Nonadherence in Dynamic Treatment Regimes – Moving Beyond Intention to Treat Analyses

Dylan Spicker

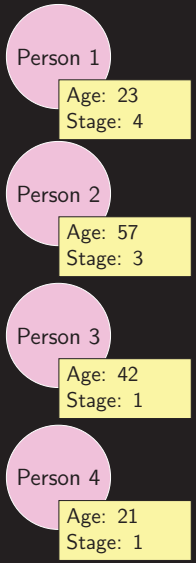
Department of Epidemiology, Biostatistics and Occupational Health
McGill University

Wednesday October 19, 2022

Precision Medicine

Treat the patient, not the disease.

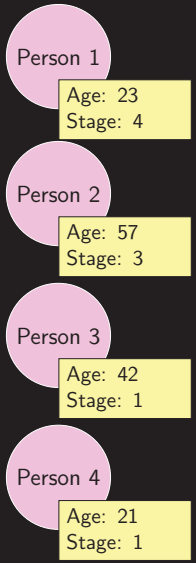
Dynamic Treatment Regimes



Experimental Treatment
($A = 1$)

Standard Treatment
($A = 0$)

Dynamic Treatment Regimes



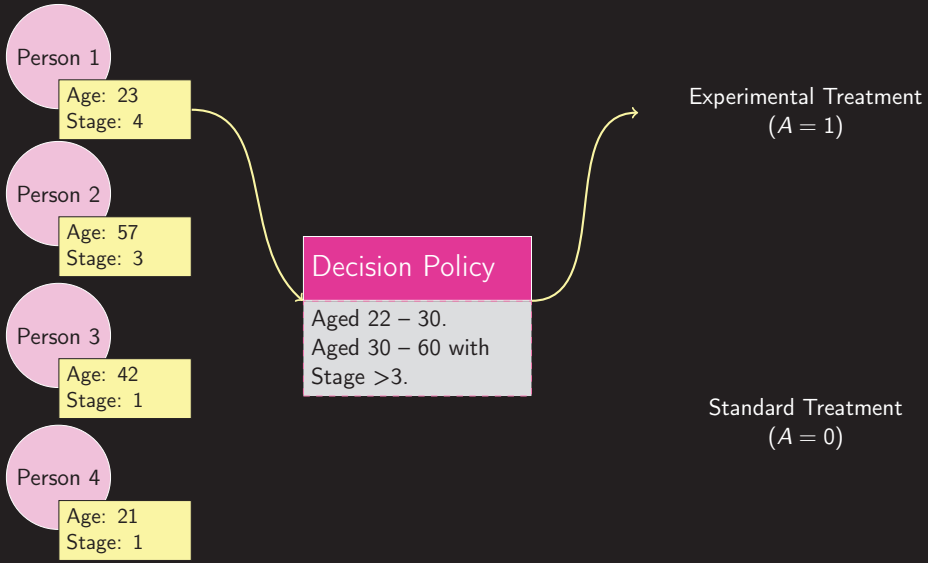
Decision Policy

Aged 22 – 30.
Aged 30 – 60 with Stage >3.

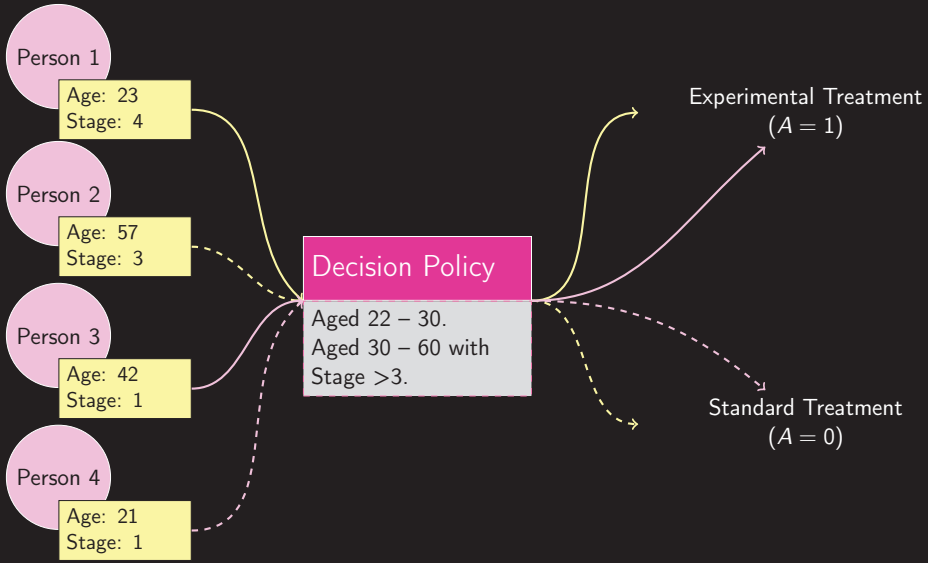
Experimental Treatment
($A = 1$)

