A Statistical Framework for Understanding Causal Effects that Vary by Treatment Initiation Time in EHR-based Studies

Luke Benz
ENAR
March 17, 2026

IBS ENAR

HARVARD T.H. CHAN
SCHOOL OF PUBLIC HEALTH

Many Collaborators

Sebastien Haneuse
Harvard

Alex Levis
Penn

Rui Wang
Harvard

Rajarshi Mukherjee
Harvard

David Arterburn
Kaiser Permante

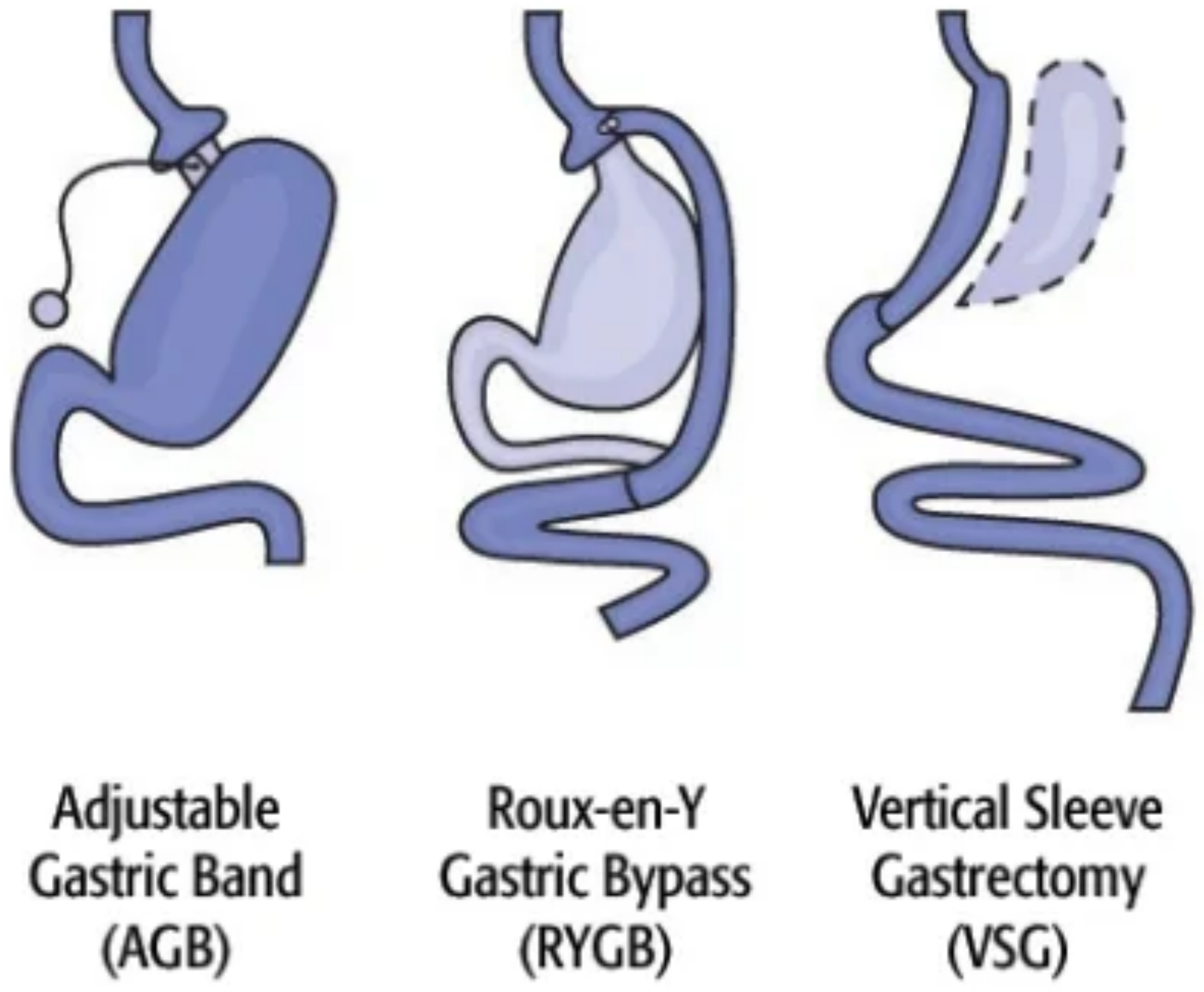
Susan Shorteed
Kaiser Permanente/
University of Washington

