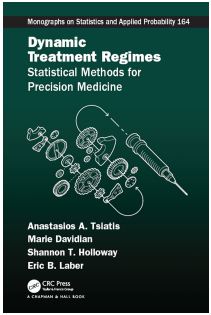
- Statistical methods for **DTRs hold great potential** to complement clinical expertise
- Informing construction of DTRs requires new ways of thinking about clinical trials
- Analysis from randomised trials simpler, but short follow-ups, selected population
- Analysis from longitudinal data is more complex, but longer follow-ups
- Statistical inference can be challenging and nonstandard
- DTRs requires multidisciplinary efforts: clinicians, behavioural scientists, statisticians, computer scientists etc..
- mHealth technology can collect data and deliver more timely, tailored interventions

Open Challenges

- The field is itself challenging
- Most application involves survival outcomes, methods less developed
- Multiple outcomes (e.g., efficacy and toxicity)
- Design considerations for SMARTs
- Achieve more effective personalisation
- Real time regimes (mHealth), online algorithms
- ....
- ....
- But there is a vast amount of stats literature being developed!

Some References (DTRs)





Some References (DTRs)

Murphy, S.A. (2003). Optimal dynamic treatment regimes. *Journal of the Royal Statistical Society, Series B*, 65.

Robins, J.M. (2004). Optimal structural nested models for optimal sequential decisions. In D. Y. Lin & P. Heagerty (Eds.), Proceedings of the second Seattle symposium on biostatistics (pp. 189-326). New York: Springer.

Lunceford J.K., Davidian M., Tsiatis A.A. (2002). Estimation of survival distributions of treatment policies in two-stage randomization designs in clinical trials. *Biometrics*, 58, 48-57.

Murphy, S.A. (2005). An experimental design for the development of adaptive treatment strategies, *Statistics in Medicine*, 24, 1455-1481. See http://people.seas.harvard.edu/ samurphy/research.html

Thall P.F., Millikan R.E., Sung H.G. (2000). Evaluating multiple treatment courses in clinical trials. *Statistics in Medicine*, 30, 1011-1128.

Wahed A.S., Tsiatis A.A. (2004). Optimal estimator for the survival distribution and related quantities for treatment policies in two-stage randomization designs in clinical trials. *Biometrics*, 60, 124-133.

Wahed A.S., Tsiatis A.A. (2006). Semiparametric efficient estimation of survival distribution for treatment policies in two-stage randomization designs in clinical trials with censored data. *Biometrika*, 93, 163-177.

Some References (JITAIs)

JITAI seminal paper:

Nahum-Shani, I., Smith, S.N. Spring, B.J., Collins, L.M., Witkiewitz, K., Tewari, A., & Murphy, S. A. (2018). Just-in-time adaptive interventions (JITAIs) in mobile health: Key components and design principles for ongoing health behavior support. *Annals of Behavioral Medicine*, 18;52(6), 446-462.

Technical papers:

Boruvka, A., Almirall, D., Witkiewitz, K., & Murphy, S.A. (2018). Assessing Time-Varying Causal Effect Moderation in Mobile Health. *Journal of the American Statistical Association*, 113(523), 1112-1121.

Qian, T., Yoo, H., Klasnja, P., Almirall, D. and Murphy, S.A. (2020). Estimating Time- Varying Causal Excursion Effects in Mobile Health with Binary Outcomes. To appear in *Biometrika*.

Liao, P., Klasnja, P., Tewari, A., & Murphy, S.A. (2016). Sample size calculations for micro-randomized trials in mHealth. *Statistics in medicine*, 35(12), 1944-1971.

Data analysis:

Liao, P., Greenewald, K., Klasnja, P. & Murphy, S.A. (2020). Personalized HeartSteps: A Reinforcement Learning Algorithm for Optimizing Physical Activity. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 4(1), 1-22.