SMART-MC: Characterizing the Dynamics of Multiple Sclerosis Therapy Transitions Using a Covariate-Based Markov Model

Priyam Das

Department of Biostatistics
School of Public Health
Virginia Commonwealth University

BASS XXXII - 32nd Annual Biopharmaceutical Applied Statistics Symposium

November 4, 2025

Overview of Multiple Sclerosis (MS)

- Chronic autoimmune disorder affecting the **central nervous system** (brain and spinal cord).
- Commonly diagnosed in **young adults (20–50 years)**, with higher prevalence in females.
- Symptoms vary widely — including fatigue, weakness, sensory loss, and mobility impairment.
- One of the **leading causes of non-traumatic neurological disability** in young adults worldwide.

MS Disease Subtypes and Progression

- **Relapsing-Remitting MS (RRMS):** Characterized by relapses (episodes of new or worsening symptoms) followed by recovery.
- **Primary Progressive MS (PPMS):** Continuous worsening of neurological function from onset, without distinct relapses.
- **Secondary Progressive MS (SPMS):** Gradual disability accumulation following an initial relapsing-remitting phase.

MS Disease-Modifying Therapies (DMTs)

Over **15 FDA-approved DMTs**, targeting distinct immunological mechanisms. Broadly categorized as:

- **High-efficacy treatment:** Natalizumab, B-cell depleting agents (rituximab, ocrelizumab).
- **Oral therapies:** Dimethyl fumarate, S1P modulators (fingolimod, teriflunomide).
- **Injectable therapies:** Interferon-beta, Glatiramer acetate (first-generation agents).
- **Aggressive/legacy therapies:** Alemtuzumab, mitoxantrone, cyclophosphamide.

Research Questions

- Our goal is to **understand how patient-level covariates influence and are associated with the treatment sequences of MS disease-modifying therapies (DMTs).**

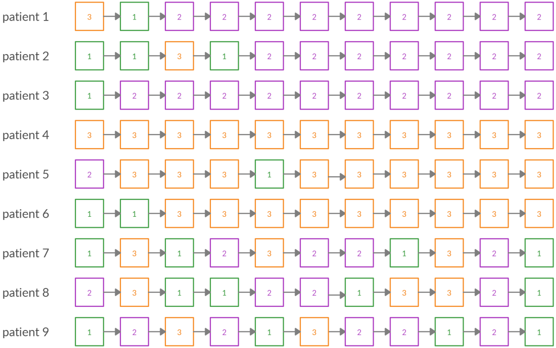
More precise research questions are formulated following the introduction of the modeling framework.

Markov Model

- Markov model is a widely used tool for modeling sequences of events from a finite state-space.
- Despite being popular for modeling sequence data, in literature, it hasn't been explored to it's full potential to model EHR treatment sequence data.

EHR treatment sequence data

- Following diagram shows the treatment prescribed to 9 patients on their first 11 doctor visits.
- Here, each patient is prescribed one of the 3 treatments (given by treatment 1, 2, 3) at each doctor visit.



patient 1	3	1	2	2	2	2	2	2	2	2	2
patient 2	1	1	3	1	2	2	2	2	2	2	2
patient 3	1	2	2	2	2	2	2	2	2	2	2
patient 4	3	3	3	3	3	3	3	3	3	3	3
patient 5	2	3	3	3	1	3	3	3	3	3	3
patient 6	1	1	3	3	3	3	3	3	3	3	3
patient 7	1	3	1	2	3	2	2	1	3	2	1
patient 8	2	3	1	1	2	2	1	3	3	2	1
patient 9	1	2	3	2	1	3	2	2	1	2	1

Markov chain

Example Markov chain with 3 treatments:

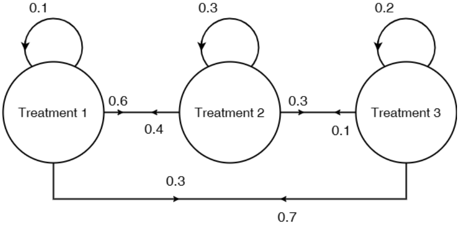


Figure: Markov chain diagram.

	Trt 1	**Trt 2**	**Trt 3**
Trt 1	0.1	0.6	0.3
Trt 2	0.4	0.3	0.3
Trt 3	0.7	0.2	0.1

Figure: Transition matrix.

MS DMT treatment sequence data

Dataset:

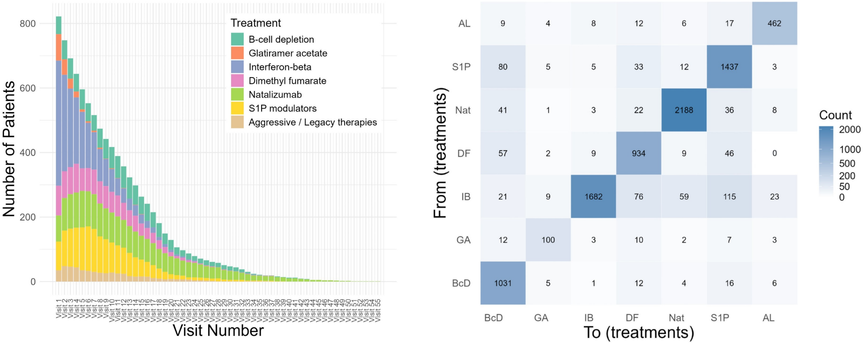
- Multiple Sclerosis (MS) disease-modifying therapy (DMT) sequence data from the `Massachusetts General and Brigham hospital system (Boston, US) EHR cohort`.
- 822 patients, observed over $\sim$ 16 years.
- **7 classes of DMTs** are considered (listed in the next slide).
- Available patient-level covariates: **Age at diagnosis**, **disease duration**, **gender (M/F)**, **race (Black/White/Others)**.

MS DMT treatment sequence data

7 classes of DMTs are given as follows :

1. B-cell depletion (rituximab,ocrelizumab)
2. Glatiramer acetate
3. Interferon-beta
4. Dimethyl fumarate
5. Natalizumab
6. S1P modulators (fingolimod, teriflunomide)
7. Aggressive/Legacy therapies (AL) (cyclophosphamide, mitoxantrone, alemtuzumab)

Summary: EHR treatment sequence data



	BcD	GA	IB	DF	Nat	S1P	AL
AL	9	4	8	12	6	17	462
S1P	80	5	5	33	12	1437	3
Nat	41	1	3	22	2188	36	8
DF	57	2	9	934	9	46	0
IB	21	9	1682	76	59	115	23
GA	12	100	3	10	2	7	3
BcD	1031	5	1	12	4	16	6

(a) (b)

Figure: (a) Stacked alluvial-style plot displaying the longitudinal treatment sequences across visits. (b) Empirical transition matrix heatmap showing the observed counts of transitions between therapies.

Goals

Research questions:

(i) How do **clinical factors**, such as *disease duration*, influence the likelihood of transitioning between first-line injectables (e.g., interferon-beta, glatiramer acetate), oral therapies (e.g., dimethyl fumarate, S1P modulators), and high-efficacy agents (e.g., natalizumab, B-cell depleting therapies)?

Goals

Research questions:

(i) How do **clinical factors**, such as *disease duration*, influence the likelihood of transitioning between first-line injectables (e.g., interferon-beta, glatiramer acetate), oral therapies (e.g., dimethyl fumarate, S1P modulators), and high-efficacy agents (e.g., natalizumab, B-cell depleting therapies)?

(ii) How do **demographic factors** such as *age*, *sex*, and *race/ethnicity* impact treatment sequencing choices?

Goals

Research questions:

(i) How do **clinical factors**, such as *disease duration*, influence the likelihood of transitioning between first-line injectables (e.g., interferon-beta, glatiramer acetate), oral therapies (e.g., dimethyl fumarate, S1P modulators), and high-efficacy agents (e.g., natalizumab, B-cell depleting therapies)?

