

Semi-parametric Bayesian Inference for Optimal Dynamic Treatment Regimes via Dynamic

Marginal Structural Models

Daniel Rodriguez Duque

daniel.rodriguezduque@mail.mcgill.ca

Department of Epidemiology, Biostatistics, and Occupational Health, McGill University, Montreal, Canada.

and

David A. Stephens

david.stephens@mcgill.ca

Department of Mathematics and Statistics, McGill University, Montreal, Canada

and

Erica E.M. Moodie

erica.moodie@mcgill.ca

Department of Epidemiology, Biostatistics, and Occupational Health, McGill University, Montreal, Canada

and

Marina B. Klein

marina.klein@mcgill.ca

Department of Medicine, McGill University, Montreal, Canada

Web Appendix A Technical Details

No Unmeasured Confounders Assumption: Consider the *unobserved history* up to time k , $\mathcal{F}_k = \{(y, z_t, x_t, u_t), t = 1, \dots, k\}$, where u_t are unobserved covariates. Furthermore, consider *observed history* up to time k is given by $\mathcal{H}_k = \{(y, z_t, x_t), t = 1, \dots, k\}$. Then, the sequence of treatments $\{z_t\}$ is unconfounded relative to latent variables $\{u_t\}$ if for each k , z_k and $\{u_t, t = 1, \dots, k\}$ are conditionally independent given (\mathcal{H}_{k-1}, x_k) . Mathematically, this may be written as $p_{\mathcal{O}}(z_k | \mathcal{F}_{k-1}, u_k, x_k) = p_{\mathcal{O}}(z_k | \mathcal{H}_{k-1}, x_k)$, $k = 1, \dots, K$.

De Finetti Representation: Below we consider a more general form of the De Finetti representation presented in the main paper. We do this by considering the vector $(y_i, \bar{x}_i, \bar{z}_i, u_i)$, where u_i are determinants of the outcome and intermediate variables. We assume that these vectors are infinitely exchangeable in order to deduce the de Finetti representation in the observational world:

$$p_{\mathcal{O}}(b_1, \dots, b_n) = \int_{\phi, \gamma, \tau} \prod_{i=1}^n \left[\int_u p_{\mathcal{O}}(y_i | \bar{x}_i, \bar{z}_i, u_i, \tau) \right. \\ \prod_{j=1}^K p_{\mathcal{O}}(x_{ij} | \bar{z}_{i(j-1)}, \bar{x}_{i(j-1)}, u_i, \phi_{1j}) p_{\mathcal{O}}(u_i | \phi_2) du_i \\ \left. \prod_{j=1}^K p_{\mathcal{O}}(z_{ij} | \bar{z}_{i(j-1)}, \bar{x}_{ij}, \gamma_j) \right] p(\phi, \gamma) d\tau d\phi d\gamma.$$

The absence of u_i in the treatment assignment probability is due to the no unmeasured confounders assumption. We can also look at the representation in the experimental measure by considering: $v_i = (b_i, g_i) \equiv (y_i, \bar{x}_i, \bar{z}_i, g_i)$, and assuming infinite exchangeability in order to obtain

$$p_{\mathcal{E}}(v_1, \dots, v_n) = \int \prod_{i=1}^n \left[\int_u p_{\mathcal{E}}(y_i | \bar{x}_i, \bar{z}_i, g_i, u_i, \tau) \right. \\ \prod_{j=1}^K p_{\mathcal{E}}(x_{ij} | z_{i(j-1)}, x_{i(j-1)}, u_i, g_i, \phi_{1j}) p_{\mathcal{E}}(u_i | \phi_2) du_i \\ \left. \prod_{j=1}^K p_{\mathcal{E}}(z_{ij} | z_{i(j-1)}, x_{i(j-1)}, g_i, \alpha_j) p(g_i) \right] p(\phi, \alpha) d\tau d\phi d\alpha.$$

Change of Measure Details Corresponding to Equation (2.3):

Let us first see how to fully develop the importance sampling argument, and then how to obtain the

form of the weights. We connect the experimental world with the observational world as follows:

$$\begin{aligned}
E_{\mathcal{E}}[U(b^*, g, \beta)|\bar{b}] &= E_{G_{\mathcal{E}}} \left[E_{b_{\mathcal{E}}|g}[U(b^*, g, \beta)|g, \bar{b}] \middle| \bar{b} \right] \\
&= E_{G_{\mathcal{E}}} \left[\int_{b^*} U(b^*, g, \beta) p_{\mathcal{E}}(b^*|g, \bar{b}) \frac{p_{\mathcal{O}}(b^*|\bar{b})}{p_{\mathcal{O}}(b^*|\bar{b})} \middle| \bar{b} \right] \\
&= E_{G_{\mathcal{E}}} \left[E_{\mathcal{O}} \left[U(b^*, g, \beta) \frac{\mathbb{1}_{g(\bar{x}^*)}(\bar{z}^*)}{\prod_{k=1}^K p_{\mathcal{O}}(z_k^*|\bar{z}_{k-1}^*, \bar{x}^*, \bar{b})} \middle| \bar{b} \right] \middle| \bar{b} \right] \\
&= E_{\mathcal{O}} \left[\frac{\frac{1}{C_G} \sum_{\{r \in \mathcal{I}\}} U(b^*, g^r, \beta) \mathbb{1}_{g^r(\bar{x}^*)}(\bar{z}^*)}{\prod_{k=1}^K p_{\mathcal{O}}(z_k^*|\bar{z}_{k-1}^*, \bar{x}^*, \bar{b})} \middle| \bar{b} \right] \\
&= E_{\mathcal{O}} \left[\frac{1}{C_G} \sum_{\{r \in \mathcal{I}\}} w^{*r} U(\bar{b}^*, g^r, \beta) \middle| \bar{b} \right].
\end{aligned}$$

Now let us examine how we may obtain the weights w^* for DTR-MSMs. Note that we need only consider the single-stage problem, as the multi-stage case follows directly.

$$\begin{aligned}
p_{\mathcal{E}}(Y = y, Z = z, X = x|G = g) &= \frac{p_{\mathcal{E}}(Y = y, Z = z, X = x|G = g)}{p_{\mathcal{O}}(Y = y, Z = z, X = x)} p_{\mathcal{O}}(Y = y, Z = z, X = x) \\
&= \frac{p_g(Y = y, Z = z, X = x)}{p_{\mathcal{O}}(Y = y, Z = z, X = x)} p_{\mathcal{O}}(Y = y, Z = z, X = x) \\
&= \frac{p_g(Y = y, g(X) = z, X = x)}{p_{\mathcal{O}}(Y = y|Z = z, X = x) p_{\mathcal{O}}(Z = z|X = x) p_{\mathcal{O}}(X = x)} p_{\mathcal{O}}(Y = y, Z = z, X = x) \\
&= \frac{p_g(Y = y|g(X) = z, X = x) p_g(g(X) = z|X = x) p_g(X = x)}{p_{\mathcal{O}}(Y = y|Z = z, X = x) p_{\mathcal{O}}(Z = z|X = x) p_{\mathcal{O}}(X = x)} p_{\mathcal{O}}(Y = y, Z = z, X = x)
\end{aligned}$$

Note that in the above argument, when we condition on $g(X) = z, X = x$, we may run into issues if $g(x)$ does not equal z . However, in practice this is not a concern as the joint probability $p_{\mathcal{E}}(Y = y, g(X) = z, X = x)$ would take the value zero in such a situation, and so this term would not contribute to the calculation. Continuing, we find:

$$\begin{aligned}
p_{\mathcal{E}}(Y = y, Z = z, X = x|G = g) &= \frac{p_g(Y = y|g(X) = z, X = x) p_g(g(X) = z|X = x)}{p_{\mathcal{O}}(Y = y|Z = z, X = x) p_{\mathcal{O}}(Z = z|X = x)} p_{\mathcal{O}}(Y = y, Z = z, X = x) \\
&= \frac{p_g(Y = y|g(X) = z, X = x) \mathbb{1}_{g(x)}(z)}{p_{\mathcal{O}}(Y = y|Z = z, X = x) p_{\mathcal{O}}(Z = z|X = x)} p_{\mathcal{O}}(Y = y, Z = z, X = x).
\end{aligned}$$

Now, we are looking for cancellation between the outcome probabilities. We have already established that when $g(x) \neq z$, the numerator is equal to zero. When $g(x) = z$, we have that $p_g(Y = y|g(X) = z, X = x) = p_{\mathcal{O}}(Y = y|Z = z, X = x)$. Thus we may finish by writing:

$$\begin{aligned} p_{\mathcal{E}}(Y = y, Z = z, X = x|G = g) \\ = \frac{\mathbb{1}_{g(x)}(z)}{p_{\mathcal{O}}(Z = z|X = x)} p_{\mathcal{O}}(Y = y, Z = z, X = x), \end{aligned}$$

yielding the weights that we were seeking.

Web Appendix B Discussion on Non-Regularity in DTRs

We note that the arguments presented in this paper are Bayesian. Thus, conditional on the posited model, the resulting inference is valid for any sample size. We emphasize that the premise of the Bayesian bootstrap is not related to attaining asymptotic consistency, but simply it is about proposing a specific model for the data, and carrying out inference conditional on this model. That being said, we may still ask how well we would expect these methods to perform as more data are observed. As noted in the main paper, the parameters of dynamic MSMs can be shown to be consistent (Orellana *and others*, 2010a; van der Laan and Petersen, 2007). We mainly make use of the estimator for the value of a specific DTR, and this is also asymptotically normal and regular as laid out by Murphy *and others* (2001). In what follows, we emphasize that estimation of dynamic MSMs do not suffer from non-regularity as is the case with other methods, like Q-Learning, G-estimation of structural nested mean models, and dynamic weighted ordinary least squares (dWOLS) (Wallace and Moodie, 2015).

We proceed by discussing the relevant literature on non-regularity in order to understand why it does not play a role in the estimation of parameters in dynamic MSMs. Additionally, we present a simulation that illustrates our point. Our simulation is similar in spirit to that of Chakraborty *and others* (2010), where we draw 1000 bootstrapped samples and evaluate whether the obtained coverage differs significantly from the nominal 95%. As we expect, the parameters in the MSM do not exhibit issues with non-regularity.

It was Robins (2004) who first raised the issue of non-regularity in methods aimed at estimating

parameters relevant to identifying optimal DTRs. The key issue is illustrated in the context of estimating the absolute value of a population mean, $|\mu|$, from n *i.i.d.* observations. A maximum likelihood approach may first estimate the mean $\hat{\mu}$, and then this may be plugged into $|\cdot|$ to obtain an estimator for $|\mu|$. What Robins (2004) emphasizes is that $|\hat{\mu}|$ has different asymptotic distributions depending on the value of μ (when $\mu = 0$ *vs.* $\mu \neq 0$). This is what yields a non-regular estimator, and the crux of this issue is in the fact that the absolute value function is discontinuous at zero. Consequently, Wald-type confidence intervals do not perform well. Chakraborty *and others* (2010) examine whether bootstrap confidence intervals yield appropriate inference in non-regular settings, but they point out that the success of the bootstrap relies on the smoothness of the estimator. Accordingly, one should not expect the bootstrap to provide adequate inference at or near the point of non-regularity.