Catherine Lee
UCSF

Heidi Fischer
Kaiser Permanente

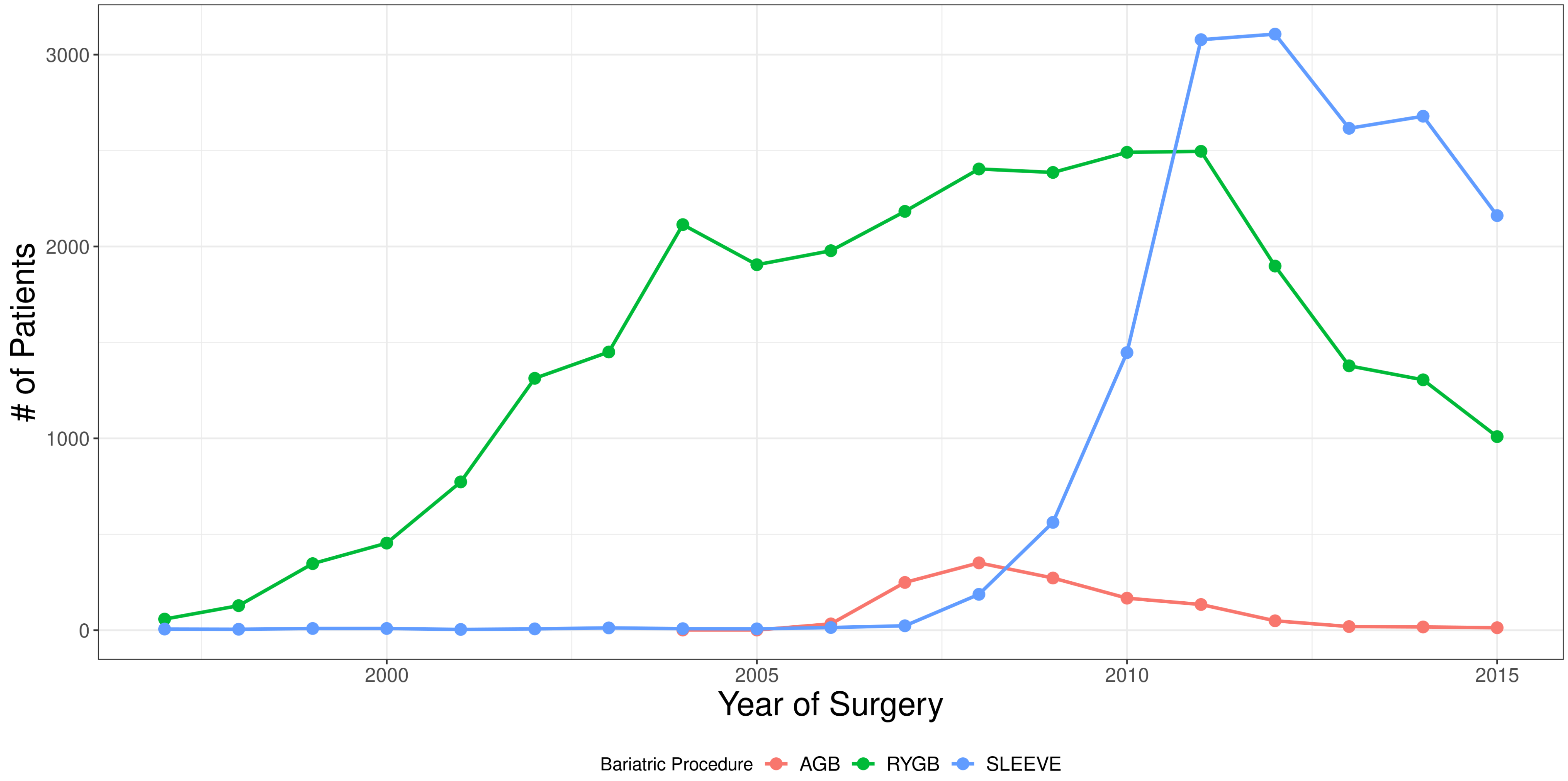
Bariatric Surgery

- Bariatric surgery is a weight-loss surgery
  - Typical candidates have BMI $\geq 35\text{kg/m}^2$
- Sleeve Gastrectomy (SG) is a newer procedure than Roux-en-Y Gastric Bypass (RYGB)
  - SG surpassed RYGB in popularity in late 2000s/ early 2010s
  - Less invasive and technically less complex
- DURABLE: NIH funded study of long-term outcomes following bariatric surgery, particularly in relation to non-surgical patients
  - Kaiser Permanente (Washington, N. California, S. California)
  - 1997-2015; ~45,000 surgical patients and 1.7 million non surgical patients