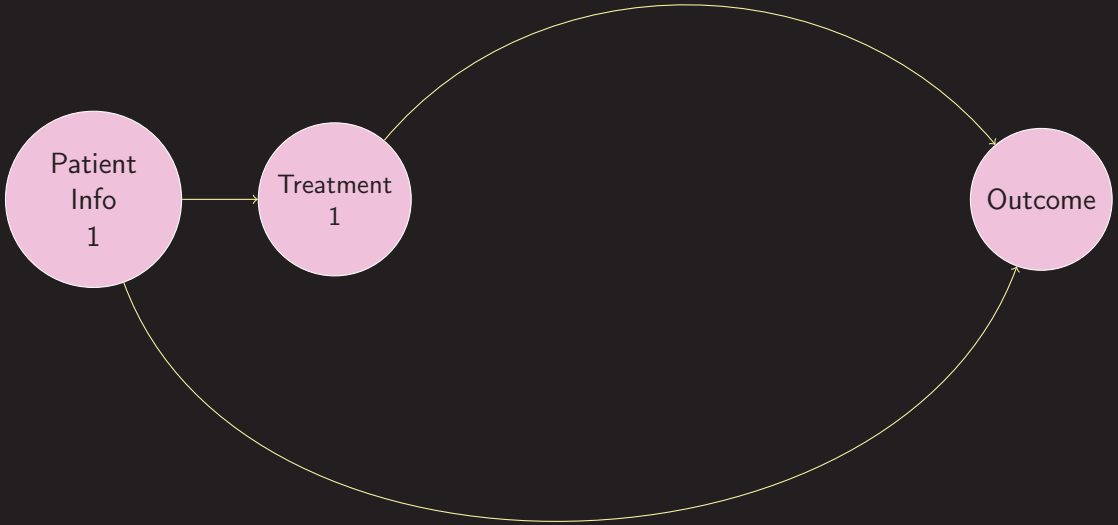
Standard Treatment
($A = 0$)