For Q-learning, it is clear where non-regularity arises. Consider a two-stage setting where the stage II model is $y_i = \gamma_{20} + \gamma_{21}z_1 + \gamma_{22}x_1z_1 + \gamma_{23}z_2 + \gamma_{24}x_2z_2$. The stage I pseudo-outcome becomes $\tilde{y}_i = \gamma_{20} + \gamma_{21}z_1 + \gamma_{22}x_1z_1 + \mathbb{1}(\gamma_{23}z_2 + \gamma_{24}x_2z_2 > 0)$. This pseudo-outcome is discontinuous at $\gamma_{23}z_2 + \gamma_{24}x_2z_2 = 0$. Therefore, we should expect that plugging-in $\hat{\gamma}_{20}, \hat{\gamma}_{21}, \hat{\gamma}_{22}, \hat{\gamma}_{23}, \hat{\gamma}_{24}$ to compute $\hat{\tilde{y}}_i$ will cause issues with the estimation of stage I parameters, as these will depend on a discontinuous function of other parameters. Non-regularity is not only an issue at $\gamma_{23}z_2 + \gamma_{24}x_2z_2 = 0$ but also near it; Chakraborty *and others* (2010) explored this via simulation and found non-regularity to impact inference. Earlier works also noted non-regularity to arise in G-estimation (Moodie and Richardson, 2010).

The parameters in dynamic MSMs do not suffer from the above-mentioned issues. Unlike G-estimation, DWOLS, and Q-learning, dynamic MSMs do not require recursively solving estimating equations, where the stage I equation has plug-in estimators obtained by solving a stage II estimating equation. Therefore, for dynamic MSMs, the estimators are not functions of discontinuous functions of other estimators. Ultimately, this means that the parameters in dynamic MSMs do not suffer from the same types of difficulties with non-regularity. Let us now examine an example in which non-regularity impacts inference in Q-learning but plays no role in the inference of parameters in dynamic MSMs. We consider a family of regimes that says treat if $x_k > \theta_k$ for $k = 1, 2$. The proposed data-generating mechanism is one that allows for straightforward marginalization so

that we can pose a correct model for $E[Y^{\theta_1\theta_2}]$. The outcome is given by:

$$Y = \gamma_0 + \gamma_1 z_1 + \gamma_2 x_1 z_1 + \gamma_3 z_2 + \gamma_4 x_2 z_2 + \epsilon \quad (1)$$

Variables are distributed as: $x_1 \sim N(0, 9)$, $x_2 \sim N(0, 4)$, $z_1, z_2 \sim \text{bern}(0.5)$. Then,

$$E[Y^{\theta_1\theta_2}] = \gamma_0 + \gamma_1 C_{11}(\theta_1) + \gamma_2 C_{12}(\theta_1) + \gamma_3 C_{21}(\theta_2) + \gamma_4 C_{22}(\theta_2) \quad (2)$$

where,

$$\begin{aligned} C_{21}(\theta_2) &= E[\mathbb{1}(x_2 > \theta_2)|x_1, z_1] = p(x_2 > \theta_2), \\ C_{22}(\theta_2) &= E[x_2 \mathbb{1}(x_2 > \theta_2)|x_1, z_1] = \frac{4}{\sqrt{2\pi}} \exp(-\theta_2^2/(2 \cdot 4^2)). \end{aligned}$$

C_{11}, C_{12} have an analogous form. Then, we have an analytic form for the marginal model. We assume further that $\gamma_3, \gamma_4 > 0$ and consider the following scenarios:

- Scenario I: $\gamma_0 = 1, \gamma_1 = 1, \gamma_2 = 1, \gamma_3 = 0, \gamma_4 = 0$.
- Scenario II: $\gamma_0 = 1, \gamma_1 = 1, \gamma_2 = 1, \gamma_3 = 0.001, \gamma_4 = 0.001$.
- Scenario III: $\gamma_0 = 1, \gamma_1 = 1, \gamma_2 = 1, \gamma_3 = 1, \gamma_4 = 1$.

Scenario I explores inference in a non-regular setting; scenario II explores a near non-regular setting, and scenario III explores a regular setting. We make use of $B = 1000$ bootstrap samples, a sample size of $n = 1000$, and $R = 500$ replications. We first examine these scenarios in the context of Q-learning. The correctly specified models that we fit are as follows:

$$\textit{Stage I} : \gamma_{10} + \gamma_{11} z_1 + \gamma_{12} x_1 z_1$$

$$\textit{Stage II} : \gamma_{20} + \gamma_{21} z_1 + \gamma_{22} x_1 z_1 + \gamma_{23} z_2 + \gamma_{24} x_2 z_2$$

The pseudo-outcome in stage I is: $\gamma_{20} + \gamma_{21} z_1 + \gamma_{22} x_1 z_1 + (\gamma_{23} + \gamma_{24} x_2) \mathbb{1}(\gamma_{23} + \gamma_{24} x_2 > 0)$. Note that because of the specific data-generating mechanism, these models are correctly specified.

Web Table 1 shows that, as expected, the parameters for the stage II model present no evidence of non-regularity as measured by coverage or bias. We note that apart from the point estimates,

stage II inference is the same for all scenarios, hence the shorter table. From Web Table 2, we see where the non-regularity becomes present. The stage I intercept exhibits coverage that is significantly different from nominal in the non-regular case. This persists even in the close-to-non-regular setting. Furthermore, as is shown in supplementary Web Table 5, evidence of non-regularity disappears in a gradient, as the data-generating mechanism gets further from the completely non-regular setting.

Web Table 1: Scenario I Coverage of 95% CI for Q-learning stage II parameters. $B = 1000; n = 1000; R = 500$.

Parameter	Coverage	Mean	Bias	SD
γ_{20}	0.946	0.9997	-0.0003	0.0112
γ_{21}	0.958	1.0004	0.0004	0.0124
γ_{22}	0.952	1.0001	0.0001	0.0029
γ_{23}	0.940	0.0000	0.0000	0.0131
γ_{24}	0.938	0.0000	0.0000	0.0045

*indicates significant difference from 0.95

Web Table 2: Coverage of 95% CI for Q-learning stage I parameters $\gamma_{10}, \gamma_{11}, \gamma_{12}$. $B = 1000; n = 1000; R = 500$.

Parameter	$\gamma_3 = \gamma_4$	Coverage	Estimate	Bias	SD
γ_{10}	0	0.884*	1.0059	0.0059	0.0098
γ_{11}		0.958	1.0004	0.0004	0.0124
γ_{12}		0.952	1.0001	0.0001	0.0030
γ_{10}	0.001	0.898*	1.0065	0.0051	0.0098
γ_{11}		0.958	1.0004	0.0004	0.0124
γ_{12}		0.952	1.0001	0.0001	0.0030
γ_{10}	1	0.944	2.3997	0.0041	0.0653
γ_{11}		0.954	0.9948	-0.0052	0.0927
γ_{12}		0.954	0.9998	-0.0002	0.0213

*indicates significant difference from 0.95

The Q-learning results are only presented for the frequentist bootstrap, as the use of the Bayesian bootstrap has not been studied in this literature. In the following, we look at the resulting inference for the Frequentist and Bayesian dynamic MSMs. The θ used to create the augmented data required for these methods are $\{-4, -2.5, -1, 0.5, 2, 3.5\}$. As expected, there are no issues with any coverage probabilities; this can be seen in Web Tables 3 and 4.

Web Table 3: Results frequentist dynamic MSM; $B = 1000$; $n = 1000$; $R = 500$.

Parameter	$\gamma_3 = \gamma_4$	Coverage	Estimate	Bias	SD
γ_0	0	0.954	1.0038	0.0038	0.4194
γ_1		0.952	0.9597	-0.0403	1.5826
γ_2		0.948	0.9913	-0.0087	0.7859
γ_3		0.958	0.0439	0.0439	1.2823
γ_4		0.944	-0.0046	-0.0046	0.8172
γ_0	0.001	0.954	1.0037	0.0037	0.4194
γ_1		0.952	0.9597	-0.0403	1.5826
γ_2		0.948	0.9913	-0.0087	0.7859
γ_3		0.958	0.0448	0.0438	1.2823
γ_4		0.944	-0.0036	-0.0046	0.8172
γ_0	1	0.956	0.9921	-0.0079	0.5344
γ_1		0.946	0.9944	-0.0056	2.0336
γ_2		0.958	1.0081	0.0081	0.9960
γ_3		0.946	1.0114	0.0114	1.6185
γ_4		0.956	0.9900	-0.0100	1.0225

*indicates significant difference from 0.95

Web Table 4: Results Bayesian dynamic MSM. $B = 1000$; $n = 1000$; $R = 500$

Parameter	$\gamma_3 = \gamma_4$	Coverage	Estimate	Bias	SD
γ_0	0.000	0.950	0.9922	-0.0078	0.4183
γ_1		0.934	1.0303	0.0303	1.6094
γ_2		0.952	1.0099	0.0099	0.7619
γ_3		0.938	-0.0270	-0.0270	1.2993
γ_4		0.952	-0.0092	-0.0092	0.7893
γ_0	0.001	0.950	0.9922	-0.0078	0.4182
γ_1		0.934	1.0302	0.0302	1.6095
γ_2		0.952	1.0099	0.0099	0.7618
γ_3		0.938	-0.0260	-0.0270	1.2994
γ_4		0.952	-0.0082	-0.0092	0.7892
γ_0	1.000	0.966	1.0130	0.0130	0.5264
γ_1		0.956	0.9613	-0.0387	2.0571
γ_2		0.956	0.9826	-0.0174	0.9608
γ_3		0.958	1.0305	0.0305	1.6333
γ_4		0.952	1.0102	0.0102	0.9893

*indicates significant difference from 0.95

In what follows, we examine how inference is impacted as $\gamma_3 = \gamma_4$ get further away from the non-regular case. For Web Table 5, we see that as we get further from non-regularity, the closer to nominal coverage becomes in Q-learning. Note that only results for the γ_{10} parameter are shown as this is the parameter that most clearly exhibits issues with non-regularity in Q-learning. From

Web Tables 6 and 7, we see that the frequentist and Bayesian bootstrap yield adequate inference with the dynamic MSM, regardless of proximity to the non-regular case.

Web Table 5: Results of Q-Learning for different levels of non-regularity; $B = 500, n = 1000, R = 500$.

Parameter	$\gamma_{23} = \gamma_{24}$	p-value	Coverage	Mean Estimate	Bias	SD
γ_{10}	0	< 0.001	0.854	1.0064	0.0064	0.0099
γ_{10}	0.001	< 0.001	0.858	1.0069	0.0056	0.0098
γ_{10}	0.005	< 0.001	0.906	1.0101	0.0031	0.0097
γ_{10}	0.010	0.031	0.928	1.0156	0.0017	0.0097
γ_{10}	0.050	0.051	0.930	1.0700	0.0002	0.0104
γ_{10}	0.100	0.473	0.942	1.1397	0.0001	0.0121
γ_{10}	1.000	0.356	0.940	2.3963	0.0007	0.0672

Web Table 6: Frequentist dynamic MSM; $B = 500, n = 1000, R = 500$.