(ii) How do **demographic factors** such as *age*, *sex*, and *race/ethnicity* impact treatment sequencing choices?

(iii) What are the **most frequent transition pathways** observed in real-world MS care, and which treatment **transitions** are **most sensitive to patient characteristics**?

Existing works: Markov Chains in EHR

Some existing works in this area are listed below:

(i) **Mixture of Markov Chains** (via EM algorithm, Helske and Helske, 2019)

(ii) Classification of *Rheumatoid Arthritis* patients' treatment sequence data with **Mixture of Markov Chains** (via Global optimization; Das et al., 2023b)

(iii) Classification of *Multiple Sclerosis* patients' treatment sequence data with **Covariate dependent Mixture of Markov Chains** (via Global optimization; Das et al., 2023a)

New Model Framework?

- The posed research questions canNOT be addressed using existing methods.

New Model Framework?

- The posed research questions canNOT be addressed using existing methods.
- One may consider covariate-dependent Markov chains, where each transition probability is modeled as a function of subject-level covariates.

New Model Framework?

- The posed research questions canNOT be addressed using existing methods.
- One may consider covariate-dependent Markov chains, where each transition probability is modeled as a function of subject-level covariates.
- However, several challenges remain . . .

Challenges

Challenges in modeling MS DMT data with covariate-dependent transitions in a Markov chain:

- **Sparse transitions:** Many treatment transitions are sparse or even empirically unobserved.

Challenges

Challenges in modeling MS DMT data with covariate-dependent transitions in a Markov chain:

- **Sparse transitions:** Many treatment transitions are sparse or even empirically unobserved.
- **Multi-modality of likelihood:** The overall likelihood may be non-concave, leading to potential multi-modality.

Challenges

Challenges in modeling MS DMT data with covariate-dependent transitions in a Markov chain:

- **Sparse transitions:** Many treatment transitions are sparse or even empirically unobserved.
- **Multi-modality of likelihood:** The overall likelihood may be non-concave, leading to potential multi-modality.
- **Constraints and identifiability:** For example, transition probabilities in each row must lie on the unit simplex.

Challenges

In order to address the aforementioned challenges, we propose the **SMART-MC** model framework, complemented by **MSCOR** — a global optimization algorithm designed to handle multimodality and constraint-related challenges.

SMART-MC: Sparse Matrix Estimation with Covariate-Based Transitions in Markov Chain Modeling

Goals

1. Explore how patients' clinical and demographical variables influence the MS patients' treatment-trajectory over doctor visits.
2. In other words, we estimate each across-treatment transition probabilities as functions of patient-level covariates.

$$\text{Transition probability}_{u,v}^{(i)} = func(X^{(i)}, \beta_{u,v})$$

- $u, v \equiv$ treatment indexes.
- $i \equiv$ patient index.

Notations

- K subjects, N treatment options.
- $\boldsymbol{X}_k = (X_{k1}, \ldots, X_{kp})$: p-dimensional covariate for k-th subject.
- $\boldsymbol{Y_k} = \{(Y_{k,1}, \ldots, Y_{k,t_k})\}_{k=1}^{K}$: treatment sequence of k^{th} subject over $t = 1, \ldots, t_k;\ Y_{k,t} \in \{1, \ldots, N\}$.
- Coefficient vectors $\{\beta_{u,v}\}$ for $u = 0, \ldots, N;\ v = 1 \ldots, N$.

$$\boldsymbol{B} = \begin{bmatrix} \boldsymbol{B}_s \\ \boldsymbol{B}_T \end{bmatrix} = \begin{bmatrix} \boldsymbol{\beta}_{0,1} & \boldsymbol{\beta}_{0,2} & \cdots & \boldsymbol{\beta}_{0,N} \\ \boldsymbol{\beta}_{1,1} & \boldsymbol{\beta}_{1,2} & \cdots & \boldsymbol{\beta}_{1,N} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\beta}_{u,1} & \boldsymbol{\beta}_{u,2} & \cdots & \boldsymbol{\beta}_{N,N} \end{bmatrix}_{(N+1)\times N} , \ ||\beta_{u,v}||_2 = 1,$$

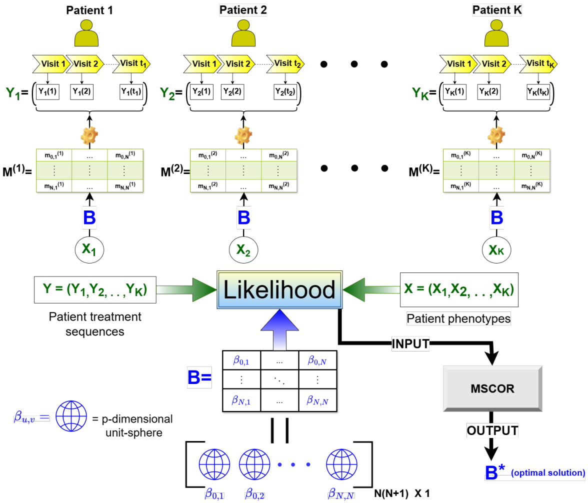
Notations

- Initial state vector (ISV)
$$M_s^{(k)} = \left[m_{0,v}^{(k)}\right]_{v=1}^{N},\ 0 \le m_{0,v}^{(k)} \le 1,\ \sum_{v=1}^{N} m_{0,v}^{(k)} = 1.$$
- Transition matrices
$$M_T^{(k)} = \left[m_{u,v}^{(k)}\right]_{u,v=1}^{N,N}.\ 0 \le m_{u,v}^{(k)} \le 1,\ \sum_{v=1}^{N} m_{u,v}^{(k)} = 1.$$
- Patient-specific transition matrices:

$$\boldsymbol{M}^{(k)} = \begin{bmatrix} \boldsymbol{M}_s^{(k)} \\ \boldsymbol{M}_T^{(k)} \end{bmatrix} = \begin{bmatrix} m_{0,1}^{(k)} & m_{0,2}^{(k)} & \dots & m_{0,N}^{(k)} \\ m_{1,1}^{(k)} & m_{1,2}^{(k)} & \dots & m_{1,N}^{(k)} \\ \vdots & \vdots & \ddots & \vdots \\ m_{N,1}^{(k)} & m_{N,2}^{(k)} & \dots & m_{N,N}^{(k)} \end{bmatrix}_{(N+1)\times N},$$

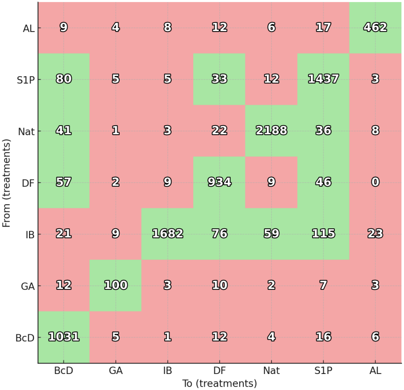
where
$$m_{u,v}^{(k)} = \frac{exp\left(\boldsymbol{X}_k \boldsymbol{\beta}_{u,v}\right)}{\sum_{n=1}^{N} exp\left(\boldsymbol{X}_k \boldsymbol{\beta}_{u,v}\right)},\ ||\beta_{u,v}||_2 = 1.$$

SMART-MC


Patient 1
Patient 2
Patient K
Visit 1
Visit 2
Visit t_1
Visit t_2
Visit t_K
$Y_1 = (Y_1(1), Y_1(2), Y_1(t_1))$
$Y_2 = (Y_2(1), Y_2(2), Y_2(t_2))$
$Y_K = (Y_K(1), Y_K(2), Y_K(t_K))$
$M^{(1)} =$
$M^{(2)} =$
$M^{(K)} =$
$m_{0,1}^{(1)}$ ... $m_{0,N}^{(1)}$
$m_{N,1}^{(1)}$ ... $m_{N,N}^{(1)}$
B
X_1
X_2
X_K
$Y = (Y_1, Y_2, \ldots, Y_K)$
Patient treatment sequences
Likelihood
$X = (X_1, X_2, \ldots, X_K)$
Patient phenotypes
INPUT
MSCOR
OUTPUT
B^* (optimal solution)
B=
$\beta_{0,1}$ ... $\beta_{0,N}$
$\beta_{N,1}$ ... $\beta_{N,N}$
$\beta_{u,v} =$ = p-dimensional unit-sphere
$\beta_{0,1}$ $\beta_{0,2}$ $\beta_{N,N}$
N(N+1) X 1

Sparse transition based adjustment

- Transition probabilities are estimated as functions of covariates ONLY if the number of empirical transitions is more than $Tol \geq p+1$ (user-defined).
  Following diagram marks the **non-rare** transitions in Red, **rare** transitions in Green, taking $p = 5$ and $Tol = 5(p+1) = 30$.