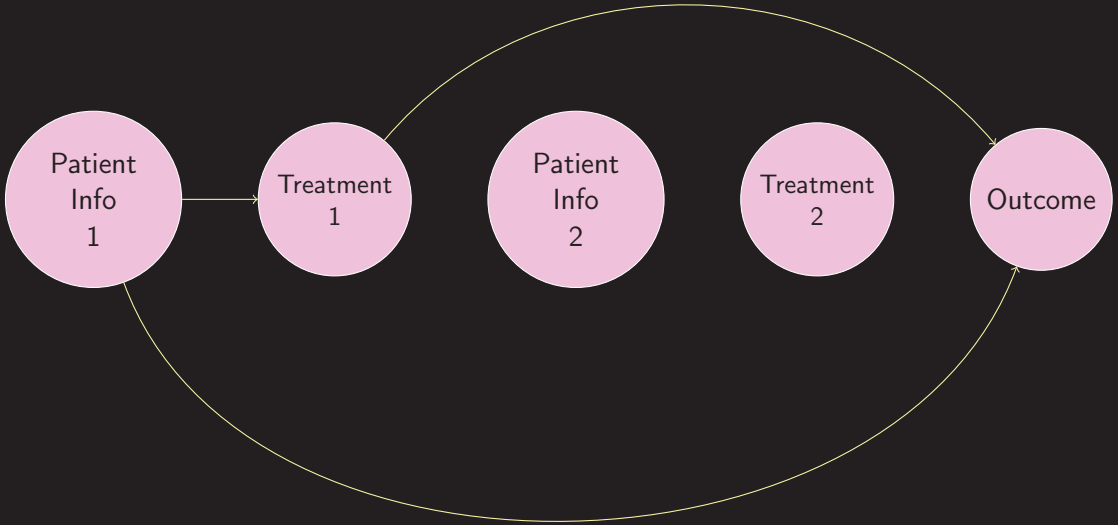
Dynamic Treatment Regimes

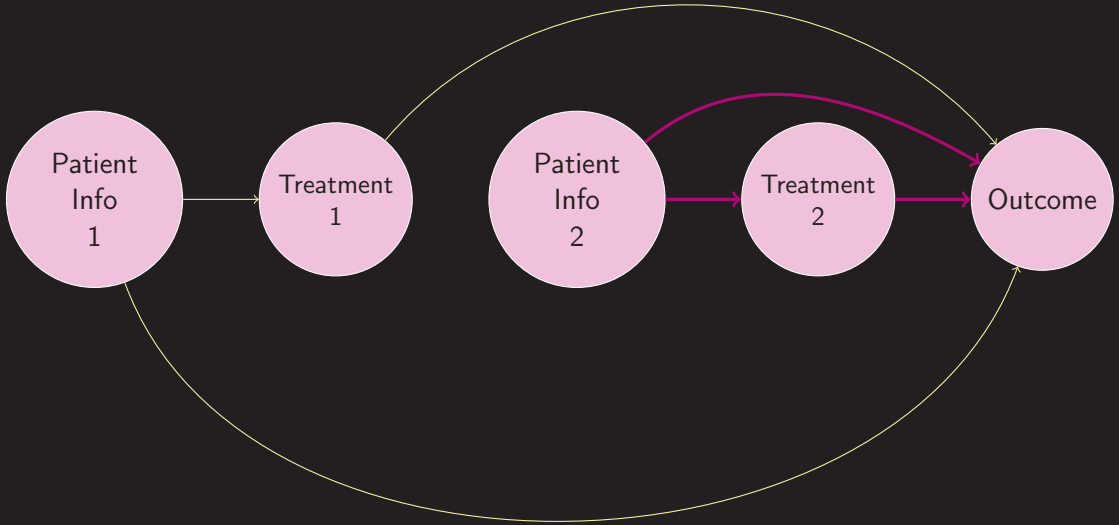


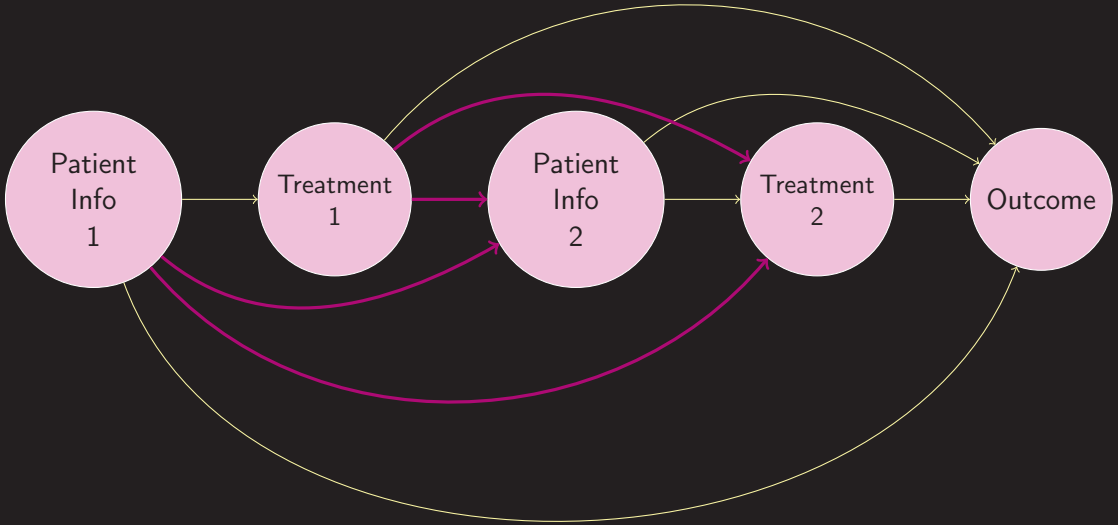
Dynamic Treatment Regimes

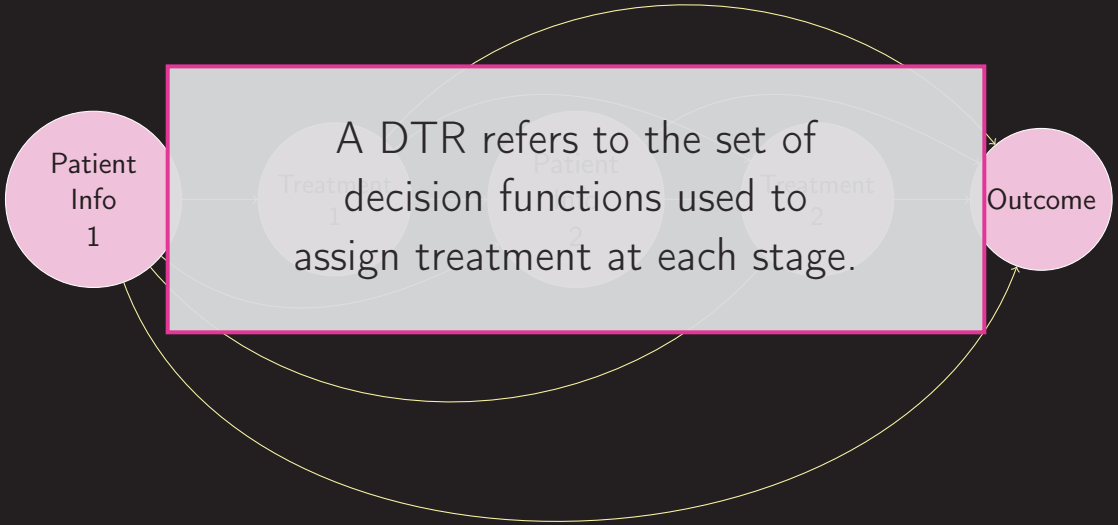



Patient
Info
1
Treatment
1
Outcome


Patient Info 1
Treatment 1
Patient Info 2
Treatment 2
Outcome


Patient Info 1
Treatment 1
Patient Info 2
Treatment 2
Outcome


Patient Info 1
Treatment 1
Patient Info 2
Treatment 2
Outcome


Patient Info 1
A DTR refers to the set of decision functions used to assign treatment at each stage.
Outcome

Simplifying Assumptions

Binary Treatments

Referring to two competing treatment options at each stage.

Known Decision Points

Finite and deterministic number and timing of decisions.

At decision point j, we take $A_j \in \{0, 1\}$.

The work today does not require this assumption, though discrete treatments are required.

Simplifying Assumptions

Binary Treatments

Referring to two competing treatment options at each stage.

Known Decision Points

Finite and deterministic number and timing of decisions.

We require decisions to be made at a finite and deterministic number of points.

We assume that these are discrete and interchangeable between patients.

Notation Summary

1. Decision points:

$$j \in \{1, 2, \ldots, K\}.$$

Notation Summary

1. Decision points:

$$j \in \{1, 2, \ldots, K\}.$$

2. Treatments (at time j):

$$A_j \in \{0, 1\}.$$

Notation Summary

1. Decision points:
$$j \in \{1, 2, \ldots, K\}.$$
2. Treatments (at time j):
$$A_j \in \{0, 1\}.$$
3. Individual information (at time j):
$$X_j \in \mathbb{R}^{\ell_j}.$$