	$\gamma_2 = \gamma_4$	p-value	Coverage	Mean Estimate	Bias	SD
γ_0	0	0.356	0.960	0.9777	-0.0223	0.4193
γ_1		0.608	0.944	1.1461	0.1461	1.5732
γ_2		0.758	0.954	1.0355	0.0355	0.7879
γ_3		0.608	0.944	-0.1176	-0.1176	1.2878
γ_4		0.608	0.944	-0.0476	-0.0476	0.8202
γ_0	0.001	0.356	0.960	0.9777	-0.0223	0.4193
γ_1		0.608	0.944	1.1461	0.1461	1.5732
γ_2		0.758	0.954	1.0355	0.0355	0.7879
γ_3		0.608	0.944	-0.1165	-0.1175	1.2878
γ_4		0.608	0.944	-0.0465	-0.0475	0.8202
γ_0	0.005	0.356	0.960	0.9778	-0.0222	0.4193
γ_1		0.608	0.944	1.1459	0.1459	1.5730
γ_2		0.758	0.954	1.0352	0.0352	0.7879
γ_3		0.608	0.944	-0.1124	-0.1174	1.2875
γ_4		0.758	0.946	-0.0423	-0.0473	0.8202
γ_0	0.010	0.259	0.962	0.9780	-0.0220	0.4193
γ_1		0.608	0.944	1.1457	0.1457	1.5728
γ_2		0.758	0.954	1.0350	0.0350	0.7879
γ_3		0.608	0.944	-0.1073	-0.1173	1.2872
γ_4		0.758	0.946	-0.0370	-0.0470	0.8203
γ_0	0.050	0.259	0.962	0.9791	-0.0209	0.4198
γ_1		0.758	0.946	1.1441	0.1441	1.5725
γ_2		0.918	0.948	1.0328	0.0328	0.7887
γ_3		0.608	0.944	-0.0663	-0.1163	1.2857
γ_4		0.918	0.948	0.0053	-0.0447	0.8214
γ_0	0.100	0.259	0.962	0.9804	-0.0196	0.4210
γ_1		0.918	0.952	1.1421	0.1421	1.5743
γ_2		0.758	0.946	1.0302	0.0302	0.7909
γ_3		0.608	0.944	-0.0149	-0.1149	1.2856
γ_4		0.608	0.944	0.0582	-0.0418	0.8240
γ_0	1.000	0.918	0.952	1.0052	0.0052	0.5519
γ_1		0.608	0.956	1.1056	0.1056	1.9877
γ_2		0.473	0.958	0.9821	-0.0179	1.0391
γ_3		0.356	0.960	0.9086	-0.0914	1.5885
γ_4		0.608	0.956	1.0101	0.0101	1.0772

Web Table 7: Bayesian dynamic MSM; $B = 500, n = 1000, R = 500$.

Parameter	$\gamma_3 = \gamma_4$	p-val	percent	Estimate	Bias	SD
γ_0	0	0.051	0.930	0.9857	-0.0143	0.4521
γ_1		0.259	0.962	1.0905	0.0905	1.5410
γ_2		0.081	0.932	1.0211	0.0211	0.8449
γ_3		0.608	0.956	-0.0709	-0.0709	1.2483
γ_4		0.356	0.940	-0.0363	-0.0363	0.8702
γ_0	0.001	0.051	0.930	0.9857	-0.0143	0.4521
γ_1		0.259	0.962	1.0906	0.0906	1.5411
γ_2		0.081	0.932	1.0211	0.0211	0.8449
γ_3		0.608	0.956	-0.0699	-0.0709	1.2484
γ_4		0.356	0.940	-0.0353	-0.0363	0.8702
γ_0	0.005	0.051	0.930	0.9856	-0.0144	0.4521
γ_1		0.259	0.962	1.0909	0.0909	1.5414
γ_2		0.081	0.932	1.0212	0.0212	0.8449
γ_3		0.608	0.956	-0.0661	-0.0711	1.2486
γ_4		0.356	0.940	-0.0314	-0.0364	0.8702
γ_0	0.010	0.051	0.930	0.9855	-0.0145	0.4521
γ_1		0.259	0.962	1.0912	0.0912	1.5418
γ_2		0.081	0.932	1.0212	0.0212	0.8449
γ_3		0.608	0.956	-0.0613	-0.0713	1.2489
γ_4		0.356	0.940	-0.0264	-0.0364	0.8701
γ_0	0.050	0.051	0.930	0.9847	-0.0153	0.4524
γ_1		0.356	0.960	1.0941	0.0941	1.5459
γ_2		0.081	0.932	1.0218	0.0218	0.8455
γ_3		0.473	0.958	-0.0233	-0.0733	1.2519
γ_4		0.473	0.942	0.0131	-0.0369	0.8704
γ_0	0.100	0.051	0.930	0.9838	-0.0162	0.4534
γ_1		0.356	0.960	1.0977	0.0977	1.5531
γ_2		0.051	0.930	1.0225	0.0225	0.8473
γ_3		0.259	0.962	0.0243	-0.0757	1.2573
γ_4		0.182	0.936	0.0624	-0.0376	0.8719
γ_0	1.000	0.051	0.930	0.9668	-0.0332	0.5651
γ_1		0.918	0.948	1.1624	0.1624	2.0210
γ_2		0.356	0.940	1.0355	0.0355	1.0660
γ_3		0.918	0.948	0.8805	-0.1195	1.6245
γ_4		0.259	0.938	0.9507	-0.0493	1.0883

Web Appendix C Considerations for Double Robust Estimator

In this section, we present additional details related to ideas discussed in Section 3 of the main paper. This includes details about how to fit outcome models in the double robust estimator.

Web Appendix C.1 Outcome Models

In this section, we provide details about how to fit outcome models for the double robust estimator.

Recall that for $k = K$, ϕ_{K+1}^* is defined as

$$\phi_{K+1}^*(\bar{x}_K^*) = E_{\mathcal{O}}[y^* | \bar{x}_K^*, \bar{z}_K^* = \bar{g}_K(\bar{x}_K), \bar{b}],$$

and for $k = K - 1, \dots, 1$, ϕ_{k+1}^* is defined as

$$\phi_{k+1}^*(\bar{x}_k^*) = E_{\mathcal{O}}[\phi_{k+2}^*(\bar{x}_{k+1}) | \bar{x}_k^*, \bar{z}_k^* = \bar{g}_k(\bar{x}_k^*), \bar{b}].$$

First, note that based on the prior we have selected (which yields the non-parametric Bayesian bootstrap as the posterior), it is enough to fit these models on the observed data, conditional on a draw from the Dirichlet weights. In a regression setting, the weights would just be incorporated into the *weights* argument in the *lm* function. We now focus on how to pose these models, based on the data generating mechanism in the single threshold simulation, which can be found in Appendix C. The outcome is generated via $y = x_1 - (-\theta^{opt} + x_1)(\mathbb{1}_{x_1 > \theta^{opt}} - z_1) - (-\theta^{opt} + x_2)(\mathbb{1}_{x_2 > \theta^{opt}} - z_2) + \sqrt{0.5}\epsilon$. Note that θ^{opt} is a constant and $\epsilon \sim N(0, 1)$. Then, we may look to fit the following model:

$$\begin{aligned} E[y | \bar{x}, \bar{z}] = & \beta_{21}x_1 + \beta_{22}\mathbb{1}_{x_1 > \theta^{opt}} + \beta_{23}z_1 + \beta_{24}x_1\mathbb{1}_{x_1 > \theta^{opt}} + \beta_{25}x_1z_1 \\ & + \beta_{26}\mathbb{1}_{x_2 > \theta^{opt}} + \beta_{27}z_2 + \beta_{28}x_2\mathbb{1}_{x_2 > \theta^{opt}} + \beta_{29}x_2z_2. \end{aligned} \tag{3}$$

We use this model to compute $\psi_2 = E[y | \bar{x}, z_1, z_2 = g(x_2)]$. We then seek to fit a model conditional on just stage one information. This requires marginalizing over x_2 when $z_2 = g(x_2)$ in equation 3.

For this, we need to compute a few quantities:

$$\begin{aligned}
1) E[\mathbb{1}_{x_2 > \theta^{opt}} | x_1, z_1] &= p(\theta^{opt} - z_1 - 0.5x_1 < \epsilon | x_1, z_1) \\
&= 1 - \Phi(\theta^{opt} - z_1 - 0.5x_1) \\
&:= T_1(x_1, z_1) \\
2) E[g(x_2) | x_1, z_1] &= E[\mathbb{1}_{x_2 > \theta} | x_1, z_1] \\
&= 1 - \Phi(\theta - z_1 - 0.5x_1) \\
&:= T_2(x_1, z_1) \\
3) E[x_2 \mathbb{1}_{x_2 > \theta^{opt}} | x_1, z_1] &= E[(z_1 + 0.5x_1 + \epsilon) \mathbb{1}_{\theta^{opt} < z_1 + 0.5x_1 + \epsilon} | x_1, z_1] \\
&:= T_3(x_1, z_1) \\
4) E[x_2 g(x_2) | x_1, z_1] &= E[(z_1 + 0.5x_1 + \epsilon) \mathbb{1}_{\theta < z_1 + 0.5x_1 + \epsilon} | x_1, z_1] \\
&:= T_4(x_1, z_1)
\end{aligned}$$

T_1, T_2, T_3 may be approximated numerically through quick draws of a normal distribution. This leads us to the model:

$$\begin{aligned}
E[\psi_2 | x_1, z_1] &= \beta_{11}x_1 + \beta_{12}\mathbb{1}_{x_1 > \theta^{opt}} + \beta_{13}z_1 + \beta_{14}x_1 \mathbb{1}_{x_1 > \theta^{opt}} + \beta_{15}x_1 z_1 \\
&+ \beta_{16}T_1(x_1, z_1) + \beta_{17}T_2(x_1, z_1) + \beta_{18}T_3(x_1, z_1) + \beta_{19}T_4(x_1, z_1).
\end{aligned}$$

When $\theta = \theta^{opt}$, then we have $T_1 = T_2$ and $T_3 = T_4$, and so two of these terms must be taken out of the model in this special case. Note that marginalization becomes slightly complex as, x_2 depends on both z_1 and x_1 . If it only dependent on z_1 which is binary, things would be simplified as stage 1 terms would absorb any marginalization terms. Of course, in practice it is difficult to correctly specify these models, but one would hope that specifying a flexible enough model would lead to improved results with regard to efficiency. Once these two models have been fit, we may compute

$$\begin{aligned}
\phi_2(x_1) &= \beta_{11}x_1 + \beta_{12}\mathbb{1}_{x_1 > \theta^{opt}} + \beta_{13}g(x_1) + \beta_{14}x_1 \mathbb{1}_{x_1 > \theta^{opt}} + \beta_{15}x_1 g(x_1) + \beta_{16}T_1(x_1, g(x_1)) \\
&+ \beta_{17}T_2(x_1, g(x_1)) + \beta_{18}T_3(x_1, g(x_1)) + \beta_{19}T_4(x_1, g(x_1))
\end{aligned}$$

and

$$\begin{aligned}\phi_3(\bar{x}_2) &= \beta_{21}x_1 + \beta_{22}\mathbb{1}_{x_1 > \theta^{opt}} + \beta_{23}g(x_1) + \beta_{24}x_1\mathbb{1}_{x_1 > \theta^{opt}} + \beta_{25}x_1g(x_1) \\ &\quad + \beta_{26}\mathbb{1}_{\theta^{opt} > x_2} + \beta_{27}g(x_2) + \beta_{28}x_2\mathbb{1}_{\theta^{opt} > x_2} + \beta_{29}x_2g(x_2).\end{aligned}$$

We then use these last two expressions in the double robust estimator.