	BcD	GA	IB	DF	Nat	S1P	AL
AL	9	4	8	12	6	17	462
S1P	80	5	5	33	12	1437	3
Nat	41	1	3	22	2188	36	8
DF	57	2	9	934	9	46	0
IB	21	9	1682	76	59	115	23
GA	12	100	3	10	2	7	3
BcD	1031	5	1	12	4	16	6

Sparse transition based adjustment

- Otherwise, those positions are estimated as constants, being equal to empirical transitional probabilities.

$$m_{u,v}^{(k)} = \widehat{m}_{u,v} * I(\widehat{c}_{u,v} \leq Tol) + \left(1 - \sum_{n=1}^{N} \widehat{m}_{u,n} * I(\widehat{c}_{u,n} \leq Tol)\right)$$

$$* \frac{exp\big(\boldsymbol{X}_k \boldsymbol{\beta}_{u,v}\big) * I(\widehat{c}_{u,v} > Tol)}{\sum_{n=1}^{N} exp\big(\boldsymbol{X}_k \boldsymbol{\beta}_{u,n}\big) * I(\widehat{c}_{u,n} > Tol)},$$

$\widehat{c}_{u,v}$ = empirical initial or transition count, $Tol = p + 1$.

Sparse transition based adjustment

Advantages

- Avoids identifiability issue (which might arise while estimating probabilities corresponding to rare transitions).
- No cross-validation.
- Occurrence of rare transitions further reduces computational burden.
- Sparsity is induced automatically at locations with no transition counts.

SMART-MC: Theoretical properties

Regularity conditions:

(A1) The data $\{(Y_k, X_k)\}_{k=1}^{K}$ are independent and identically distributed (i.i.d.) draws from the underlying population.

(A2) The true transition probabilities satisfy

$$m_{u,v}^{(k)} = \begin{cases} m_{u,v}^{*}, & \text{if } v \in \mathcal{V}_u^c, \\ \left(1 - \sum_{v' \in \mathcal{V}_u^c} m_{u,v'}^{*}\right) \cdot \dfrac{\exp(X_k^\top \beta_{u,v}^{*})}{\sum_{v' \in \mathcal{V}_u} \exp(X_k^\top \beta_{u,v'}^{*})}, & \text{if } v \in \mathcal{V}_u, \end{cases}$$

where $\beta_{u,v}^{*} \in \mathbb{R}^{p+1}$ satisfy $\|\beta_{u,v}^{*}\|_2 = 1$, and $m_{u,v}^{*} \in (0, 1)$.

(A3) There exists a constant $C > 0$ such that $\|X_k\| \le C$ for all k.

(A4) For each $(u, v) \in \mathcal{V}_u$, the Fisher information matrix is full rank and covariates are not perfectly collinear (justified in details).

SMART-MC: Theoretical properties

Theorem (Consistency of SMART-MC)

Under assumptions (A1)–(A4), *as* $K \to \infty$, *the maximum likelihood estimator of SMART-MC satisfies:*

(a) *For all rare transitions* $v \in \mathcal{V}_u^c$, $\hat{m}_{u,v} \xrightarrow{P} m^*_{u,v}$.

(b) *For all non-rare transitions* $v \in \mathcal{V}_u$, $\hat{\beta}_{u,v} \xrightarrow{P} \beta^*_{u,v}$.

SMART-MC: Theoretical properties

Theorem (Asymptotic normality of SMART-MC)

Under assumptions (A1)–(A4)*, for each fixed origin state u and each non-rare destination state $v \in \mathcal{V}_u$, let $\hat{\beta}_{u,v}$ denote the maximum pseudo-likelihood estimator under the unit-norm constraint:*

$$\hat{\beta}_{u,v} := \arg \max_{\beta \in \mathbb{R}^{p+1}, \|\beta\|_2 = 1} \ell_u(\beta),$$

where $\ell_u(\beta)$ is the partial log-pseudo-likelihood defined over transitions from state u. Then,

$$\sqrt{n_u}\, P_{u,v}^\top \left(\hat{\beta}_{u,v} - \beta^*_{u,v} \right) \xrightarrow{d} \mathcal{N}(0, \Sigma_{u,v}),$$

where $P_{u,v} \in \mathbb{R}^{(p+1)\times p}$ is an orthonormal basis matrix for the tangent space $\mathcal{T}_{\beta^*_{u,v}} := \{h \in \mathbb{R}^{p+1} : \beta^{*\top}_{u,v} h = 0\}$; $\mathcal{I}_{u,v}$ is the Fisher information matrix evaluated at $\beta^*_{u,v}$ and $\Sigma_{u,v} := \left(P_{u,v}^\top \mathcal{I}_{u,v} P_{u,v}\right)^{-1}$.

SMART-MC: Inference via Wald-Type Test Statistics

For transition $(u \to v)$, pseudo-likelihood estimator $\hat{\beta}_{u,v}$ satisfies $\|\hat{\beta}_{u,v}\|_2 = 1$.

$$\sqrt{n_u}\, P_{u,v}^{\top}(\hat{\beta}_{u,v} - \beta_{u,v}^{*}) \xrightarrow{d} \mathcal{N}\big(0,\, (P_{u,v}^{\top}\mathcal{I}_{u,v}P_{u,v})^{-1}\big)$$

Projected coefficients:

$$\theta_{u,v} = P_{u,v}^{\top}\hat{\beta}_{u,v}, \quad \widehat{\Sigma}_{u,v} = (P_{u,v}^{\top}\widehat{\mathcal{I}}_{u,v}P_{u,v})^{-1}$$

Wald statistic:

$$z_j = \frac{(\theta_{u,v})_j}{\sqrt{[\widehat{\Sigma}_{u,v}]_{jj}/n_u}}, \quad p_j = 2\Phi(-|z_j|)$$

CI: $(\theta_{u,v})_j \pm z_{\alpha/2}\sqrt{[\widehat{\Sigma}_{u,v}]_{jj}/n_u}$

Challenges with Maximizing likelihood

Challenges with Maximizing likelihood

- Sparsity-adjusted likelihood is NOT necessarily concave.
- Parameter-space is given by a collection of unit-spheres.
- Ideally, a global optimization algorithm should be used to find MLE.

MSCOR: Multiple Spherically Constrained Optimization Routine

‘Exploitation vs exploration’ dilemma

- The explore-exploit trade-off addresses the balance between acquiring new information (*exploration*) and utilizing it for performance improvement (*exploitation*).
- Gradient-based methods $\equiv$ Full *exploitation*.
- Grid search $\equiv$ Full *exploration*.

‘Exploitation vs exploration’ dilemma

- Grid search (i.e., full *exploration*) can be very expensive.

‘Exploitation vs exploration’ dilemma

- Grid search (i.e., full *exploration*) can be very expensive.
- Gradient-based methods (i.e., full *exploitation*) lack ability to jump out of local solutions, might not exploit the full potential of parallel threading.