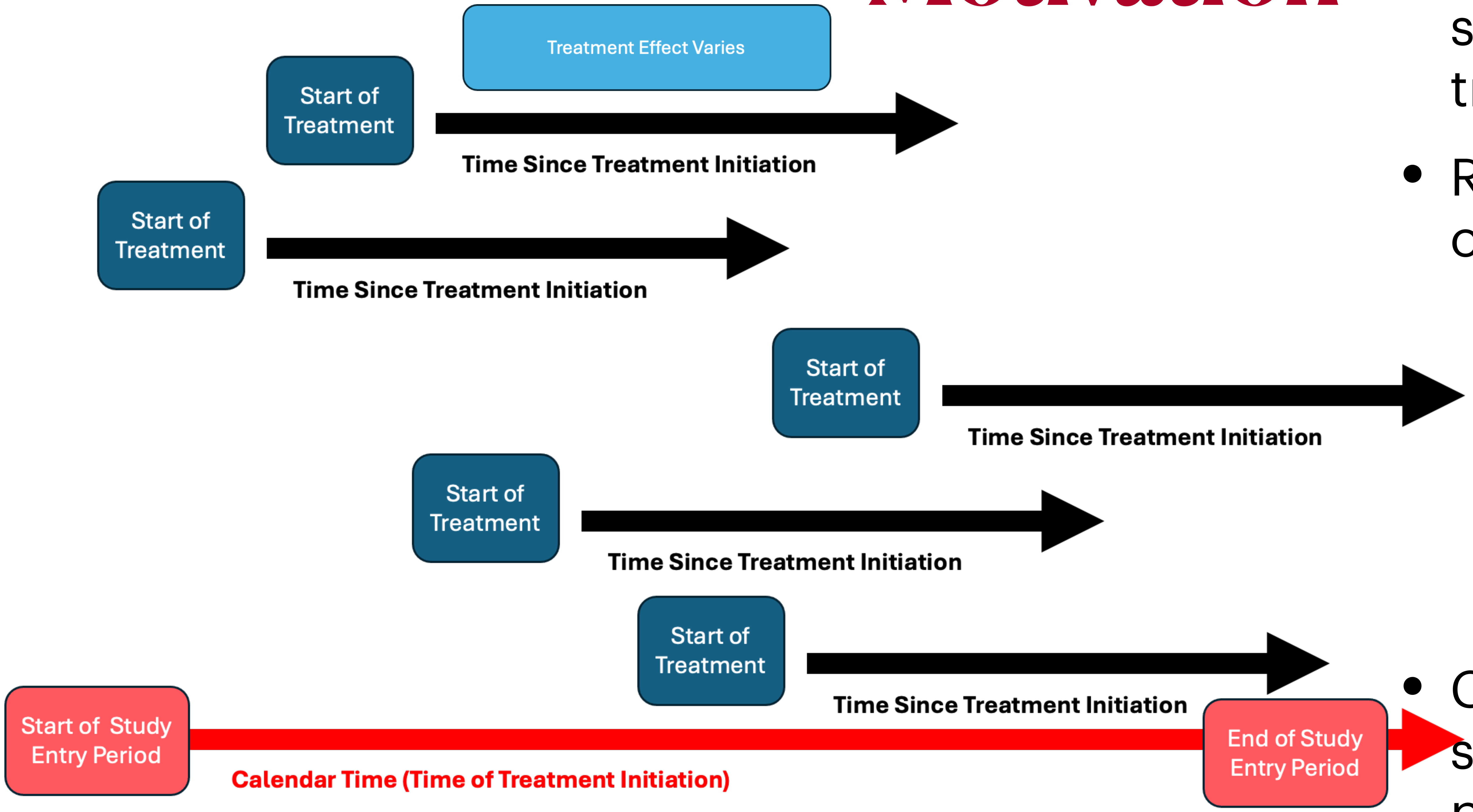

Adjustable Gastric Band (AGB)
Roux-en-Y Gastric Bypass (RYGB)
Vertical Sleeve Gastrectomy (VSG)

Distribution of Bariatric Surgery Procedures

DURABLE Database: 1997-2015



Motivation



Treatment Effect Varies

Start of Treatment

Time Since Treatment Initiation

Start of Treatment

Time Since Treatment Initiation

Start of Treatment

Time Since Treatment Initiation

Start of Treatment

Time Since Treatment Initiation

Start of Treatment

Time Since Treatment Initiation

Start of Study Entry Period

End of Study Entry Period

Calendar Time (Time of Treatment Initiation)

- **Time-varying effects** usually speaks to time **since** starting a treatment
- Real-world treatments evolve over time
  - Standard of care (best practice)
  - Procedures where techniques can change (bariatric surgery)
- Causal effects in EHR-based studies with long study entry periods may vary over the course **of treatment initiation time (calendar time)**

Motivation

- Common viewpoint in pharamcoepi is that this is a potential source of bias
    - Towards estimating an implicitly assumed **constant** effect (across calendar time)
    - Reasonable viewpoint if treatments (e.g., drugs or medications) are not changing themselves
    - But what if our treatment strategies have to be more loosely defined?
        - Calendar-time varying effects = fundamental changes in treatment efficacy?
- In this project, we introduce a framework for
    1. Efficient estimation of causal effects that may vary by calendar time
    2. Determining **how** effects vary by treatment time
    3. Determining **why** effects vary by treatment time