Web Appendix C.2 Bayesian Double Robustness

If we are able to show the equivalence between expressions (10) and (11) in the main article, then we will have demonstrated the double robustness property. Consider:

$$\begin{aligned}\phi_2^*(\bar{x}_0^*) &+ \sum_{k=2}^K w_{k-1}^*(\phi_{k+1}^*(\bar{x}_k^*) - \phi_k^*(\bar{x}_{k-1}^*)) + w_K^*(y^* - \phi_{K+1}^*(\bar{x}_K^*)) \\ &= \phi_2^*(\bar{x}_0^*) + w_K^*(y^* - \phi_{K+1}^*(\bar{x}_K^*)) + \sum_{k=2}^K w_{k-1}^*\phi_{k+1}^*(\bar{x}_k^*) - \sum_{k=1}^{K-1} w_k^*\phi_{k+1}^*(\bar{x}_k^*) \\ &= \phi_2^*(\bar{x}_0^*) + w_K^*(y^* - \phi_{K+1}^*(\bar{x}_K^*)) + w_{K-1}^*\phi_{K+1}^*(\bar{x}_{K-1}^*) - w_1^*\phi_2^*(\bar{x}_1^*) - \sum_{k=2}^{K-1} (w_k^* - w_{k-1}^*)\phi_{k+1}^*(\bar{x}_k^*) \\ &= w_K^*y^* - \sum_{k=1}^K (w_k^* - w_{k-1}^*)\phi_{k+1}^*(\bar{x}_k^*) - \sum_{k=2}^{K+1} w_{k-1}^*(h(\bar{B}) - h(\bar{b})) \\ &= w_K^*y^* - w_K^*h(\bar{b}) + w_0^*h(\bar{b}) - \sum_{k=1}^K (w_k^* - w_{k-1}^*)\phi_{k+1}^*(\bar{x}_k^*) + \sum_{k=1}^K (w_k^* - w_{k-1}^*)h(\bar{b}) \\ &= h(\bar{b}) + w_K^*(y^* - h(\bar{b})) - \sum_{k=1}^K (w_k^* - w_{k-1}^*)(\phi_{k+1}^*(\bar{x}_k^*) - h(\bar{b})),\end{aligned}$$

recalling that $w_0^* = 0$ and that $h(\bar{b}) = E_g[y^*|\bar{b}]$. From the first expression we may see that this is an unbiased estimator when the conditional means are correctly specified. This is obtained from an iterated expectation argument, and by showing that $E[w_{k-1}^*(\phi_{k+1}^*(\bar{x}_k^*) - \phi_k^*(\bar{x}_{k-1}^*))|\bar{x}_{k-1}] = 0$. The full argument can be seen in Orellana *and others* (2010b). The last expression allows us to see this is unbiased when the treatment assignment models are correctly specified by noting that $E[w_k^*|\bar{x}_k^*, \bar{z}_{k-1}] = w_{k-1}^*$ and by again using an iterated expectation.

Web Appendix D Simulation Details

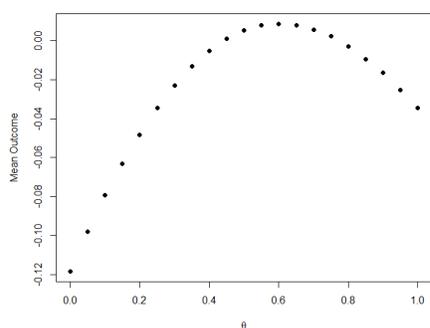
This appendix explores inference for the following regimes types: 1) single threshold regimes, 2) double threshold regimes, 3) weighted regimes, and 4) weighted regimes where the threshold depends on a binary baseline covariate. Simulations 2) and 4) are those in the main paper.

Web Appendix D.1 Thresholding DTRs

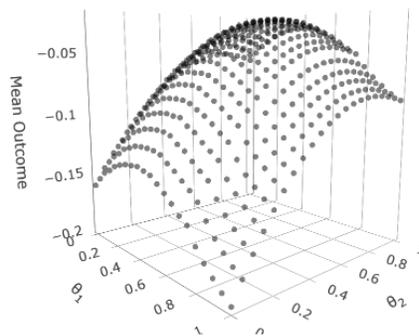
The data generating mechanism is given by:

- $x_1 \sim N(0, 1)$, $x_2 \sim N(0, z_1 + 0.5x_1)$
- $z_1 \sim \text{Bern}(p = \text{expit}(1.5x_1))$, $z_2 \sim \text{Bern}(p = \text{expit}(2x_2 - 0.5z_1))$
- $y = x_1 - (-\theta_{1opt} + x_1)(\mathbb{1}_{\theta_{1opt} > z_1} - z_1) - (-\theta_{2opt} + x_2)(\mathbb{1}_{\theta_{2opt} > z_2} - z_2) + \sqrt{0.5}\epsilon$, $\epsilon \sim N(0, 1)$
and $\theta_{1opt}, \theta_{2opt}$ the location of the desired optima. In the single threshold simulation, we have that $\theta_{1opt} = \theta_{2opt} = \theta_{opt}$.

Note that “expit” is the inverse logit function. For the out of sample prediction, we used a population of $n = 10,000$ and $x_1 \sim N(0.6, 1)$, $x_2 \sim N(0.1 + z_1, 1)$. Web Figure 1 plots the expected outcome under the DTRs considered in this section.



(a)



(b)

Web Figure 1: (a) Response surface for single threshold simulation with Normal covariates; (b) Response surface for double threshold simulation with Normal covariates.

Web Appendix D.1.1 Single Threshold Simulation

Web Table 8 shows results for single threshold simulation, under a sample size of $n = 500$. This contrasts the sample size of $n = 1000$ shown in Web Table 9. Here, $\theta_{opt} = 0.6$, and the value at the optimum is 0. Generally, the results follow the same pattern though with an overall loss of precision corresponding to the reduction in sample size.

Web Table 8: Results for single threshold simulation (Normal covariates; $n = 500$; 500 Monte Carlo replicates).

Method	Model Correct	$\hat{\theta}$	Estimated Outcome Train Pop.	Coverage Probability θ	Mean Outcome Test Pop.
Frequentist	None	0.416 (0.110)	0.217 (0.120)	—	0.587 (0.013)
Frequentist	Treat	0.637 (0.189)	0.038 (0.070)	—	0.589 (0.014)
Frequentist	Outcome	0.580 (0.176)	0.015 (0.069)	—	0.591 (0.014)
Frequentist	Both	0.618 (0.159)	0.013 (0.057)	—	0.593 (0.010)
Frequentist	IPW	0.638 (0.183)	0.029 (0.066)	—	0.590 (0.013)
Bayesian	None	0.414 (0.118)	0.232 (0.119)	0.664	0.586 (0.014)
Bayesian	Treat	0.648 (0.201)	0.057 (0.068)	0.976	0.587 (0.015)
Bayesian	Outcome	0.573 (0.188)	0.026 (0.068)	0.980	0.590 (0.016)
Bayesian	Both	0.624 (0.168)	0.021 (0.057)	0.972	0.592 (0.012)
Bayesian	IPW	0.641 (0.196)	0.045 (0.065)	0.976	0.588 (0.015)

Web Table 9: Results for single threshold simulation (Normal covariates; $n = 1000$; 500 Monte Carlo replicates).

Method	Model Correct	$\hat{\theta}$	Estimated Outcome Train Pop.	Coverage Probability θ	Mean Outcome Test Pop.
Frequentist	None	0.419 (0.093)	0.209 (0.084)	—	0.588 (0.011)
Frequentist	Treat	0.635 (0.172)	0.024 (0.047)	—	0.591 (0.013)
Frequentist	Outcome	0.599 (0.122)	0.012 (0.044)	—	0.596 (0.006)
Frequentist	Both	0.608 (0.122)	0.010 (0.038)	—	0.596 (0.007)
Frequentist	IPW	0.624 (0.155)	0.018 (0.045)	—	0.593 (0.011)
Bayesian	None	0.418 (0.097)	0.218 (0.083)	0.516	0.588 (0.012)
Bayesian	Treat	0.642 (0.178)	0.038 (0.045)	0.976	0.590 (0.014)
Bayesian	Outcome	0.597 (0.132)	0.018 (0.044)	0.980	0.595 (0.008)
Bayesian	Both	0.611 (0.128)	0.016 (0.038)	0.972	0.596 (0.008)
Bayesian	IPW	0.634 (0.172)	0.030 (0.044)	0.968	0.591 (0.013)

Note: Standard deviations are Monte Carlo standard deviations

We also investigate the results when intermediary covariates are Gamma-distributed as follows:

$x_1 \sim \text{Gamma}(\alpha = 2, \beta = 2)$, $x_2 \sim \text{Gamma}(\alpha = z_1 + 0.5x_1, \beta = 1)$. The known mean outcome under the optimal threshold is 1 in the training population. In the test population, the distribution of intermediary covariates was changed to be $x_1 = \text{Gamma}(\alpha = 1.5, \beta = 1)$ and $x_2 = \text{Gamma}(\alpha = z_1 + 0.5x_1, \beta = 2)$. The exploration grid was the same as the Normal setup except that the thresholds started at 0.05, given that Gamma covariates are positive. The results mostly parallel the already observed results, see Web Table 10 and 11. Notable is that the resulting credible intervals appear to be slightly more conservative.

Web Table 10: Results for single threshold simulation (Gamma covariates; $n = 500$; 500 Monte Carlo replicates).

Method	Model Correct	$\hat{\theta}$	Estimated Outcome Train Pop.	Coverage Probability θ	Mean Outcome Test Pop.
Frequentist	None	0.127 (0.165)	1.065 (0.058)	—	1.444 (0.016)
Frequentist	Treat	0.609 (0.160)	1.024 (0.054)	—	1.489 (0.011)
Frequentist	Outcome	0.578 (0.192)	1.044 (0.084)	—	1.485 (0.014)
Frequentist	Both	0.624 (0.136)	1.020 (0.047)	—	1.490 (0.010)
Frequentist	IPW	0.654 (0.167)	1.044 (0.059)	—	1.486 (0.015)
Bayesian	None	0.181 (0.219)	1.132 (0.062)	0.774	1.450 (0.021)
Bayesian	Treat	0.632 (0.167)	1.092 (0.052)	0.998	1.487 (0.014)
Bayesian	Outcome	0.614 (0.182)	1.155 (0.087)	1	1.486 (0.013)
Bayesian	Both	0.649 (0.138)	1.080 (0.046)	0.998	1.489 (0.011)
Bayesian	IPW	0.706 (0.173)	1.124 (0.061)	0.962	1.482 (0.018)

Web Table 11: Results for simulation I (Gamma covariates; $n = 1000$; 500 Monte Carlo replicates).