Notation Summary

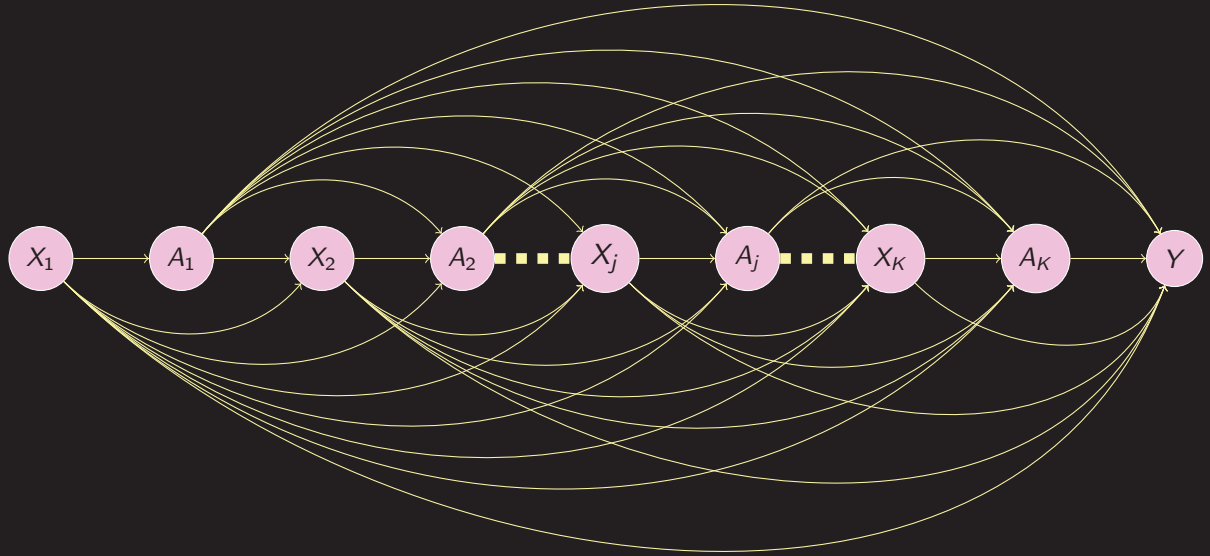
1. Decision points:
$$j \in \{1, 2, \ldots, K\}.$$
2. Treatments (at time j):
$$A_j \in \{0, 1\}.$$
3. Individual information (at time j):
$$X_j \in \mathbb{R}^{\ell_j}.$$
4. Outcome (observed at time K):
$$Y \in \mathbb{R}.$$

Our goal is to determine

$$d = \{d_1, d_2, \ldots, d_K\}, \quad d_j : \mathbb{R}^{\ell_j^*} \longrightarrow \{0, 1\},$$

such that Y is maximized in expectation.

Why is DTR Estimation Hard?



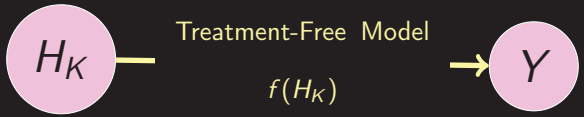
Why is DTR Estimation Hard?

Modelling the Outcome

$$E[Y|A_K, H_K] = f(H_K) + A_K C(H_K)$$

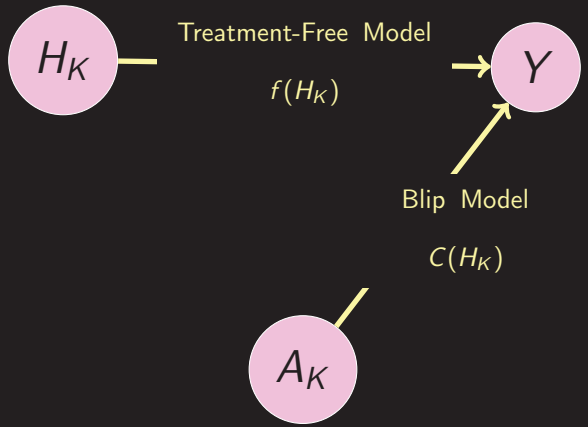
Modelling the Outcome

$$E[Y|A_K, H_K] = f(H_K) + A_K C(H_K)$$



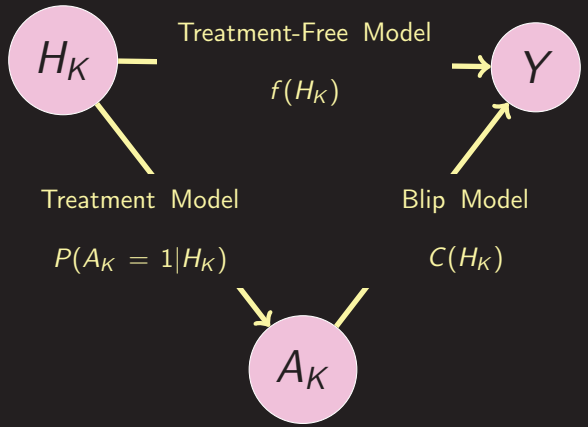
Modelling the Outcome

$$E[Y|A_K, H_K] = f(H_K) + A_K C(H_K)$$



Modelling the Outcome

$$E[Y|A_K, H_K] = f(H_K) + A_K C(H_K)$$



G-Estimation: Mathematically (Robins 2004)

1. Specify the required models at stage K.

$$C_K(H_K;\psi_K); \quad f_K(H_K;\beta_K); \quad \pi_K(H_K;\alpha_K).$$

G-Estimation: Mathematically (Robins 2004)

1. Specify the required models at stage K.
2. Solve for ψ_K in

$$\sum_{i=1}^{n} \lambda_K(H_{K,i}) \left\{ A_{K,i} - \underbrace{\pi_K(A_{K,i}; \alpha_K)}_{\text{Treatment Model}} \right\} \times \left\{ Y_i - A_{K,i} \underbrace{C_K(H_{K,i}; \psi_K)}_{\text{Blip Model}} - \underbrace{f_K(H_{K,i}; \beta_K)}_{\text{Treatment-Free Model}} \right\} = 0.$$

G-Estimation: Mathematically (Robins 2004)

1. Specify the required models at stage K.
2. Solve for ψ_K in estimating equations.
3. Compute a pseduo-outcome for each individual.

$$\widetilde{V}_j = \begin{cases} \widetilde{V}_{j+1} + (A_j^{\text{opt}} - A_j) C_j(H_j; \psi_j) & j \leq K \\ Y & j = K+1 \end{cases}.$$

G-Estimation: Mathematically (Robins 2004)

1. Specify the required models at stage K.
2. Solve for ψ_K in estimating equations.
3. Compute a pseduo-outcome for each individual.
4. Repeat the previous steps, replacing Y with $\widetilde{V}_{j+1}$ at each stage j.