Global optimization

Key aspects of typical Global optimization algorithms:

- Global optimization methods balance these extremes, conceptually modeled as

  $$\gamma \cdot \textit{exploitation} + (1-\gamma) \cdot \textit{exploration}, \quad \gamma \in (0,1),$$

  bridging gradient-based methods ($\gamma = 1$) and grid search ($\gamma = 0$).
- Global optimization algorithms typically adopt a strategy to escape local solutions.
- Search-space exploration strategy can be parallelized (depends on the algorithm).

Recursive Modified Pattern Search

Recursive Modified Pattern Search is a well-established Blackbox optimization technique earlier developed on following parameter spaces:

- Multiple Simplexes (MSiCOR; Das et al. (2023a), *JCGS*)
- Hyper-rectangular (RMPSH; Das (2023), *Indian J. Stat*)
- Spherical (SCOR; Das et al. (2022), *BMC Bioinformatics*)
- Unit Simplex (RMPSS; Das (2021), *Indian J. Stat*)

Fermi's principal

Pattern-search strategy is based on *Fermi's Principle* (Fermi and Metropolis (1952))

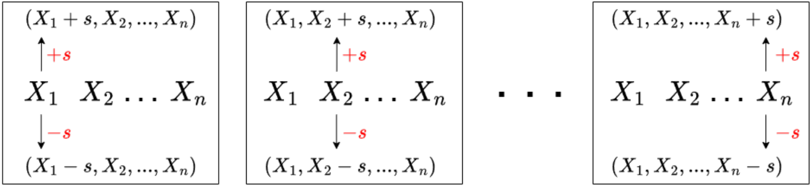


Figure: Fermi's principle : Possible $2n$ movements starting from initial point $(x_1, \dots, x_n)$ inside an iteration with fixed step-size s, while optimizing any n-dimensional objective function over unconstrained parameter space.

Fermi's principal on unit sphere

Figure: Candidate solutions obtained via Fermi's principle roaming around *j*-th coordinate on a unit-sphere.

$$\begin{tikzcd}
(X_1 + t_j^+, \dots, X_{j-1} + t_j^+, X_j + s, X_{j+1} + t_j^+, \dots, X_n + t_j^+) \\
(X_1, X_2, \dots, X_j, \dots, X_n) \arrow[u, "{(j,+)}"] \arrow[d, "{(j,-)}"'] \\
(X_1 + t_j^-, \dots, X_{j-1} + t_j^-, X_j - s, X_{j+1} + t_j^-, \dots, X_n + t_j^-)
\end{tikzcd}$$

$$t_j = \frac{-2\sum_{q=1,q\neq j}^{n} X_q + \sqrt{D_i(s)}}{2(n-1)}; \text{ where } D_j(s) = \left(2\sum_{q=1,q\neq j}^{n} X_q\right)^2 - 4(n-1)(2sX_j + s^2).$$

MSCOR with B spheres, each of length n (i.e., dim $n-1$), will yield $2Bn$

candidate solutions at each iterations.

MSCOR: Overall strategy

- MSCOR consists of a sequence of *runs*. Within each *run*, a set of iterations are performed.

MSCOR: Overall strategy

- MSCOR consists of a sequence of *runs*. Within each *run*, a set of iterations are performed.

- At the beginning of a *run*, the step-size is set to a large step-size, e.g., $s = 1$. As iterations within the *run* proceed, depending on the improvement of the objective function value, the step-size decays.

MSCOR: Overall strategy

- MSCOR consists of a sequence of *runs*. Within each *run*, a set of iterations are performed.
- At the beginning of a *run*, the step-size is set to a large step-size, e.g., $s = 1$. As iterations within the *run* proceed, depending on the improvement of the objective function value, the step-size decays.
- At the end of each iteration, the best solution among the $2Bn + 1$ candidate solutions (including the solution from the previous iteration) is selected to move forward with.

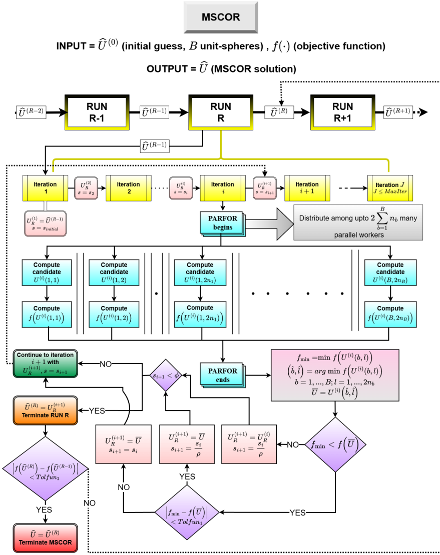
MSCOR: Overall strategy

- MSCOR consists of a sequence of *runs*. Within each *run*, a set of iterations are performed.
- At the beginning of a *run*, the step-size is set to a large step-size, e.g., $s = 1$. As iterations within the *run* proceed, depending on the improvement of the objective function value, the step-size decays.
- At the end of each iteration, the best solution among the $2Bn + 1$ candidate solutions (including the solution from the previous iteration) is selected to move forward with.
- A *run* ends when step-size becomes significantly small (say $< 10^{-6}$).

MSCOR: Overall strategy

- The next *run* starts from the solution returned by the previous *run*, but with large step-size (which again decays across iterations in that *run*). Sudden increment in step-size encourages distant exploration, enabling **jumping-out-of-location-solution.**
- MSCOR terminates when two consecutive *runs* yield the same solution (in words, that strategy translates into, *"even jumping out didn't yield better solution"*).

Parallel MSCOR



MSCOR

INPUT = $\widehat{U}^{(0)}$ (initial guess, B unit-spheres) , $f(\cdot)$ (objective function)

OUTPUT = $\widehat{U}$ (MSCOR solution)

$\widehat{U}^{(R-2)}$ → RUN R-1 → $\widehat{U}^{(R-1)}$ → RUN R → $\widehat{U}^{(R)}$ → RUN R+1 → $\widehat{U}^{(R+1)}$

$\widehat{U}^{(R-1)}$

Iteration 1 — $U_R^{(2)}$, $s = s_2$ → Iteration 2 …… $U_R^{(i)}$, $s = s_i$ → Iteration i — $U_R^{(i+1)}$, $s = s_{i+1}$ → Iteration $i+1$ - - → Iteration J, $J \leq MaxIter$

$U_R^{(1)} = \widehat{U}^{(R-1)}$, $s = s_{initial}$

PARFOR begins

Distribute among upto $2\sum_{b=1}^{B} n_b$ many parallel workers

Compute candidate $U^{(i)}(1,1)$ → Compute $f\left(U^{(i)}(1,1)\right)$

Compute candidate $U^{(i)}(1,2)$ → Compute $f\left(U^{(i)}(1,2)\right)$

Compute candidate $U^{(i)}(1,2n_1)$ → Compute $f\left(U^{(i)}(1,2n_1)\right)$

. . . . . .

Compute candidate $U^{(i)}(B,2n_B)$ → Compute $f\left(U^{(i)}(B,2n_B)\right)$

PARFOR ends

$$f_{\min} = \min f\left(U^{(i)}(b,l)\right)$$
$$\left(\hat{b},\hat{l}\right) = arg\min f\left(U^{(i)}(b,l)\right)$$
$$b = 1,\dots,B;\ l = 1,\dots,2n_b$$
$$\overline{U} = U^{(i)}\left(\hat{b},\hat{l}\right)$$

$f_{\min} < f\left(\overline{U}\right)$

NO: $U_R^{(i+1)} = U_R^{(i)}$, $s_{i+1} = \dfrac{s_i}{\rho}$

YES: $\left|f_{\min} - f\left(\overline{U}\right)\right| < Tolfun_1$

YES: $U_R^{(i+1)} = \overline{U}$, $s_{i+1} = \dfrac{s_i}{\rho}$

NO: $U_R^{(i+1)} = \overline{U}$, $s_{i+1} = s_i$

$s_{i+1} < \phi$

NO: Continue to iteration $i+1$ with $U_R^{(i+1)}, s = s_{i+1}$

YES: $\widehat{U}^{(R)} = U_R^{(i+1)}$ Terminate RUN R

$\left|f\left(\widehat{U}^{(R)}\right) - f\left(\widehat{U}^{(R-1)}\right)\right| < Tolfun_2$

YES: $\widehat{U} = \widehat{U}^{(R)}$ Terminate MSCOR

NO

MSCOR

MSCOR: Multiple Spherically Constrained Optimization Routine

- Black-box optimization technique over multiple spherically constrained parameter space.
- Functions can be multi-modal/non-convex/non-closed-form/non-differentiable/discontinuous.
- Employs a strategy of jumping out of local optimum points.
- CPU and GPU parallelizable.