Method	Model Correct	$\hat{\theta}$	Estimated Outcome Train Pop.	Coverage Probability θ	Mean Outcome Test Pop.
Frequentist	None	0.076 (0.079)	1.051 (0.045)	—	1.439 (0.008)
Frequentist	Treat	0.608 (0.131)	1.015 (0.037)	—	1.491 (0.008)
Frequentist	Outcome	0.587 (0.146)	1.027 (0.065)	—	1.490 (0.009)
Frequentist	Both	0.610 (0.106)	1.010 (0.033)	—	1.493 (0.005)
Frequentist	IPW	0.632 (0.153)	1.025 (0.042)	—	1.488 (0.012)
Bayesian	None	0.106 (0.146)	1.085 (0.044)	0.546	1.442 (0.014)
Bayesian	Treat	0.625 (0.131)	1.062 (0.036)	1	1.491 (0.009)
Bayesian	Outcome	0.608 (0.143)	1.107 (0.065)	1	1.490 (0.009)
Bayesian	Both	0.626 (0.111)	1.052 (0.032)	1	1.492 (0.007)
Bayesian	IPW	0.671 (0.160)	1.080 (0.044)	0.974	1.486 (0.014)

Web Appendix D.1.2 Double Threshold Simulation

In Web Table 12 we examine the results of the double threshold simulation with normal covariates and a larger sample size than presented in the main paper. Here, $\theta_{1opt} = 0.4, \theta_{2opt} = 0.8$, and the value at the optimum is 0. There is a general gain in precision due to the larger sample size; additionally, the coverage of the confidence intervals deviates slightly farther from the nominal coverage.

Web Table 12: Results for double threshold simulation (Normal covariates; $n = 1000$; 500 Monte Carlo replicates).

Method	Model Correct	$\hat{\theta}_1$	$\hat{\theta}_2$	Estimated Outcome Train Pop.	Coverage Probability θ_1, θ_2	Mean Outcome Test Pop.
Frequentist	None	0.254 (0.097)	0.677 (0.142)	0.236 (0.086)	—	0.591 (0.008)
Frequentist	Treat	0.470 (0.204)	0.788 (0.175)	0.031 (0.045)	—	0.588 (0.015)
Frequentist	Outcome	0.393 (0.164)	0.783 (0.156)	0.016 (0.043)	—	0.593 (0.008)
Frequentist	Both	0.416 (0.152)	0.801 (0.145)	0.013 (0.037)	—	0.594 (0.007)
Frequentist	IPW	0.443 (0.179)	0.790 (0.180)	0.023 (0.044)	—	0.590 (0.012)
Bayesian	None	0.252 (0.104)	0.682 (0.154)	0.250 (0.085)	0.770, 0.918	0.590 (0.008)
Bayesian	Treat	0.473 (0.217)	0.795 (0.179)	0.047 (0.043)	0.970, 0.988	0.587 (0.016)
Bayesian	Outcome	0.390 (0.171)	0.787 (0.179)	0.026 (0.043)	0.986, 0.992	0.591 (0.010)
Bayesian	Both	0.419 (0.159)	0.809 (0.148)	0.021 (0.037)	0.982, 0.982	0.593 (0.008)
Bayesian	IPW	0.456 (0.191)	0.798 (0.183)	0.036 (0.043)	0.978, 0.988	0.589 (0.014)

Note: Standard deviations are Monte Carlo standard deviations

Next, we can examine the results when intermediary covariates are Gamma-distributed as described in the previous section. Web Tables 13 and 14 show the results for this setup. Overall, we observe that the optimal threshold are unbiasedly estimated, and that credible intervals are somewhat conservative leading to higher coverage probabilities. Part of this is due to the choice of increments: larger increments leading to higher coverage. The value at the optimal thresholds is also unbiased.

Web Table 13: Results for double threshold simulation (Gamma covariates; $n = 500$; 500 Monte Carlo replicates).

Method	Model Correct	$\hat{\theta}_1$	$\hat{\theta}_2$	Estimated Outcome Train Pop.	Coverage Probability θ_1, θ_2	Mean Outcome Test Pop.
Frequentist	None	0.129 (0.074)	0.751 (0.210)	1.145 (0.076)	—	1.481 (0.005)
Frequentist	Treat	0.379 (0.181)	0.791 (0.173)	1.038 (0.049)	—	1.488 (0.010)
Frequentist	Outcome	0.401 (0.211)	0.757 (0.177)	1.055 (0.068)	—	1.485 (0.014)
Frequentist	Both	0.406 (0.168)	0.792 (0.149)	1.024 (0.043)	—	1.490 (0.009)
Frequentist	IPW	0.456 (0.197)	0.785 (0.188)	1.050 (0.052)	—	1.485 (0.016)
Bayesian	None	0.136 (0.087)	0.757 (0.216)	1.197 (0.072)	0.810, 0.974	1.481 (0.005)
Bayesian	Treat	0.393 (0.190)	0.806 (0.177)	1.099 (0.049)	0.998, 0.964	1.487 (0.012)
Bayesian	Outcome	0.446 (0.208)	0.760 (0.186)	1.131 (0.069)	0.994, 0.988	1.484 (0.017)
Bayesian	Both	0.426 (0.171)	0.807 (0.150)	1.076 (0.043)	1.000, 0.994	1.489 (0.010)
Bayesian	IPW	0.494 (0.213)	0.800 (0.186)	1.128 (0.053)	1.000, 0.952	1.482 (0.021)

Web Table 14: Results for double threshold simulation (Gamma covariates; $n = 1000$; 500 Monte Carlo replicates).

Method	Model Correct	$\hat{\theta}_1$	$\hat{\theta}_2$	Estimated Outcome Train Pop.	Coverage Probability θ_1, θ_2	Mean Outcome Test Pop.
Frequentist	None	0.109 (0.034)	0.780 (0.163)	1.131 (0.060)	—	1.480 (0.002)
Frequentist	Treat	0.386 (0.150)	0.805 (0.150)	1.024 (0.037)	—	1.491 (0.006)
Frequentist	Outcome	0.383 (0.180)	0.790 (0.140)	1.035 (0.052)	—	1.489 (0.009)
Frequentist	Both	0.401 (0.125)	0.809 (0.128)	1.014 (0.032)	—	1.493 (0.005)
Frequentist	IPW	0.419 (0.150)	0.784 (0.164)	1.030 (0.040)	—	1.490 (0.009)
Bayesian	None	0.110 (0.038)	0.785 (0.175)	1.159 (0.058)	0.542 0.980	1.480 (0.003)
Bayesian	Treat	0.387 (0.162)	0.821 (0.152)	1.066 (0.036)	1.000 0.970	1.490 (0.007)
Bayesian	Outcome	0.426 (0.188)	0.786 (0.151)	1.088 (0.053)	1.000 0.990	1.488 (0.013)
Bayesian	Both	0.408 (0.136)	0.819 (0.127)	1.049 (0.031)	1.000 0.990	1.492 (0.006)
Bayesian	IPW	0.451 (0.164)	0.805 (0.168)	1.082 (0.040)	1.000 0.962	1.489 (0.011)

An analogous individualized decision rule graph can be produced for the thresholds in this simulation, however this is no more instructive than the figure for the single threshold rule.

Web Appendix D.2 Weighted DTRs Simulation

Next, we explore one additional simulation. For this family of regimes, patients are treated in stage one if $\psi_1 x_{1,1} + \psi_2 x_{1,2} > 0.5$ and in stage two if $\psi_1 x_{2,1} + \psi_2 x_{2,2} > 0.5$. Here, $\psi_1, \psi_2 > 0$ such that $\psi_1 + \psi_2 = 1$. The optimal parameters are chosen to be $\psi_{1opt} = \psi_{2opt} = 0.5$. The response surface in this setting is similar to that in Simulation II. The data generating mechanism proceeds as follows:

- $x_{1,1} \sim N(1, 1)$, $x_{1,2} \sim N(0, 1)$
- $z_1 \sim \text{Bern}(\text{expit}(1.5x_{1,2} + 2x_{1,1}))$
- $x_{2,1} \sim N(0.2z_1 + 0.1x_{1,1}, 1)$, $x_{2,2} \sim N(0.5z_1 + 0.1x_{1,2}, 1)$
- $z_2 \sim \text{Bern}(p = \text{expit}(1.5x_{2,2} - 0.6z_1 + 2x_{2,1}))$
- $z_{1,opt} = 0.5x_{1,1} + 0.5x_{1,2} > 0.5$, $z_{2,opt} = 0.5x_{2,1} + 0.5x_{2,2} > 0.5$,
- $y = x_{11} + x_{12} - (0.5x_{11} + 0.5x_{12} - 0.5)(z_{1,opt} - z_1) - (0.5x_{21} + 0.5x_{22} - 0.5)(z_{2,opt} - z_2) + \sqrt{0.5}\epsilon$,
 $\epsilon \sim N(0, 1)$

The value at the optimal threshold can be seen to be 1. For the test population, we used a population size of $n = 10,000$ and $x_{1,1} \sim N(0.1, 1)$, $x_{1,2} \sim N(0.5, 1)$, $x_{2,1} \sim N(0.1 + 0.2z_1 + 0.1x_{1,1}, 1)$, $x_{2,2} \sim N(0.5 + 0.5z_1 + 0.1x_{1,2}, 1)$. Results are presented in the Web Table 15 and 16, and we observed that we obtain unbiased results. Surprisingly, even with both nuisance models misspecified, the estimator performs quite well in terms of coverage, though it estimates the outcome under the optimal regime with high bias. Note that although there are two-parameters in this decision rule, the condition that $\psi_1 + \psi_2 = 1$, makes it so that it is enough to evaluate the coverage probability of only one parameter; this is also why the Monte Carlo standard errors in the ψ_1, ψ_2 columns are the same.

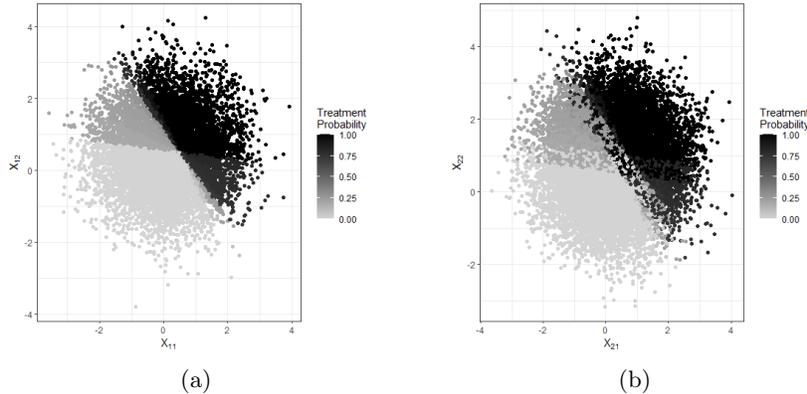
Web Table 15: Frequentist and Bayesian results ($n = 500$; 500 Monte Carlo replicates).