G-Estimation: Mathematically (Robins 2004)

1. Specify the required models at stage K.
2. Solve for ψ_K in estimating equations.
3. Compute a pseduo-outcome for each individual.
4. Repeat the previous steps, replacing Y with $\widetilde{V}_{j+1}$ at each stage j.

G-estimation produces doubly robust estimators for the blip parameters.

Causal Assumptions

SUTVA

Stable Unit Treatment Value Assumption

NUC

No Unmeasured Confounders.

Positivity

No Extrapolation Outside the Data.

There is only one version of each treatment option, and there is no interference between patients in the data.

Causal Assumptions

SUTVA

Stable Unit Treatment Value Assumption

NUC

No Unmeasured Confounders.

Positivity

No Extrapolation Outside the Data.

There are no factors which influence both the treatment assignment as well as the outcome, which are not measured.

Causal Assumptions

SUTVA

Stable Unit Treatment Value Assumption

NUC

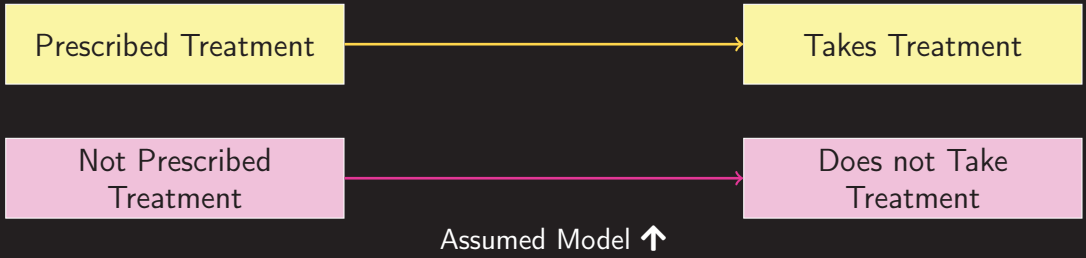
No Unmeasured Confounders.

Positivity

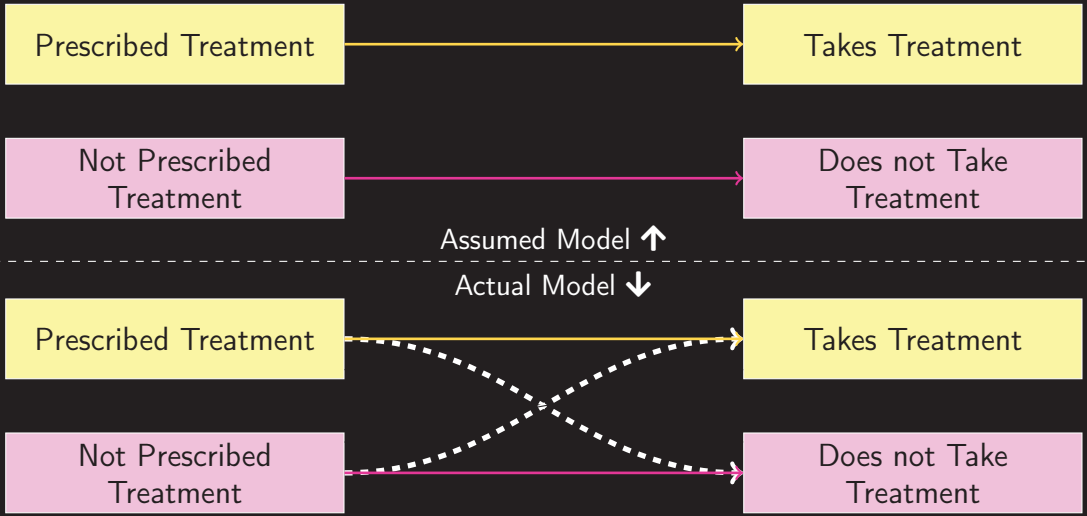
No Extrapolation Outside the Data.

All treatment regimes under consideration must have been possible for all individuals in the data.

The Problem of Nonadherence



The Problem of Nonadherence



Is the intervention or the intended intervention of more interest.

Treat the assigned intervention as the causal quantity of interest.

Treat the assigned intervention as the causal quantity of interest.

Random Assignment

Treatment assignment is the randomized quantity, and so non-confounded.

Treat the assigned intervention as the causal quantity of interest.

Random Assignment

Treatment assignment is the randomized quantity, and so non-confounded.

Real-World

Nonadherence exists in the "real-world", and so should be factored-in.

Treat the assigned intervention as the causal quantity of interest.

Random Assignment

Treatment assignment is the randomized quantity, and so non-confounded.

Real-World

Nonadherence exists in the "real-world", and so should be factored-in.

Standard Practice

It is standard practice, and better than as-treated or per-protocol.

ITTs can be very valuable ...
but this is not the whole story.

Problems with ITTs

Altered Causal Estimand

We estimate the effect of treatment assignment not treatment itself.

Non-Attenuation

Unpredictable as to whether an ITT correctly ranks multiple treatments.

Transportability Concerns

Are there differences in how patients adhere inside and outside the study?