MSCOR convergence

Definition

The ‘*shadow*’ of a point $\boldsymbol{W}$ (denoted by $\boldsymbol{W}^C$) belonging to the closure of $\boldsymbol{S}$ (i.e., $\bar{\boldsymbol{S}}$) is the point of intersection of the straight line connecting the origin to $\boldsymbol{W}$ with $\boldsymbol{S}$, where $\boldsymbol{S} = O^{n_1-1} \times \cdots \times O^{n_B-1}$, and $O^{w-1} = \{(x_1, \cdots, x_w) \in \mathrm{R}^w \ : \ \sum_{i=1}^{w} x_i^2 = 1, i = 1, \cdots, w\}$.

MSCOR convergence

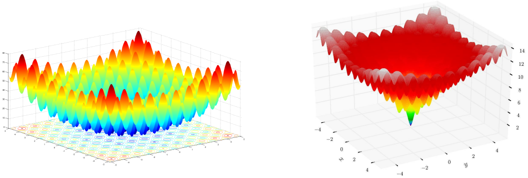
Theorem

Suppose $f : \boldsymbol{S} \mapsto \mathbb{R}$ *is convex, continuous and differentiable with extended definition on* $\bar{\boldsymbol{S}}$*, such that,* $f(\boldsymbol{W}) = f(\boldsymbol{W}^C)$ *when* $\boldsymbol{W} \in interior(\boldsymbol{S})$*. Consider a sequence* $\delta_{j,k} = \frac{s_j}{\rho^k}$ *for* $k \in \mathrm{N}$ *and* $s_j > 0, \rho > 1$*. Suppose* $\mathbf{U} \in \boldsymbol{S}$ *given by*

$$\mathbf{U} = (\mathbf{u}_1, \ldots, \mathbf{u}_B) \text{ where } \mathbf{u}_b = (u_{j,1}, \ldots, u_{j,n_b}) \in O^{n_j - 1},\ j = 1, \ldots, B.$$

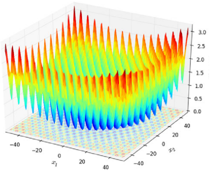
Define, $\mathbf{u}_{j,k}^{(i+)} = (u_{j,1} + t_i(\delta_{j,k}), \ldots, u_{j,i-1} + t_i(\delta_{j,k}), u_{j,i} + \delta_{j,k}, u_{j,i+1} + t_i(\delta_{j,k}), \ldots, u_{j,n_j} + t_i(\delta_{j,k}))$, $\mathbf{u}_{j,k}^{(i-)} = (u_{j,1} + t_i(-\delta_{j,k}), \ldots, u_{j,i-1} + t_i(-\delta_{j,k}), u_{j,i} - \delta_{j,k}, u_{j,i+1} + t_i(-\delta_{j,k}), \ldots, u_{j,n_j} + t_i(-\delta_{j,k}))$ *for* $j = 1, \ldots, B$, $i = 1, \cdots, n_j$*, where* $t_i(s)$ *denotes the adjustment step size corresponding to step size* s*. Under some regularity conditions, if for all sufficiently large* $k \in \mathrm{N}$*, for* $j = 1, \ldots, B$, $i = 1, \ldots, n_j - 1$*, then the global minimum of* f *over* $\mathbf{S}$ *occurs at* $\mathbf{U}$.

Example benchmark functions: 2-D Rastrigin, Ackley's and Griewank function



2D Rastrigin

2D Ackley



2D Griewank

Figure: 2-D Rastrigin, Ackley's and Griewank function.

Non-convex function based benchamrk study

Functions	Algorithms	$B = 5, n_b = 5$			$B = 10, n_b = 20$			$B = 100, n_b = 5$		
		min. value	se of solution	mean time (se)	min. value	se of solution	mean time (se)	min. value	se of solution	mean time (se)
Ackley's (modified)	MSCOR	**2.22e - 14**	0.029	1.64 (0.008)	**3.61e - 13**	0.000	312.58 (0.385)	**1.65e - 09**	0.288	3600.04^*(0.006)
	GA	1.51e + 01	0.169	16.34 (0.769)	2.59e + 01	0.080	78.81 (0.315)	3.88e + 02	2.214	357.20 (3.682)
	SA	4.70e + 00	0.081	1.84 (0.092)	2.16e + 01	0.039	53.48 (2.490)	2.70e + 02	1.175	371.77 (23.565)
	IP	**7.51e - 12**	0.347	0.06 (0.003)	**3.67e - 03**	0.023	0.09 (0.002)	2.17e + 02	6.179	0.42 (0.026)
	SQP	9.50e - 04	0.414	0.03 (0.001)	1.28e - 02	0.000	0.41 (0.001)	**8.33e + 01**	9.219	5.15 (0.028)
	AS	2.35e + 00	0.328	0.03 (0.001)	1.53e + 00	0.401	0.47 (0.003)	1.56e + 02	4.564	5.60 (0.010)
Griewank (modified)	MSCOR	<**1e - 16**	0.000	1.54 (0.007)	**1.78e - 15**	0.000	204.51 (0.444)	**1.46e - 09**	0.000	3600.07^*(0.010)
	GA	8.04e - 01	0.040	19.59 (0.962)	1.12e + 00	0.021	88.70 (0.287)	3.60e + 01	0.400	461.57 (4.188)
	SA	1.06e - 01	0.008	2.03 (0.101)	7.99e - 01	0.004	54.12 (2.392)	2.72e + 01	0.166	372.25 (11.450)
	IP	2.47e - 13	0.000	0.02 (0.002)	6.53e - 04	0.000	0.10 (0.002)	2.03e + 00	0.175	0.50 (0.025)
	SQP	**1.98e - 13**	0.000	0.01 (0.000)	**5.96e - 12**	0.000	0.24 (0.001)	**3.80e - 12**	0.000	1.69 (0.015)
	AS	3.50e - 08	0.022	0.03 (0.002)	2.77e - 07	0.005	0.43 (0.015)	4.54e - 07	0.464	5.79 (0.722)
Neg. sum of squares (modified)	MSCOR	<**1e - 16**	0.000	0.45 (0.005)	<**1e - 16**	0.000	43.81 (0.413)	**1.51e - 14**	0.000	1602.09 (15.515)
	GA	5.17e + 00	0.198	16.47 (0.805)	8.27e + 01	0.648	74.74 (0.258)	1.89e + 02	2.398	325.61 (2.558)
	SA	2.19e + 00	0.044	1.85 (0.087)	7.10e + 01	0.126	50.59 (2.549)	1.65e + 02	0.435	358.06 (16.27)
	IP	**7.99e - 15**	0.000	0.02 (0.000)	1.26e + 00	0.100	0.09 (0.002)	3.83e + 00	1.520	0.40 (0.023)
	SQP	1.07e - 14	0.000	0.02 (0.000)	**4.26e - 07**	0.000	0.41 (0.002)	**9.09e - 12**	0.000	3.78 (0.102)
	AS	1.92e - 09	0.093	0.02 (0.001)	1.60e + 01	0.714	0.45 (0.003)	2.42e + 01	3.595	5.53 (0.093)
Rastrigin (modified)	MSCOR	<**1e - 16**	0.762	2.08 (0.417)	**8.53e - 13**	0.000	135.99 (0.255)	**1.02e + 02**	5.544	3600.04^*(0.011)
	GA	9.90e + 01	5.792	18.21 (0.835)	1.59e + 03	9.215	79.37 (0.262)	4.98e + 03	73.999	412.85 (51.696)
	SA	8.64e + 00	0.302	1.76 (0.082)	3.47e + 01	2.006	93.74 (3.322)	4.72e + 02	10.532	935.30 (60.66)
	IP	6.72e + 00	0.725	0.04 (0.001)	**1.68e - 04**	5.922	0.10 (0.001)	5.14e + 02	111.633	0.41 (0.010)
	SQP	8.18e + 00	0.637	0.03 (0.000)	7.04e + 00	3.435	0.42 (0.002)	**4.71e + 02**	10.107	5.32 (0.075)
	AS	**1.20e + 00**	0.969	0.03 (0.000)	2.19e + 02	21.095	0.46 (0.001)	8.49e + 02	104.721	5.78 (0.058)