Method	Model Correct	$\hat{\psi}_1$	$\hat{\psi}_2$	Estimated Outcome Train Pop.	Coverage Probability ψ_1	Mean Outcome Test Pop.
Frequentist	None	0.719 (0.251)	0.282 (0.251)	0.208 (0.206)	—	0.509 (0.078)
Frequentist	Treat	0.477 (0.193)	0.523 (0.193)	1.110 (0.274)	—	0.551 (0.045)
Frequentist	Outcome	0.518 (0.122)	0.482 (0.122)	1.022 (0.096)	—	0.571 (0.027)
Frequentist	Both	0.474 (0.117)	0.526 (0.117)	1.038 (0.124)	—	0.571 (0.027)
Frequentist	IPW	0.464 (0.208)	0.536 (0.208)	1.215 (0.559)	—	0.545 (0.049)
Bayesian	None	0.754 (0.258)	0.247 (0.258)	0.258(0.201)	0.944	0.496 (0.078)
Bayesian	Treat	0.473 (0.199)	0.527 (0.199)	1.142(0.272)	0.964	0.550 (0.048)
Bayesian	Outcome	0.523 (0.131)	0.477 (0.131)	1.034(0.095)	0.980	0.569 (0.030)
Bayesian	Both	0.476 (0.119)	0.524 (0.119)	1.047(0.119)	0.968	0.571 (0.027)
Bayesian	IPW	0.460 (0.211)	0.540 (0.211)	1.264(0.563)	0.950	0.544 (0.050)

Web Table 16: Frequentist and Bayesian results ($n = 1000$; 500 Monte Carlo replicates).

Method	Model Correct	$\hat{\psi}_1$	$\hat{\psi}_2$	Estimated Outcome Train Pop.	Coverage Probability ψ	Mean Outcome Test Pop.
Frequentist	None	0.731 (0.249)	0.269 (0.249)	0.184 (0.140)	—	0.507 (0.079)
Frequentist	Treat	0.462 (0.170)	0.538 (0.170)	1.061 (0.140)	—	0.557 (0.041)
Frequentist	Outcome	0.521 (0.104)	0.479 (0.104)	1.015 (0.062)	—	0.575 (0.024)
Frequentist	Both	0.469 (0.109)	0.531 (0.109)	1.021 (0.066)	—	0.573 (0.026)
Frequentist	IPW	0.454 (0.190)	0.545 (0.190)	1.117 (0.242)	—	0.551 (0.046)
Bayesian	None	0.759 (0.250)	0.241 (0.250)	0.217 (0.136)	0.958	0.497 (0.079)
Bayesian	Treat	0.455 (0.173)	0.545 (0.173)	1.086 (0.138)	0.988	0.556 (0.042)
Bayesian	Outcome	0.526 (0.112)	0.474 (0.112)	1.023 (0.062)	0.972	0.573 (0.027)
Bayesian	Both	0.471 (0.107)	0.529 (0.107)	1.029 (0.064)	0.970	0.573 (0.025)
Bayesian	IPW	0.448 (0.193)	0.552 (0.193)	1.156 (0.236)	0.968	0.549 (0.047)

Now, we may examine the individualized inference for this scenario. Web Figure 2 shows us that there are combinations of x_{k1}, x_{k2} where there is high certainty about following the optimal regime and areas of low certainty.



Web Figure 2: Weighted DTR simulation individualized treatment probabilities using double robust estimator; (a) Stage 1 treatment (b) Stage 2 treatment.

Web Appendix D.3 Weighted DTRs with Binary Covariate Simulation

Here, we provide the details for simulation II in the main paper. In this family of regimes, patients are treated if $\psi_1 x_{k1} + \psi_2 x_{k2} > 0.5 - 3\psi_3 u$, $k = 1, \dots, 4$, where $\psi_1 + \psi_2 = 1$, $\psi_1, \psi_2 > 0$. The exploration grid is given by $\psi_1, \psi_2 \in [0.2, 0.8]$ in increments of 0.05 and $\psi_3 \in [-0.3, 0.3]$ in increments of 0.1. This yields a grid of 91 points, with known optima $\psi_{1opt} = 0.5$, $\psi_{2opt} = 0.5$, $\psi_{3opt} = 0.1$. The specific data generating mechanism used is given by:

- $x_{11} \sim N(1, 1)$, $x_{12} \sim N(0, 1)$, $u \sim \text{Bern}(0.5)$, $z_1 \sim \text{Bern}(\text{expit}(0.5x_{12} + x_{11}))$
- $x_{k1} \sim N(0.2z_{k-1} + 0.1x_{k-1,1}, 1)$, $x_{k2} \sim N(0.5z_{k-1} + 0.1x_{k-1,2}, 1)$, $k = 2, 3, 4$
- $z_k \sim \text{Bern}(p = \text{expit}(0.5x_{k2} - 0.6z_{k-1} + x_{k1}))$, $k = 2, 3, 4$
- $z_{k,opt} = 0.5x_{k1} + 0.5x_{k2} + 0.3u > 0.5$, $k = 1, \dots, 4$
- $y = x_{11} + x_{12} - \sum_{k=1}^4 (0.5x_{k1} + 0.5x_{k2} + 0.3u - 0.5)(z_{k,opt} - z_k) + \sqrt{0.1}\epsilon$, $\epsilon \sim N(0, 1)$

Web Table 17 shows the results for a sample size of $n = 1000$. Generally, we observe a gain in precision as compared to the $n = 500$ table in the main paper. Additionally, we note that when all models are correct, we estimate ψ_{3opt} very well. This reflects the fact that the value function is more peaked in this direction as compared to other parameters. For the test population, x_{k1}, x_{k2} were shifted by 0.1 and 0.5, respectively and $u \sim \text{Bern}(0.7)$; we observe that the double robust estimator yields the highest value, as expected.

Web Table 17: Frequentist and Bayesian results ($n = 1000$; 500 Monte Carlo replicates).

Method	Model Correct	$\hat{\psi}_1$	$\hat{\psi}_3$	Estimated Outcome Train Pop.	Coverage Probability ψ_1, ψ_3	Mean Outcome Test Pop.
Freq.	None	0.570 (0.099)	0.092 (0.091)	1.871 (0.282)	—	0.546 (0.051)
Freq.	Treat	0.472 (0.136)	0.107 (0.101)	1.096 (0.111)	—	0.544 (0.051)
Freq.	Outcome	0.503 (0.040)	0.100 (0.004)	1.002 (0.048)	—	0.583 (0.006)
Freq.	Both	0.502 (0.025)	0.100 (0.000)	0.999 (0.045)	—	0.585 (0.004)
Freq.	IPW	0.478 (0.137)	0.099 (0.110)	1.120 (0.139)	—	0.543 (0.051)
Bayes.	None	0.571 (0.108)	0.097 (0.085)	1.995 (0.272)	0.95 0.996	0.558 (0.021)
Bayes.	Treat	0.465 (0.133)	0.105 (0.103)	1.164 (0.100)	0.986 1	0.547 (0.026)
Bayes.	Outcome	0.501 (0.036)	0.100 (0.000)	1.006 (0.049)	0.992 1	0.590 (0.003)
Bayes.	Both	0.499 (0.022)	0.100 (0.000)	1.001 (0.045)	1 1	0.592 (0.002)
Bayes.	IPW	0.459 (0.142)	0.102 (0.105)	1.206 (0.117)	0.984 1	0.544 (0.026)

Web Appendix E Details of the NA-ACCORD Analysis

In what follows, we describe the procedure used to create the data, the analysis plan, the specific models utilized, and we address questions of positivity, individualized inference, and balance.

Web Appendix E.1 Data Creation

Study Start: Study initiation (time zero) is the first instance of ART treatment on or after 2004 in the NA-ACCORD database.

- Study start is not enrollment date as many patients have a long lag between cohort enrollment and ART initiation.

Censoring: Last ART record that has continuous follow-up from study start and that has CD4 and viral load measurements available. This entails the following:

1. There is a monthly ART record from month one up until the month of study exit.
 - Note: some patients have no records for several months and then continuous follow-up resumes. Study exit for these patients is the last month of the first instance of continuous follow-up.
 - There is one exception to the above: If patients have four or fewer months of ART records missing and then continuous follow-up begins again, these months are filled with the last observed treatment. This approach is reasonable as patients do not switch treatment very often.
2. Each record can be associated with a viral load and CD4 cell count measurement.
 - Associate each ART record with CD4 and viral load measurement by taking closest measurement date to ART record date, and using last observation carried forward.
 - With the exception of missing baseline lab values, patients who have missing lab values are censored at the first instance of missingness.
 - Patients who have missing lab values at study start are kept in the study and we create a *status* variable which indicates baseline missingness.

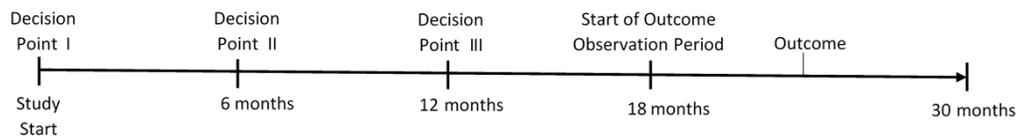
Stage-specific Censoring Details:

- **Stage 1:** Patients lost to follow up after stage 1 covariates are observed but before stage 2 covariates are observed are censored at stage 1.

- **Stage 2:** Patients lost to follow up after stage 2 covariates are observed but before stage 3 covariates are observed are censored at stage 2.
- **Stage 3:** Patients lost to follow-up after stage 3 covariates are observed but before final outcome is observed are censored at stage 3.

Study End: Study end is 18 months after study start; the outcome FIB4 is taken to be the first FIB4 measurement recorded after study end, within 12 months.

Details of the follow-up can be observed in the following diagram:



Web Figure 3: Study Stages

Treatment: We dichotomize all ART treatments to PI based or another ART medication. Some patients receive dual therapy in combination with PI; these are included in “other ART” group.

Treatment Decisions: We consider 6-month observation intervals thereby leading to three treatment decision points: one in the first month of the study, one in the 7th month of the study, and one at 13th month.

Augmented Data Creation

The regimes we explore are of the family: start on a non-PI based ART therapy and switch into PI when $FIB4 > \theta$. Refer to a dataset with information about patients adhering to a regime in this family by R_θ . In addition to the censoring described in the section above, we must take care to keep track of artificial censoring in R_θ . A patient is artificially censored with respect to a regime with threshold θ when they stop adhering to the regime. If they never adhere to the regime, then they are artificially censored at baseline. Adherence to R_θ can be determined based on the following category of patients:

1. Indicated to Switch but did Not Switch (ISNS): Artificial censoring at Indicated switch date.
2. Indicated to Switch and Switched (ISS): No artificial censoring. If patient switches more than

once during the study period, then they are artificially censored at the time of their second switch.

3. Not Indicated to Switch and did Not Switch (NISNS): No artificial censoring.
4. Not Indicated to Switch but Switched (NISS): Artificial censoring at switch date.
5. No Regime (NR): Initial therapy was PI; artificial censoring at baseline.

Note on creating R_θ :

- Each R_θ dataset will contain all patients in the study population. Even patients who are artificially censored at baseline will contribute to fitting outcome models, and toward the fit of the double robust estimator.
- To determine the θ that will be used in the data augmentation, look at the distribution of FIB4 measurements at baseline and create equally spaced increments of 0.2. Based on the data, it turned out that the starting value was 0.4.

Final Datasets: At the end of the above data creation we should have two datasets:

- DATA in long format constitutes of patients in the study population up until their censoring or the study end date. This dataset does not contain any variables that reference regime adherence.
- AUGDATA is the stacked R_θ datasets. Each R_θ datasets is a long-format dataset of patients who adhere to regime R_θ with threshold θ , for the full follow-up period. Each of these dataset have an additional variable providing the regime *index* θ .

Web Appendix E.2 Analysis

For simplicity, we first describe the frequentist analysis, and then describe the Bayesian adaptation.