There is likely substantial scientific interest in the biological efficacy of treatment, if it had been adhered to.

ITTs cannot estimate treatment efficacy in an unbiased manner.

Problems with ITTs

Altered Causal Estimand

We estimate the effect of treatment assignment not treatment itself.

Non-Attenuation

Unpredictable as to whether an ITT correctly ranks multiple treatments.

Transportability Concerns

Are there differences in how patients adhere inside and outside the study?

Despite claims of effect attenuation under nonadherence, there is no statistical guarantee that treatment rankings are preserved.

ITTs cannot be used to compare treatments based on treatment efficacy.

Problems with ITTs

Altered Causal Estimand

We estimate the effect of treatment assignment not treatment itself.

Non-Attenuation

Unpredictable as to whether an ITT correctly ranks multiple treatments.

Transportability Concerns

Are there differences in how patients adhere inside and outside the study?

The adherence rates may differ between the sample and population.

ITT Effect = Adherence $\times$ Treatment Efficacy.

Does adherence remain the same in the population?

Should we conduct research into improving adherence?

Should we use treatments in a supervised setting?

ITTs can be very valuable ...
but this is not the whole story.

See for instance Sheiner and Rubin (1995) or Shrier, Verhagen, and Stovitz (2017).

What other options do we really have?

ITTs for DTRs: A Justification

1. In a DTR, we are predicting the optimal treatment regime, from observed data.

ITTs for DTRs: A Justification

1. In a DTR, we are predicting the optimal treatment regime, from observed data.
2. Suppose that the outcome is such that

$$E[Y|A, A^*, X] = X'\beta + A \cdot X'\psi.$$

ITTs for DTRs: A Justification

1. In a DTR, we are predicting the optimal treatment regime, from observed data.
2. Suppose that the outcome is such that

$$E[Y|A, A^*, X] = X'\beta + A \cdot X'\psi.$$

3. Optimal treatment depends only on $\text{sign}(X'\psi)$.

ITTs for DTRs: A Justification

1. In a DTR, we are predicting the optimal treatment regime, from observed data.
2. Suppose that the outcome is such that

$$E[Y|A, A^*, X] = X'\beta + A \cdot X'\psi.$$

3. Optimal treatment depends only on sign$(X'\psi)$.
4. Under certain assumptions, a "naive" OLS estimator for ψ has

$$\widehat{\psi}(X, A^*) \xrightarrow{p} \psi^* = \psi \left[P(A = 1|A^* = 1) - P(A = 1|A^* = 0)\right].$$

ITTs for DTRs: A Justification

1. In a DTR, we are predicting the optimal treatment regime, from observed data.
2. Suppose that the outcome is such that

$$E[Y|A, A^*, X] = X'\beta + A \cdot X'\psi.$$

3. Optimal treatment depends only on $\text{sign}(X'\psi)$.
4. Under certain assumptions, a "naive" OLS estimator for ψ has

$$\widehat{\psi}(X, A^*) \xrightarrow{p} \psi^* = \psi \left[P(A = 1|A^* = 1) - P(A = 1|A^* = 0)\right].$$

5. This quantity is such that $\text{sign}(X'\psi^*) = \text{sign}(X'\psi)$.

This is only true under the exclusion restriction, with constant non-adherence probabilities, in the single-stage setting.

Additional Problems with Nonadherence in DTRs

Stable Treatments

Do treatment options remain binary?

Reported Treatments

What if we do not measure prescribed treatment?

If a patient with $A^* = 1$ is non-adherent, is that the same as a patient $A = 0$?

When considering nonadherence, it becomes possible that new treatment options are introduced.

Additional Problems with Nonadherence in DTRs

Stable Treatments

Do treatment options remain binary?

Reported Treatments

What if we do not measure prescribed treatment?

We have focused on prescribed treatment as it compares to actual treatment. What about reported treatment?

$$A^* \longrightarrow A \longrightarrow A^\dagger.$$

Summary

To justify use of an ITT, we need to answer:

1. Is it the right causal effect or is the model simple enough to correspond with treatment-efficacy?
2. Does the SUTVA still hold, even when considering the adherence data?
3. Does the NUC assumption still hold, even when considering the reported data?

If not, we need another alternative.

Proposed Solution: A modified version of G-estimation

Treatment Prescription

Alter the treatment model to be a treatment-prescription model.

Additional Model

Add in a fourth component for adherence probabilities.

Alter Existing Models

Change the existing models to be computable, given observed data.

Instead of specifying a treatment model, giving the probability of receiving treatment, we specify a treatment-prescription model, giving the probability of treatment assignment.

Proposed Solution: A modified version of G-estimation

Treatment Prescription

Alter the treatment model to be a treatment-prescription model.

Additional Model

Add in a fourth component for adherence probabilities.

Alter Existing Models

Change the existing models to be computable, given observed data.

In addition to the blip, treatment-free, and treatment-prescription models, we must now also specify a misclassification model.

Proposed Solution: A modified version of G-estimation

Treatment Prescription

Alter the treatment model to be a treatment-prescription model.

Additional Model

Add in a fourth component for adherence probabilities.

Alter Existing Models

Change the existing models to be computable, given observed data.

When using treatment indicators in the existing models, we need to update the terms to be conditioned on treatment assignment instead. This involves making use of the misclassification model throughout.