Table: Comparison study: MSCOR, genetic algorithm (GA), simulated annealing (SA), interior-point (IP), sequential quadratic programming (SQP), and active-set (AS).

Simulation Study

SMART-MC Simulation study

- Number of treatments $N = 10$, number of patients $K = 1000$ patients, number of covariates $p = 5$ (excluding intercept).
- Treatment sequence length for each patient $t_k = 20$.
- **Sparse True Transition matrix:** Subject to (i) 67% true transitions probabilities are taken to be 0 (similar to real dataset). (ii) Each row must have at least 2 non-zero probabilities including the (always non-zero) diagonal elements.
- Compared with Naive model (that does NOT incorporate special treatment for sparsity)

SMART-MC Simulation study

Transitions	Transition counts	Scenario	β_0	β_1	β_2	β_3	β_4	β_5
4 (initial state)	203 (20.30%)	True	0.23	-0.87	0.00	-0.35	-0.04	-0.27
		SMART-MC	0.24 (0.019)	-0.80 (0.014)	0.04 (0.020)	-0.45 (0.021)	-0.25 (0.014)	-0.20 (0.025)
		Naive	0.88 (0.019)	-0.29 (0.039)	-0.01 (0.037)	-0.07 (0.037)	-0.35 (0.035)	-0.11 (0.035)
$6 \mapsto 6$	2342 (12.33%)	True	0.37	-0.82	-0.16	-0.18	-0.37	0.05
		SMART-MC	0.40 (0.032)	-0.83 (0.017)	-0.13 (0.025)	-0.12 (0.042)	-0.34 (0.017)	0.08 (0.032)
		Naive	0.82 (0.033)	-0.34 (0.034)	0.23 (0.038)	-0.03 (0.035)	-0.34 (0.032)	-0.20 (0.033)
$4 \mapsto 4$	1796 (9.45%)	True	0.30	-0.57	-0.45	0.32	0.27	0.44
		SMART-MC	0.32 (0.011)	-0.57 (0.012)	-0.45 (0.018)	0.31 (0.011)	0.27 (0.023)	0.44 (0.011)
		Naive	0.83 (0.030)	-0.41 (0.034)	-0.22 (0.036)	0.27 (0.033)	0.05 (0.037)	0.15 (0.030)
$7 \mapsto 7$	1307 (6.88%)	True	0.16	0.29	-0.60	0.65	0.32	-0.10
		SMART-MC	0.11 (0.023)	0.22 (0.029)	-0.64 (0.016)	0.64 (0.013)	0.27 (0.019)	-0.21 (0.023)
		Naive	0.73 (0.035)	0.03 (0.035)	-0.32 (0.035)	0.59 (0.035)	0.02 (0.037)	-0.13 (0.030)
$8 \mapsto 8$	1261 (6.64%)	True	0.02	0.58	0.42	-0.06	-0.69	-0.01
		SMART-MC	0.15 (0.029)	0.58 (0.017)	0.40 (0.013)	-0.08 (0.022)	-0.69 (0.023)	-0.07 (0.029)
		Naive	0.74 (0.034)	0.33 (0.038)	0.48 (0.033)	-0.29 (0.037)	-0.16 (0.039)	-0.02 (0.034)
$6 \mapsto 7$	1070 (5.63%)	True	0.10	-0.01	-0.71	0.05	0.51	0.47
		SMART-MC	0.07 (0.032)	0.03 (0.017)	-0.67 (0.025)	0.09 (0.041)	0.54 (0.0017)	0.50 (0.032)
		Naive	0.77 (0.038)	0.37 (0.036)	-0.27 (0.033)	0.11 (0.038)	0.39 (0.040)	0.18 (0.038)
$5 \mapsto 4$	900 (4.74%)	True	0.50	0.05	-0.24	-0.71	-0.23	-0.36
		SMART-MC	0.53 (0.032)	0.16 (0.017)	-0.16 (0.028)	-0.73 (0.023)	-0.24 (0.035)	-0.27 (0.032)
		Naive	0.95 (0.034)	-0.08 (0.034)	-0.12 (0.033)	-0.15 (0.035)	0.14 (0.032)	-0.18 (0.034)
$8 \mapsto 2$	872 (4.59%)	True	0.33	-0.19	-0.81	0.42	-0.14	0.08
		SMART-MC	0.34 (0.030)	-0.18 (0.019)	-0.80 (0.014)	0.43 (0.022)	-0.15 (0.022)	0.09 (0.030)
		Naive	0.84 (0.033)	-0.21 (0.035)	-0.40 (0.036)	0.08 (0.036)	0.25 (0.040)	0.12 (0.033)
$2 \mapsto 5$	754 (3.97%)	True	0.20	-0.59	-0.25	0.05	-0.10	-0.73
		SMART-MC	0.20 (0.025)	-0.54 (0.019)	-0.26 (0.016)	0.04 (0.024)	-0.10 (0.030)	-0.77 (0.025)
		Naive	0.83 (0.033)	-0.29 (0.036)	-0.05 (0.033)	-0.07 (0.038)	0.32 (0.036)	-0.33 (0.033)
$4 \mapsto 2$	684 (3.60%)	True	-0.43	0.45	-0.09	-0.61	0.06	-0.49
		SMART-MC	-0.40 (0.010)	0.50 (0.010)	-0.10 (0.018)	-0.58 (0.011)	0.14 (0.022)	-0.48 (0.010)
		Naive	0.53 (0.039)	0.48 (0.038)	0.15 (0.039)	-0.43 (0.036)	-0.04 (0.036)	-0.53 (0.039)
$3 \mapsto 5$	678 (3.57%)	True	0.26	0.44	0.57	0.39	0.37	-0.36
		SMART-MC	0.24 (0.031)	0.42 (0.015)	0.54 (0.026)	0.37 (0.047)	0.39 (0.026)	-0.43 (0.031)
		Naive	0.90 (0.034)	0.29 (0.034)	0.14 (0.032)	0.06 (0.030)	0.19 (0.034)	-0.21 (0.034)

Table: The true and estimated coefficients of the covariates corresponding to the most frequent initial treatment and the top 10 most frequent treatment transitions (empirically) are reported.

SMART-MC Simulation study

Trt. seq.	Number of patients		
length	1000	2000	3000
20	0.0296 (0.0017)	0.0209 (0.0011)	0.0171 (0.0009)
40	0.0209 (0.0010)	0.0150 (0.0010)	0.0120 (0.0007)
60	0.0169 (0.0011)	0.0121 (0.0006)	0.0096 (0.0007)

Table: The mean absolute deviation (MAD) between the true and estimated coefficients for the top 10 most frequent treatment transitions.

SMART-MC Simulation study

- Parallel MSCOR (using MATLAB with 12 CPU cores) ran up to ∼**7x** faster compared to MSCOR .