Treatment Propensity Models: Use DATA to fit a logistic regression model for each stage. Possible time-varying confounders include CD4 Cell count and Viral load. These variables are certainly used to assign treatment, and were proxies for level of HIV infection. There is some evidence

to suggest that HIV is associated with decreased liver health. Therefore, these variables may also mediate previous treatment effects.

Censoring Models: Fit censoring models for each decision point.

Outcome Models: The first conditions on baseline information; the second conditions on information up to stage 2; the third conditions on information up stage 3.

Weight Construction:

- Estimate stage-specific treatment and censoring models.
- For all patients in AUGDATA use the treatment propensity model to compute the probability that they received their observed treatment at each time point.
- Invert each of these probabilities to obtain a weight for each patient for each decision point. Collapse AUGDATA into one observation per patient per regime, and multiply all patient weights in order to create a final weight variable for each patient.

Inverse Probability Weighting Analysis: This analysis is only performed on the subset of cases who are neither censored nor artificially censored. Fit a weighted regression with FIB4 as the outcome and with regime *index* as the predictor. The weights are the ones calculated in the above step. This fit yields the normalized IPW estimator.

Double Robust Analysis: Make use of double robust estimator. This estimator makes use of all observations censored or uncensored (up to the censoring point).

Bayesian Inference Adaptation:

- Draw a vector of Dirichlet weights for as many patients as in DATA. Assign one of these weights to each patient by adding a Dirichlet weight variable to DATA. Note that this variable will not have variation within patients. Additionally, merge these weights into AUGDATA.
- Fit the treatment propensity, censoring, and outcome models as above, but this time incorporate the Dirichlet weights into the fitting. Construct the weights for the collapsed data as before, using the predictions from the treatment propensity model.
- For the IPW analysis, fit the marginal mean model by multiplying the final weights in the collapsed AUGDATA by the Dirichlet weights of each person in AUGDATA.

- For the double robust analysis, run regression, where outcome for each patient is the person-specific contribution to equation (12) in the main paper, and where the predictor is *index*.
- Repeat this over many iterations in order to obtain the posterior distribution of interest.

Analysis Models:

We now specify the models used for analysis.

Censoring Models:

$$\text{Stage 1 : } status + status \times rcs(\log(CD4)) + AgeBaseline + Insurance + AtRiskAlcohol \\ + Sex + Smoking + DrugUse + Race + CalendarYear$$

$$\text{Stage 2 : } rcs(\log(CD4)) + AtRiskAlcohol + Smoking + DrugUse + Race + CalendarYear$$

$$\text{Stage 3 : } rcs(\log(CD4)) + AtRiskAlcohol + Smoking + DrugUse + Race + CalendarYear$$

Treatment Models:

$$\text{Stage 1 : } status + status \times rcs(\log(CD4)) + AgeBaseline + Insurance + AtRiskAlcohol \\ + Sex + HCV + Race + CalendarYear$$

$$\text{Stage 2 : } rcs(\log(CD4)) + Sex + Insurance + HCV + Stage1Treat + Race + CalendarYear$$

$$\text{Stage 3 : } rcs(\log(CD4)) + Sex + Insurance + HCV + Stage2Treat + Race + CalendarYear$$

Note: *rcs* denotes a restricted cubic spline; Stage1Treat denotes stage 1 treatment and Stage2Treat denotes stage 2 treatment. Some patients have missing lab values at baseline; this is indicated by the *status* variable in the models above.

Outcome Models:

$$\text{Stage 1 : } index + index \times (Sex + AgeBaseline + Smoking + DrugUse + HBV + HCV \\ + Insurance + Treat + status \times rcs(\log(CD4)) + status \times rcs(\log(ViralLoad)))$$

$$\text{Stage 2 : } index + index \times (Sex + AgeBaseline + Smoking + DrugUse + HBV + HCV \\ + Insurance + Stage1Treat + Treat + rcs(\log(CD4)) + rcs(\log(ViralLoad)))$$

$$\text{Stage 3 : } Sex + AgeBaseline + Smoking + DrugUse + HBV + HCV + Insurance \\ + Stage1Treat + Stage2Treat + Treat + rcs(\log(CD4)) + rcs(\log(ViralLoad))$$

Note: The *index* variable in the models above is fit as a categorical variable, denoting the regime

index.

Sensitivity Analyses: The following sensitivity analyses were performed:

- Sensitivity Analysis I: All models the same, except that outcome model restricted cubic splines are replaced with $\log(CD4)$ and $\log(ViralLoad)$ terms.
- Sensitivity Analysis II: All models the same, except for outcome model restricted cubic splines are replaced with $rcs(\log(TimeBetween \times CD4))$ and $rcs(\log(TimeBetetween \times ViralLoad))$ terms. This model attempts to account for the fact that not all lab measurements are taken within the same amount of time of the decision point.
- Sensitivity Analysis III: All models the same, except for outcome model restricted cubic splines are replaced with $\log(TimeBetween \times CD4)$ and $\log(TimeBetetween \times ViralLoad)$ terms.

Conclusion of sensitivity analysis: results changed only minimally across models.

Web Appendix E.3 Positivity

Two types of positivity violations are of concern: structural positivity and practical positivity (Petersen *and others*, 2012). The former refers to when patients with specific sets of characteristics are precluded from receiving a treatment; we do not think this is an issue here. The latter refers to the fact that we do not observe all treatments covariate combinations, due to a finite sample size. This is of concern in our setting, as therapeutic switches were infrequent. Zhu *and others* (2021) mention that if propensity scores (PS) are used for achieving balance, then the focus should be on assessing PS overlap between treatment groups. We assessed positivity *for each candidate regime* by checking whether the distribution of the propensity score at each interval for the modeled treatment are similar in the regime adherent group and the regime non-adherent group. This must be done separately for each regime of interest (each θ). In the first stage, all regimes start by evaluating they hypothetical world in which all patients start on a non-PI regimen. Therefore, at this stage the treatment was the probability of receiving PI. For this reason, we only need to perform one comparison across all regimes for this stage (there is no dependence on θ at this stage). We observe that there is overlap from Web Table 18. For the second and third stage, the propensity

of interested was in those who switched treatment. Therefore, we compared the probability that a patient switched into PI in the adherent group vs. the non-adherent group; these comparisons are specific to a threshold θ and are presented for a subset of regimes in Web Table 18. Propensity score overlap indicated that patients who adhered have similar covariate distributions to those who did not adhere. Therefore the types of patients who switch in the regime-enforced world are well represented in the observational world. The propensity to switch treatment was generally small, highlighting that relatively few individuals contribute to the estimation of our regime of interest – a limitation that must be acknowledged.

Web Table 18: Propensity score overlap between patients who adhered to a specific regime and patients who did not adhere for a subset of regimes. (Adh.=“Adherent”)

Regime θ	Group	0%	10%	25%	50%	75%	90%	100%
	Adh. Stage 1	0.198	0.440	0.515	0.606	0.693	0.747	0.835
	Non-Adh. Stage 1	0.179	0.383	0.453	0.537	0.628	0.698	0.820
0.4	Adh. Stage 2	0.012	0.018	0.023	0.037	0.056	0.071	0.119
0.4	Non-Adh. Stage 2	0.011	0.018	0.023	0.035	0.052	0.067	0.130
0.4	Adh. Stage 3	0.006	0.010	0.013	0.018	0.030	0.041	0.050
0.4	Non-Adh. Stage 3	0.006	0.010	0.012	0.017	0.027	0.036	0.062
1.0	Adh. Stage 2	0.012	0.017	0.022	0.033	0.051	0.067	0.123
1.0	Non-Adh. Stage 2	0.011	0.019	0.024	0.037	0.053	0.070	0.130
1.0	Adh. Stage 3	0.006	0.010	0.013	0.018	0.027	0.037	0.062
1.0	Non-Adh. Stage 3	0.006	0.011	0.013	0.020	0.029	0.037	0.062
1.6	Adh. Stage 2	0.011	0.018	0.022	0.034	0.051	0.066	0.123
1.6	Non-Adh. Stage 2	0.012	0.020	0.026	0.039	0.056	0.073	0.130
1.6	Adh. Stage 3	0.006	0.010	0.013	0.019	0.028	0.037	0.062
1.6	Non-Adh. Stage 3	0.007	0.011	0.014	0.021	0.030	0.039	0.062
2.2	Adh. Stage 2	0.011	0.018	0.022	0.034	0.051	0.067	0.123
2.2	Non-Adh. Stage 2	0.012	0.021	0.028	0.041	0.059	0.075	0.130
2.2	Adh. Stage 3	0.006	0.010	0.013	0.019	0.028	0.037	0.062
2.2	Non-Adh. Stage 3	0.007	0.011	0.015	0.022	0.032	0.040	0.061
2.8	Adh. Stage 2	0.011	0.018	0.022	0.034	0.051	0.066	0.123
2.8	Non-Adh. Stage 2	0.012	0.021	0.029	0.043	0.062	0.077	0.130
2.8	Adh. Stage 3	0.006	0.010	0.013	0.019	0.028	0.037	0.062
2.8	Non-Adh. Stage 3	0.007	0.012	0.016	0.023	0.032	0.042	0.061

Web Appendix E.4 Normalization of Weights

In real data analyses, the variability of the estimators is an important consideration. One approach to arrive at more robust estimates is to use weight normalization, as this can reduce the variability of the resulting weights. A discussion of weight normalization can be found in Chapter 12 of Hernán

and Robins (2020), and it has been further explored in the literature for example in Xiao *and others* (2010). For a sample of Dirichlet weights $\pi = (\pi_1, \dots, \pi_n)$, the normalized IPW estimator for the value of a regime g^r is:

$$\frac{\sum_{i=1}^n \frac{\pi_i \mathbb{1}_{g^r(\bar{x}_i)}(\bar{z}_i) y_i}{\prod_{j=1}^K p_{\mathcal{O}}(z_{ij} | \bar{z}_{ij-1}, \bar{x}_{ij-1})}}{\sum_{i=1}^n \frac{\pi_i \mathbb{1}_{g^r(\bar{x}_i)}(\bar{z}_i)}{\prod_{j=1}^K p_{\mathcal{O}}(z_{ij} | \bar{z}_{ij-1}, \bar{x}_{ij-1})}}$$

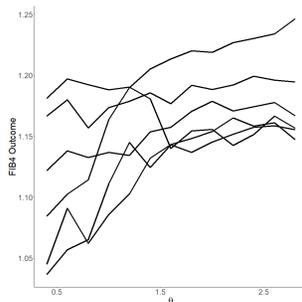
Taking the expectation in the numerator and the denominator across Π , yields the familiar frequentist estimator. The same approach can be taken with the weights in the double robust estimator.

Web Appendix E.5 Balance Diagnostics

Next, we assess the balance obtained from the resulting weighting. We used standardized mean differences to assess balance. Web Table 20 shows the treatment balance assessment at each stage, using the full weights. Some standardized mean differences are moderately large, even after weighting, but this must be considered in the context of having a finite sample size and several probabilities contributing to the weighting of each observation.

Web Appendix E.6 Results for Individualized Inference

By looking at Figure 3 in the main paper, it may be tempting to conclude that there is no benefit to tailoring. This is actually not the case. We remind the reader that we are after the computation: $\theta_{min} = \operatorname{argmin}(E_{\theta_1}[Y], \dots, E_{\theta_{13}}[Y])$. From Web Figure 4, we note that across draws of Π , the expected outcome under regime θ follows a predictable pattern. That is, for small θ the outcome tends to be lower than for high values of θ . We conclude that Figure 3 in the main paper does not display all necessary information.