G-Estimation with Nonadherence: Mathematically

The blip parameters are estimated by solving

$$\sum_{i=1}^{n} \lambda_K^*(H_{K,i}^*) \left\{ A_{K,i}^* - \underbrace{\pi_K^*(A_{K,i}^*; \alpha_K^*)}_{\text{Treatment Prescription Model}} \right\}$$

$$\times \left\{ Y_i - \underbrace{P(A_{K,i} = 1 | A_{K,i}^*, H_{K,i})}_{\text{Misclassification Model}} \underbrace{C_K^*(H_{K,i}^*; \psi_K^*)}_{\text{Blip Model}} - \underbrace{f_K^*(H_{K,i}^*; \beta_K^*)}_{\text{Treatment-Free Model}} \right\} = 0.$$

Modified G-Estimation

1. The resulting estimators are doubly robust.

Modified G-Estimation

1. The resulting estimators are doubly robust.
2. Can make use of either treatment prescription or reported treatment.

Modified G-Estimation

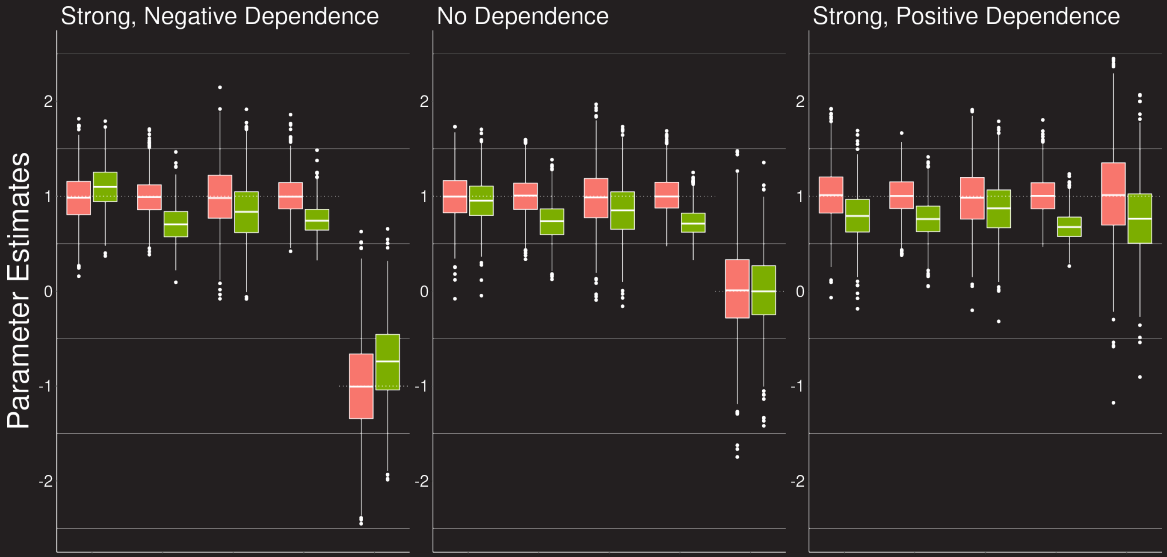
1. The resulting estimators are doubly robust.
2. Can make use of either treatment prescription or reported treatment.
3. Easily accommodates multiple treatments for those who are non-adherent.

Modified G-Estimation

1. The resulting estimators are doubly robust.
2. Can make use of either treatment prescription or reported treatment.
3. Easily accommodates multiple treatments for those who are non-adherent.
4. Flexibility in the estimation of misclassification models. Can make use of:
   - Internal validation;
   - External validation;
   - Sensitivity analyses.

Modified G-Estimation

1. The resulting estimators are doubly robust.
2. Can make use of either treatment prescription or reported treatment.
3. Easily accommodates multiple treatments for those who are non-adherent.
4. Flexibility in the estimation of misclassification models. Can make use of:
   - Internal validation;
   - External validation;
   - Sensitivity analyses.
5. Provides estimates of both treatment efficacy and adherence probabilities, allowing ITT results to be recovered.


Strong, Negative Dependence
No Dependence
Strong, Positive Dependence
Parameter Estimates
2
1
0
-1
-2
A1
X1:A1
A2
A2:X2
A1:A2

Drawbacks to Modified G-Estimation

Despite the benefits to the modified G-estimation approach, there are further considerations to make.

Drawbacks to Modified G-Estimation

Despite the benefits to the modified G-estimation approach, there are further considerations to make.

Complexity

The modified method requires more complexity in the modelling than the standard approach.

Drawbacks to Modified G-Estimation

Despite the benefits to the modified G-estimation approach, there are further considerations to make.

Complexity

The modified method requires more complexity in the modelling than the standard approach.

Extra Modelling

Both the blip and misclassification models need to be correct for valid estimation.

Drawbacks to Modified G-Estimation

Despite the benefits to the modified G-estimation approach, there are further considerations to make.

Complexity	Extra Modelling	Independence Assumptions
The modified method requires more complexity in the modelling than the standard approach.	Both the blip and misclassification models need to be correct for valid estimation.	There are formal independence assumptions required for causal conclusions, which may be violated.

Conclusions

ITTs play an important role, but should be considered critically.

Approaches that estimate treatment efficacy directly are possible but may require further modelling, complicating the assumptions.