Num. covariates	Num. treatments	Number of parameters	$K = 1000, t_k = 10$			$K = 1000, t_k = 20$			$K = 2000, t_k = 20$		
			MSCOR time (sec)	par-MSCOR time (sec)	Speed improvement	MSCOR time (sec)	par-MSCOR time (sec)	Speed improvement	MSCOR time (sec)	par-MSCOR time (sec)	Speed improvement
$p = 3$	$N = 6$	168	38	10	3.8x	43	11	3.9x	95	31	3.1x
	$N = 9$	360	204	32	6.4x	252	40	6.3x	587	96	6.1x
	$N = 12$	624	1198	178	6.7x	1502	211	7.1x	3253	501	6.5x
$p = 5$	$N = 6$	252	53	10	5.3x	57	11	5.2x	136	22	6.2x
	$N = 9$	540	328	52	6.3x	443	71	6.2x	1014	163	6.2x
	$N = 12$	936	2158	344	6.3x	2825	455	6.2x	7077	1082	6.5x
$p = 8$	$N = 6$	378	120	24	5.0x	159	29	5.5x	315	54	5.8x
	$N = 9$	810	744	119	6.3x	923	143	6.5x	2057	337	6.1x
	$N = 12$	1404	4127	634	6.5x	4881	765	6.4x	12697	1931	6.6x

Table: Computation times: MSCOR vs parallel MSCOR.

SMART-MC analysis of MS DMTs

MS DMTs data (Recap)

Dataset:

- Multiple Sclerosis (MS) disease-modifying therapy (DMT) sequence data from the `Massachusetts General and Brigham hospital system (Boston, US) EHR cohort`.
- 822 patients, observed over $\sim$ 16 years.
- **7 classes of DMTs** are considered (listed in the next slide).
- Available patient-level covariates: **Age at diagnosis**, **disease duration**, **gender (M/F)**, **race (Black/White/Others)**.

MS DMT treatment sequence data (Recap)

7 classes of DMTs are given as follows :

1. B-cell depletion (**BcD**; includes rituximab,ocrelizumab)
2. Glatiramer acetate (**GA**)
3. Interferon-beta (**IB**)
4. Dimethyl fumarate (**DF**)
5. Natalizumab (**Nat**)
6. S1P modulators (includes fingolimod, teriflunomide)
7. Aggressive/Legacy therapies (**AL**; includes cyclophosphamide, mitoxantrone, alemtuzumab)

MS DMTs analysis

Transitions	Transition counts	Intercept	Age at diagnosis	Disease duration	Sex (Female)	Race: White	Race: Black
Nat ↦ Nat	2188 (24.84%)	0.70 (0.088)	0.27 (0.075)	0.13 (0.098)	0.21 (0.106)	0.57 (0.108)	0.24 (0.102)
IB ↦ IB	1682 (19.45%)	0.76 (0.072)	0.12 (0.091)	0.35 (0.107)	-0.20 (0.119)	0.33 (0.125)	0.36 (0.137)
S1P ↦ S1P	1437 (16.62%)	0.51 (0.099)	0.19 (0.131)	0.18 (0.142)	-0.27 (0.160)	0.77 (0.170)	-0.11 (0.231)
BcD ↦ BcD	1031 (11.92%)	NA	NA	NA	NA	NA	NA
DF ↦ DF	934 (10.80%)	0.61 (0.129)	0.49 (0.116)	-0.18 (0.144)	0.55 (0.172)	-0.12 (0.198)	0.17 (0.232)
AL ↦ AL	462 (5.34%)	NA	NA	NA	NA	NA	NA
IB ↦ S1P	115 (1.33%)	-0.71 (0.193)	-0.31 (0.119)	0.33 (0.131)	0.05 (0.181)	-0.04 (0.212)	-0.54 (0.218)
GA ↦ GA	100 (1.16%)	NA	NA	NA	NA	NA	NA
S1P ↦ BcD	80 (0.93%)	-0.69 (0.152)	-0.14 (0.151)	-0.04 (0.168)	-0.48 (0.219)	0.24 (0.239)	-0.47 (0.356)
IB ↦ DF	76 (0.88%)	-0.53 (0.072)	-0.14 (0.091)	0.28 (0.103)	-0.69 (0.125)	-0.36 (0.113)	0.08 (0.127)
IB ↦ Nat	59 (0.68%)	-0.73 (0.317)	-0.48 (0.124)	-0.20 (0.170)	0.08 (0.222)	0.32 (0.249)	-0.28 (0.325)
DF ↦ BcD	57 (0.66%)	-0.79 (0.298)	0.16 (0.145)	-0.15 (0.195)	0.18 (0.275)	-0.55 (0.233)	0.05 (0.380)
DF ↦ S1P	46 (0.53%)	-0.82 (0.158)	0.03 (0.124)	-0.37 (0.154)	0.06 (0.189)	0.18 (0.198)	-0.39 (0.128)
Nat ↦ BcD	41 (0.47%)	-0.82 (0.064)	-0.08 (0.088)	-0.20 (0.119)	-0.12 (0.126)	-0.48 (0.089)	-0.20 (0.119)
Nat ↦ S1P	36 (0.41%)	-0.65 (0.088)	-0.12 (0.112)	-0.23 (0.149)	-0.32 (0.152)	-0.51 (0.112)	-0.38 (0.131)
S1P ↦ DF	33 (0.38%)	-0.80 (0.090)	-0.21 (0.123)	-0.15 (0.141)	-0.27 (0.152)	-0.36 (0.142)	-0.30 (0.161)

Table: SMART-MC estimated coefficient values corresponding to the most frequent treatment transitions (with at least 30 transitions) are reported, along with the corresponding transition counts (transition proportions). red ≡ ‘negative’.

MS DMTs analysis

Transitions	Age at diagnosis	Disease duration	Sex (Female)	Race: White	Race: Black
Nat ↦ Nat	<0.001*	0.185	0.048*	<0.001*	0.019*
IB ↦ IB	0.187	0.001*	0.093	0.008*	0.009*
S1P ↦ S1P	0.147	0.205	0.092	<0.001*	0.634
BcD ↦ BcD	NA	NA	NA	NA	NA
DF ↦ DF	<0.001*	0.211	0.001*	0.544	0.464
AL ↦ AL	NA	NA	NA	NA	NA
IB ↦ S1P	0.009*	0.012*	0.782	0.850	0.013*
GA ↦ GA	NA	NA	NA	NA	NA
S1P ↦ BcD	0.354	0.812	0.028*	0.315	0.187
IB ↦ DF	0.124	0.007*	<0.001*	<0.001*	0.529
IB ↦ Nat	<0.001*	0.239	0.719	0.199	0.389
DF ↦ BcD	0.270	0.442	0.513	0.018*	0.895
DF ↦ S1P	0.809	0.016*	0.751	0.363	0.002*
Nat ↦ BcD	0.363	0.093	0.341	<0.001*	0.093
Nat ↦ S1P	0.284	0.123	0.035*	<0.001*	0.004*
S1P ↦ DF	0.088	0.287	0.076	0.011*	0.062

Table: Bootstrap-based two-sided p-values for estimated SMART-MC coefficients corresponding to treatment transitions with at least 30 observed cases. red ≡ ‘negative association’.