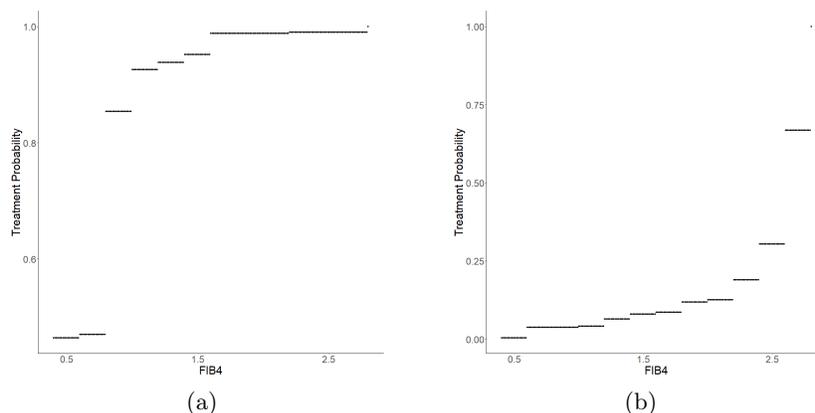
Web Figure 4: Values for six different samples of the posterior distribution

To enrich our analysis, we consider the posterior distribution of two types of θ : one is θ_{min} , which was the original target and which is thought to minimize end-stage FIB4; the second is θ_{max} , which corresponds to the worst decision rule we can obtain by maximizing end-stage FIB4. We now see from Web Table 19 that the outcome-minimizing and outcome-maximizing threshold are not equiprobable. Consequently this does allow us to consider individualized inference, though we should realize that even if we can identify an optimal threshold, it is still clear that the expected change in final FIB4 is minimal and therefore the resulting optimal decision rule will have limited clinical value.

Web Table 19: Posterior Distribution of outcome minimizing/maximizing regimes (500 posterior draws).

Threshold	0.4	0.6	0.8	1.0	1.2	1.4	1.6	1.8	2	2.2	2.4	2.6	2.8
θ_{min}	232	3	192	36	6	7	18	0	0	1	0	0	5
θ_{max}	2	17	0	2	11	8	3	16	4	32	57	182	166

From Web Figure 5(a), we see that the when a patient’s FIB4 score is at 0.8 or greater, they should switch into PI if they hope to follow optimal therapy. From figure 5(b), we see that we should be careful regarding when to switch into PI. Operationalizing a rule that says switch when FIB4 is greater than 2.6 means that we might actually be following the least optimal regime. Of course, we remind the reader that the difference in effect size that each of these regimes yield is small.



Web Figure 5: Cases study individualized treatment probabilities using double robust estimator; (b) Treatment based on θ_{min} (a) Treatment based on θ_{max} .

Web Table 20: Balance diagnostics on the weighted sample: NA-ACCORD.

Stage 1	No PI	PI	SMD
n	7438.8	5182.2	
Smoking (%)	4112.9 (55.3)	2888.7 (55.7)	0.009
At Risk Alcohol (%)	1971.3 (26.5)	1478.1 (28.5)	0.045
Drug Use (%)	1495.4 (20.1)	992.7 (19.2)	0.024
Sex (%)	1258.4 (16.9)	1240.3 (23.9)	0.175
Age at Baseline (mean (SD))	40.07 (11.05)	40.89 (10.56)	0.076
Race Group(%)			0.067
Black	2874.1 (38.6)	1875.4 (36.2)	
Missing	533.6 (7.2)	444.5 (8.6)	
Other	405.3 (5.4)	273.6 (5.3)	
White	3625.7 (48.7)	2588.7 (50.0)	
Insurance (%)	3148.4 (42.3)	1946.2 (37.6)	0.097
HCV at Baseline (%)	736.3 (9.9)	772.8 (14.9)	0.153
HBV at Baseline (%)	381.5 (5.1)	359.6 (6.9)	0.076
Stage 2	No PI	PI	SMD
n	7138.2	5482.8	
Smoking (%)	3921.4 (54.9)	2921.8 (53.3)	0.033
At Risk Alcohol (%)	1871.6 (26.2)	1466.8 (26.8)	0.012
Drug Use (%)	1381.3 (19.4)	900.8 (16.4)	0.076
Sex (%)	1267.1 (17.8)	1333.8 (24.3)	0.162
Age at Baseline (mean (SD))	40.77 (11.02)	41.41 (10.37)	0.060
Race Group (%)			0.100
Black	2792.0 (39.1)	2063.4 (37.6)	
Missing	500.8 (7.0)	535.6 (9.8)	
Other	386.5 (5.4)	301.1 (5.5)	
White	3458.9 (48.5)	2582.8 (47.1)	
Insurance (%)	2959.3 (41.5)	1975.0 (36.0)	0.112
HCV at Baseline (%)	723.5 (10.1)	847.7 (15.5)	0.160
HBV at Baseline (%)	366.9 (5.1)	398.0 (7.3)	0.088
Stage 3	No PI	PI	SMD
n	7156.6	5464.4	
Smoking (%)	3946.3 (55.1)	2895.3 (53.0)	0.043
At Risk Alcohol(%)	1863.6 (26.0)	1445.0 (26.4)	0.009
druguse (%)	1393.1 (19.5)	884.4 (16.2)	0.086
Sex(%)	1287.5 (18.0)	1365.2 (25.0)	0.171
Age at Baseline (mean (SD))	40.78 (11.00)	41.24 (10.34)	0.043
Race Group (%)			0.105
Black	2792.9 (39.0)	2111.3 (38.6)	
Missing	504.9 (7.1)	541.4 (9.9)	
Other	391.9 (5.5)	298.8 (5.5)	
White	3466.9 (48.4)	2513.0 (46.0)	
Insurance(%)	3010.2 (42.1)	1940.8 (35.5)	0.135
HCV at Baseline (%)	732.6 (10.2)	844.8 (15.5)	0.157
HBV at Baseline (%)	372.2 (5.2)	387.4 (7.1)	0.079

Web Appendix F Acknowledgements

The content of this manuscript is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The NA-ACCORD was supported by National Institutes of Health grants U01AI069918, F31AI124794, F31DA037788, G12MD007583, K01AI093197, K01AI131895, K23EY013707, K24AI065298, K24AI118591, K24DA000432, KL2TR000421, N01CP01004, N02CP055504, N02CP91027, P30AI027757, P30AI027763, P30AI027767, P30AI036219, P30AI050409, P30AI050410, P30AI094189, P30AI110527, P30MH62246, R01AA016893, R01DA011602, R01DA012568, R01 AG053100, R24AI067039, U01AA013566, U01AA020790, U01AI038855, U01AI038858, U01AI068634, U01AI068636, U01AI069432, U01AI069434, U01DA03629, U01DA036935, U10EY008057, U10EY008052, U10EY008067, U01HL146192, U01HL146193, U01HL146194, U01HL146201, U01HL146202, U01HL146203, U01HL146204, U01HL146205, U01HL146208, U01HL146240, U01HL146241, U01HL146242, U01HL146245, U01HL146333, U24AA020794, U54MD007587, UL1RR024131, UL1TR000004, UL1TR000083, Z01CP010214 and Z01CP010176; contracts CDC-200-2006-18797 and CDC-200-2015-63931 from the Centers for Disease Control and Prevention, USA; contract 90047713 from the Agency for Healthcare Research and Quality, USA; contract 90051652 from the Health Resources and Services Administration, USA; grants CBR-86906, CBR-94036, HCP-97105 and TGF-96118 from the Canadian Institutes of Health Research, Canada; Ontario Ministry of Health and Long Term Care; and the Government of Alberta, Canada. Additional support was provided by the National Institute Of Allergy And Infectious Diseases (NIAID), National Cancer Institute (NCI), National Heart, Lung, and Blood Institute (NHLBI), Eunice Kennedy Shriver National Institute Of Child Health & Human Development (NICHD), National Human Genome Research Institute (NHGRI), National Institute for Mental Health (NIMH) and National Institute on Drug Abuse (NIDA), National Institute On Aging (NIA), National Institute Of Dental & Craniofacial Research (NIDCR), National Institute Of Neurological Disorders And Stroke (NINDS), National Institute Of Nursing Research (NINR), National Institute on Alcohol Abuse and Alcoholism (NIAAA), National Institute on Deafness and Other Communication Disorders (NIDCD), and National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK).

References

- CHAKRABORTY, BIBHAS, MURPHY, SUSAN A. AND STRECHER, VICTOR. (2010). Inference for non-regular parameters in optimal dynamic treatment regimes. *Statistical Methods in Medical Research* **19**(3), 317–343.
- HERNÁN, MIGUEL A AND ROBINS, JAMES M. (2020). *Causal inference: what if*. Boca Raton: Chapman & Hall/CRC.
- MOODIE, ERICA E.M. AND RICHARDSON, THOMAS S. (2010). Estimating optimal dynamic regimes: Correcting bias under the null. *Scandinavian Journal of Statistics* **37**(1), 126–146.
- MURPHY, SUSAN A., VAN DER LAAN, MARK J. AND ROBINS, JAMES M. (2001). Marginal mean models for dynamic regimes. *Journal of the American Statistical Association* **96**(456), 1410–1423.
- ORELLANA, LILIANA, ROTNITZKY, ANDREA AND ROBINS, JAMES M. (2010a). Dynamic regime marginal structural mean models for estimation of optimal dynamic treatment regimes, part I: Main content. *The International Journal of Biostatistics* **6**(2).
- ORELLANA, LILIANA, ROTNITZKY, ANDREA AND ROBINS, JAMES M. (2010b). Dynamic regime marginal structural mean models for estimation of optimal dynamic treatment regimes, part II: Proof of results. *The International Journal of Biostatistics* **6**(2).
- PETERSEN, MAYA L., PORTER, KRISTIN E., GRUBER, SUSAN, WANG, YUE AND VAN DER LAAN, MARK J. (2012). Diagnosing and responding to violations in the positivity assumption. *Statistical Methods in Medical Research* **21**(1), 31–54.
- ROBINS, JAMES M. (2004). Optimal structural nested models for optimal sequential decisions. In: *Proceedings of the second Seattle Symposium in Biostatistics*. Springer. pp. 189–326.
- VAN DER LAAN, MARK J. AND PETERSEN, MAYA L. (2007). Causal effect models for realistic individualized treatment and intention to treat rules. *The International Journal of Biostatistics* **3**(1).
- WALLACE, MICHAEL P. AND MOODIE, ERICA E.M. (2015). Doubly-robust dynamic treatment regimen estimation via weighted least squares. *Biometrics* **71**(3), 636–644.

XIAO, YONGLING, ABRAHAMOWICZ, MICHAL AND MOODIE, ERICA E.M. (2010). Accuracy of conventional and marginal structural cox model estimators: a simulation study. *The International Journal of Biostatistics* **6**(2).

ZHU, YAQIAN, HUBBARD, REBECCA A., CHUBAK, JESSICA, ROY, JASON AND MITRA, NANDITA. (2021). Core concepts in pharmacoepidemiology: Violations of the positivity assumption in the causal analysis of observational data: Consequences and statistical approaches. *Pharmacoepidemiology and Drug Safety*.