Thank You!

www.dylanspicker.com | dylan.spicker@uwaterloo.ca

- Robins, James M. (2004). "Optimal Structural Nested Models for Optimal Sequential Decisions". In: Proceedings of the Second Seattle Symposium in Biostatistics: Analysis of Correlated Data Ed. by D. Y. Lin and P. J. Heagerty. New York, NY: Springer New York, pp. 189–326. ISBN: 978-1-4419-9076-1. DOI: 10.1007/978-1-4419-9076-1_11. URL: https://doi.org/10.1007/978-1-4419-9076-1_11.
- Sheiner, Lewis B. and Donald B. Rubin (1995). "Intention-to-treat analysis and the goals of clinical trials". In: Clinical Pharmacology &amp: Therapeutics 57.1, pp. 6–15. DOI: 10.1016/0009-9236(95)90260-0. URL: https://doi.org/10.1016%2F0009-9236%2895%2990260-0.
- Shrier, Ian, Evert Verhagen, and Steven D. Stovitz (2017). "The Intention-to-Treat Analysis Is Not Always the Conservative Approach". In: The American Journal of Medicine 130.7, pp. 867–871. DOI: 10.1016/j.amjmed.2017.03.023. URL: https://doi.org/10.1016%2Fj.amjmed.2017.03.023.

Theorem 7.5.1 (Consistency of Modified G-Estimation)

Theorem

Suppose that for $j = 1, \ldots, K$ *and* $i = 1, \ldots, n$, *we know* $P(A^*_{i,j}|H^*_{i,j})$ *and* $\pi^*_j(H^*_{i,j}, A^*_{i,j})$, *and we correctly specify the form of* $C^*_j(H^*_{i,j}; \psi_j)$. *Then the* $\widehat{\psi}_j$ *which are estimated by solving* $U^*_j(\widehat{\psi}_j) = 0$ *are consistent for the true* ψ_j, *under the following independence assumptions (I.A.):*

I.A. (1): $E[V_{j+1}(H_j)|H_j, A_j, \overline{A}^*_j] = E[V_{j+1}(H_j)|H_j, A_j]$ *for all* $j = 1, \ldots, K$.

I.A. (2): $E[C_j(H_j)|A_j = 1, H^*_j, A^*_j] = E[C_j(H_j)|H^*_j, A^*_j]$ *for all* $j = 1, \ldots, K$.

I.A. (3): $E[\nu_j(H_j)|H^*_j, A^*_j] = E[\nu_j(H_j)|H^*_j]$ *for all* $j = 1, \ldots, K$.

Theorem 7.7.1 (Asymptotic Normality of Modified G-Estimation)

Theorem (Asymptotic Normality of Modified G-Estimation)

Suppose that for $j = 1, \ldots, K$ *and* $i = 1, \ldots, n$, *we consistently estimate* $P(A^*_{i,j}|H^*_{i,j})$ *and* $\pi^*_j(H^*_{i,j}, A^*_{i,j})$ *through corresponding unbiased estimating equations, and we correctly specify the form of* $C^*_j(H^*_{i,j}; \psi_j)$. *Then the* $(\widehat{\psi}_1, \ldots, \widehat{\psi}_K)$ *which are estimated as components when solving* $U^*_j = 0$ *(Equation (??)) are asymptotically normal, under the independence assumptions from Theorem 1, and the regularity conditions set out by Robins 2004 surrounding exceptional laws. Denoting* $\widehat{\Psi} = (\widehat{\psi}_1, \ldots, \widehat{\psi}_K)$, *we get that, as* $n \to \infty$,

$$\sqrt{n}\left(\widehat{\Psi} - \Psi\right) \xrightarrow{d} N\left(0, \Sigma_\Psi\right).$$

Here $\Sigma_\Psi = I_\Psi \Sigma_\Theta I_\Psi$, I_Ψ *is the diagonal matrix with* 1*'s on the diagonal entries corresponding to the locations of the* Ψ *parameters in* Θ, Θ *is the solution to* $E[U^*(\Theta)] = 0$, *and* Σ_Θ *is sandwich variance matrix based on* U^*.

Pseudo Outcome Justification (Nonadherence)

In the event that there are no treatment indicators in the blip function, then the blip function is exactly known when ψ_j is known, and $A_{i,j}^{\text{opt}}$ will be correctly specified. Suppose that for C_{k+1}, only A_k is involved in the computation. Then, knowing the form of C_{k+1} we can say that

$$
\begin{aligned}
&E\left[A_{k+1}^{\text{opt}} C_{k+1}(H_{k+1}) \middle| H_{k+1}^*, A_{k+1}^*\right] \\
&= P(A_k = 1 | H_{k+1}^*, A_{k+1}^*) E\left[A_{k+1}^{\text{opt}} C_{k+1}(H_{k+1}) \middle| H_{k+1}^*, A_{k+1}^*, A_k = 1\right] \\
&\quad + \left(1 - P(A_k = 1 | H_{k+1}^*, A_{k+1}^*)\right) E\left[A_{k+1}^{\text{opt}} C_{k+1}(H_{k+1}) \middle| H_{k+1}^*, A_{k+1}^*, A_k = 0\right] \\
&= \pi_k^*(H_{k+1}^*) I\left\{C_{k+1}(H_{k+1}^*, A_k = 1) > 0\right\} C_{k+1}(H_{k+1}^*, A_k = 1) \\
&\quad + \left[1 - \pi_k^*(H_{k+1}^*)\right] I\left\{C_{k+1}(H_{k+1}^*, A_k = 0) > 0\right\} C_{k+1}(H_{k+1}^*, A_k = 0).
\end{aligned}
$$

We can take

$$
\begin{aligned}
\widetilde{V}_j &= \widetilde{V}_{j+1} + \pi_{j-1}^*(H_j^*) I\left\{C_j(H_j^*, A_j = 1) > 0\right\} C_j(H_j^*, A_j = 1) \\
&\quad + \left[1 - \pi_{j-1}^*(H_j^*)\right] I\left\{C_j(H_j^*, A_j = 0) > 0\right\} C_j(H_j^*, A_j = 0) - \pi_j^*(H_{i,j}^*) C_j^*(H_{i,j}^*).
\end{aligned}
$$