MS DMTs analysis

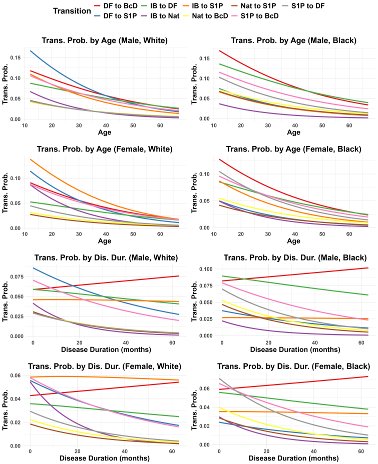
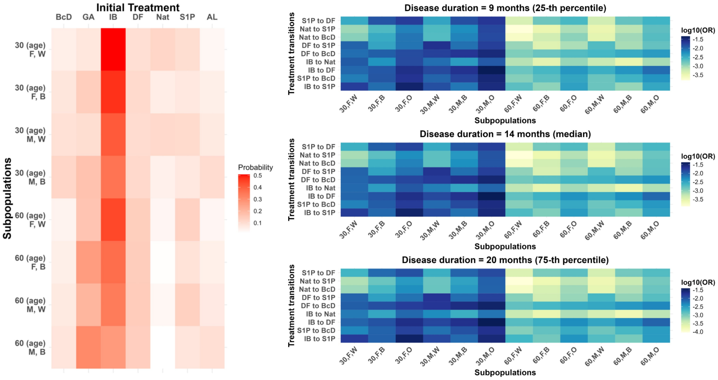


Figure: Estimated transition probabilities for non-rare across-DMT transitions across age and disease duration, stratified by key patient subgroups, as derived from the SMART-MC model.

MS DMTs analysis



(a) (b)

Figure: (a) Estimated initial treatment probabilities across subpopulations defined by age (30 or 60 years), sex (M = Male, F = Female), and race (W = White, B = Black). (b) Odds ratios (OR) of transitioning to a different treatment versus continuing the same treatment for the top 7 most frequent MS DMT transitions (to a different one).

Effect of Clinical Factors: Disease Duration

Effect of Clinical Factors: Disease Duration

- IB $\mapsto$ IB ($p = 0.001$; positive effect),
- IB $\mapsto$ DF ($p = 0.007$; positive effect),
- IB $\mapsto$ S1P ($p = 0.012$; positive effect),
- DF $\mapsto$ S1P ($p = 0.016$; negative effect).

Effect of Demographic Factors: Age at diagnosis

Effect of Demographic Factors: Age at diagnosis

- Nat $\mapsto$ Nat ($p < 0.001$; positive effect),
- DF $\mapsto$ DF ($p < 0.001$; positive effect),
- IB $\mapsto$ S1P ($p = 0.009$; negative effect),
- IB $\mapsto$ Nat ($p < 0.001$; negative effect).

Effect of Demographic Factors: Sex

Effect of Demographic Factors: Sex (Female)

- Nat $\mapsto$ Nat ($p = 0.048$; positive effect),
- DF $\mapsto$ DF ($p = 0.001$; positive effect),
- IB $\mapsto$ DF ($p < 0.001$; negative effect),
- Nat $\mapsto$ S1P ($p = 0.035$; negative effect),
- S1P $\mapsto$ BcD ($p = 0.028$; negative effect).

Effect of Demographic Factors: Race

Effect of Demographic Factors: Race (Black)

- Nat $\mapsto$ Nat ($p = 0.019$; positive effect),
- IB $\mapsto$ IB ($p = 0.009$; positive effect),
- IB $\mapsto$ S1P ($p = 0.013$; negative effect),
- DF $\mapsto$ S1P ($p = 0.002$; negative effect),
- Nat $\mapsto$ S1P ($p = 0.004$; negative effect).

Effect of Demographic Factors: Race

Effect of Demographic Factors: Race (White)

- Nat $\mapsto$ Nat ($p < 0.001$; positive effect),
- IB $\mapsto$ IB ($p = 0.008$; positive effect),
- S1P $\mapsto$ S1P ($p < 0.001$; positive effect),
- IB $\mapsto$ DF ($p < 0.001$; negative effect),
- DF $\mapsto$ BcD ($p = 0.018$; negative effect),
- Nat $\mapsto$ BcD ($p < 0.001$; negative effect),
- Nat $\mapsto$ S1P ($p < 0.001$; negative effect),
- S1P $\mapsto$ DF ($p = 0.011$; negative effect)

Key findings

Key findings:

1. Older patients were more likely to persist on existing treatments such as DF and Nat, but less likely to escalate from IB to Nat or S1P.
2. Female patients were more likely to stay on DF and Nat, but less likely to transition from IB to DF, Nat to S1P, or S1P to BcD.
3. Longer disease history and male sex increase the likelihood of injectable-to-oral switch, while White patients were less likely to undergo it.
4. Younger non-Black patients with longer disease duration were more likely to escalate to S1P modulators.

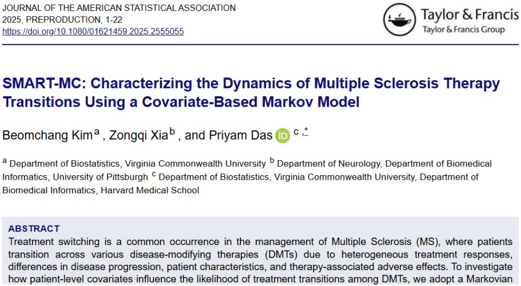
Key findings

Key findings (contd.):

5. Patients with longer disease course, and Black patients were less likely to escalate from fumarate (DF) to S1P.
6. White patients were less likely to escalate from dimethyl fumarate to B-cell therapies.
7. Females and racial minorities (non-Black, non-White) were less likely to de-escalate from natalizumab to S1P modulators.
8. Patients with longer disease course, and Black patients were less likely to escalate from fumarate (DF) to S1P.

Availibility

- Published in *Journal of the American Statistical Association*.
- Codes available on github at github.com/priyamdas2/SMART-MC-MSCOR.



JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION
2025, PREPRODUCTION, 1-22
https://doi.org/10.1080/01621459.2025.2555055
Taylor & Francis
Taylor & Francis Group

SMART-MC: Characterizing the Dynamics of Multiple Sclerosis Therapy Transitions Using a Covariate-Based Markov Model

Beomchang Kim[a], Zongqi Xia[b], and Priyam Das[c,*]

[a] Department of Biostatistics, Virginia Commonwealth University [b] Department of Neurology, Department of Biomedical Informatics, University of Pittsburgh [c] Department of Biostatistics, Virginia Commonwealth University, Department of Biomedical Informatics, Harvard Medical School

ABSTRACT
Treatment switching is a common occurrence in the management of Multiple Sclerosis (MS), where patients transition across various disease-modifying therapies (DMTs) due to heterogeneous treatment responses, differences in disease progression, patient characteristics, and therapy-associated adverse effects. To investigate how patient-level covariates influence the likelihood of treatment transitions among DMTs, we adopt a Markovian

Acknowledgements

Beomchang Kim
(PhD student,
Department of
Biostatistics, VCU)

Dr. Xongqi Zia
(Neurologist and Asst.
Prof., Department of
Neurology, University of
Pittsburgh)

Website and profile

- Website: www.drpriyamdas.com
- Github id: priyamdas2
- Email: dasp4@vcu.edu

References I

Das, P. (2021). Recursive modified pattern search on high-dimensional simplex : A blackbox optimization technique. *The Indian Journal of Statistics - Sankhya B*, 83:440–483.

Das, P. (2023). Black-box optimization on hyper-rectangle using recursive modified pattern search and application to ROC-based classification problem. *Sankhya B*, 85:365–404.

Das et al. (2022). Estimating the optimal linear combination of predictors using spherically constrained optimization. *BMC Bioinformatics*, 23(Suppl 3):436.

Das et al. (2023a). Clustering sequence data with mixture markov chains with covariates using multiple simplex constrained optimization routine (msicor). *Journal of Computational and Graphical Statistics*, 33(2):379–392.

Das et al. (2023b). Utilizing biologic disease-modifying anti-rheumatic treatment sequences to subphenotype rheumatoid arthritis. *Arthritis Research and Therapy*, 25(1):1–7.

Fermi, E. and Metropolis, N. (1952). Numerical solution of a minimum problem. los alamos unclassified report la–1492. *Los Alamos National Laboratory, Los Alamos, USA*.

Helske, S. and Helske, J. (2019). Mixture hidden Markov models for sequence data: the seqHMM package in R. *Journal of Statistical Software*, 88(3).

Thanks! Questions?