An EHR-based Study of Weight Loss Following Bariatric Surgery

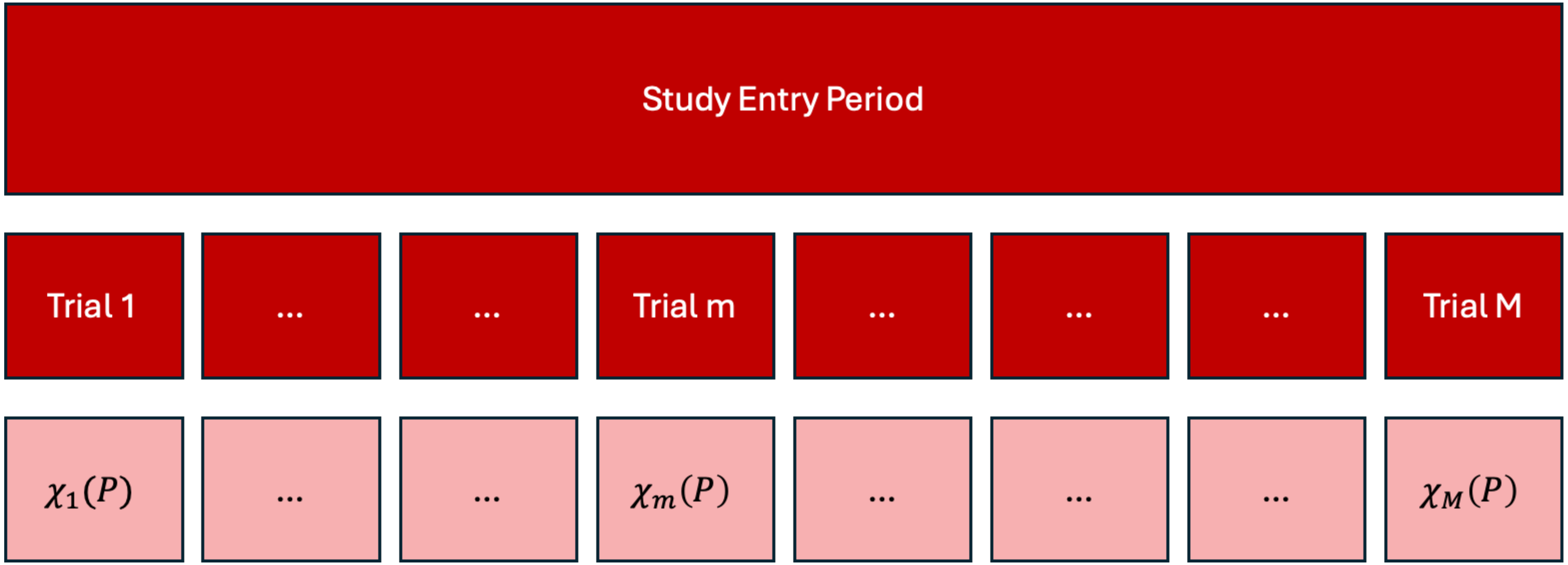
- Examine relative weight change at 4 pre-specified times post-surgery
    - {6 months, 1 year, 2 years, 3 years}
- DURABLE database between 2005-2011
    - 17,905 surgical patients
    - 933,044 non-surgical patients
- Sequence of 84 target trials (one per month between Jan. 2005 and Dec. 2011)
    - Necessary to define “time zero”
- Treatment comparisons
    - Surgery vs. No Surgery
    - {RYGB, SG} vs. No-Surgery (separately)
    - RYGB vs. SG

Notation

- Trial index $m \in \{1,\ldots,M\}$ ($M = 84$ in our running example)
- Y_m: continuous outcome (e.g., % weight loss from baseline of trial m)
- A_m: Binary treatment
- $\boldsymbol{L}_m$: Covariates (confounders, effect modifiers)
- E_m : Eligibility indicator
    - As we will see, we can treat the periods that someone isn't in the EHR as ineligible ($E_m = 0$)

$$O_1, \ldots O_n \overset{iid}{\sim} P$$
$$O = (E_1, E_1\boldsymbol{L}_1, E_1\mathrm{A}_1, E_1Y_1, \ldots, E_M, E_M\boldsymbol{L}_M, E_M\mathrm{A}_M, E_MY_M)$$

Calendar Time-Specific Average Treatment Effect



Study Entry Period

Trial 1	...	...	Trial m	...	...	...	Trial M
$\chi_1(P)$	...	...	$\chi_m(P)$	...	...	...	$\chi_M(P)$

$$\chi_m(P) = \mathbb{E}_P[Y_m(1) - Y_m(0) \mid E_m = 1]$$

- Leverage the sequential nature of the design to define initiator cohorts across calendar time
- Implicit in the definition of $\chi_m(P)$ is that comparisons are being made between whatever "versions" of treatment are in use at trial m
  - ITT analogue — don't care what happens after trial m

Nonparametric identification + efficiency theory

1. **Consistency** $Y_m(A_m) = Y_m$ when $E_m = 1$, almost surely, for all m
2. **Positivity** $0 < \epsilon \leq P(A_m = 1 \mid \boldsymbol{L}_m, E_m = 1) \leq 1 - \epsilon < 1$ almost surely, for all m
3. **No Unmeasured Confounding** $Y_m(a_m) \perp\!\!\!\perp A_m \mid \boldsymbol{L}_m,\ E_m = 1$ for all m

At the outset, estimation of $\chi_m(P)$ may seem straightforward, for example on the basis of it's nonparametric influence function:

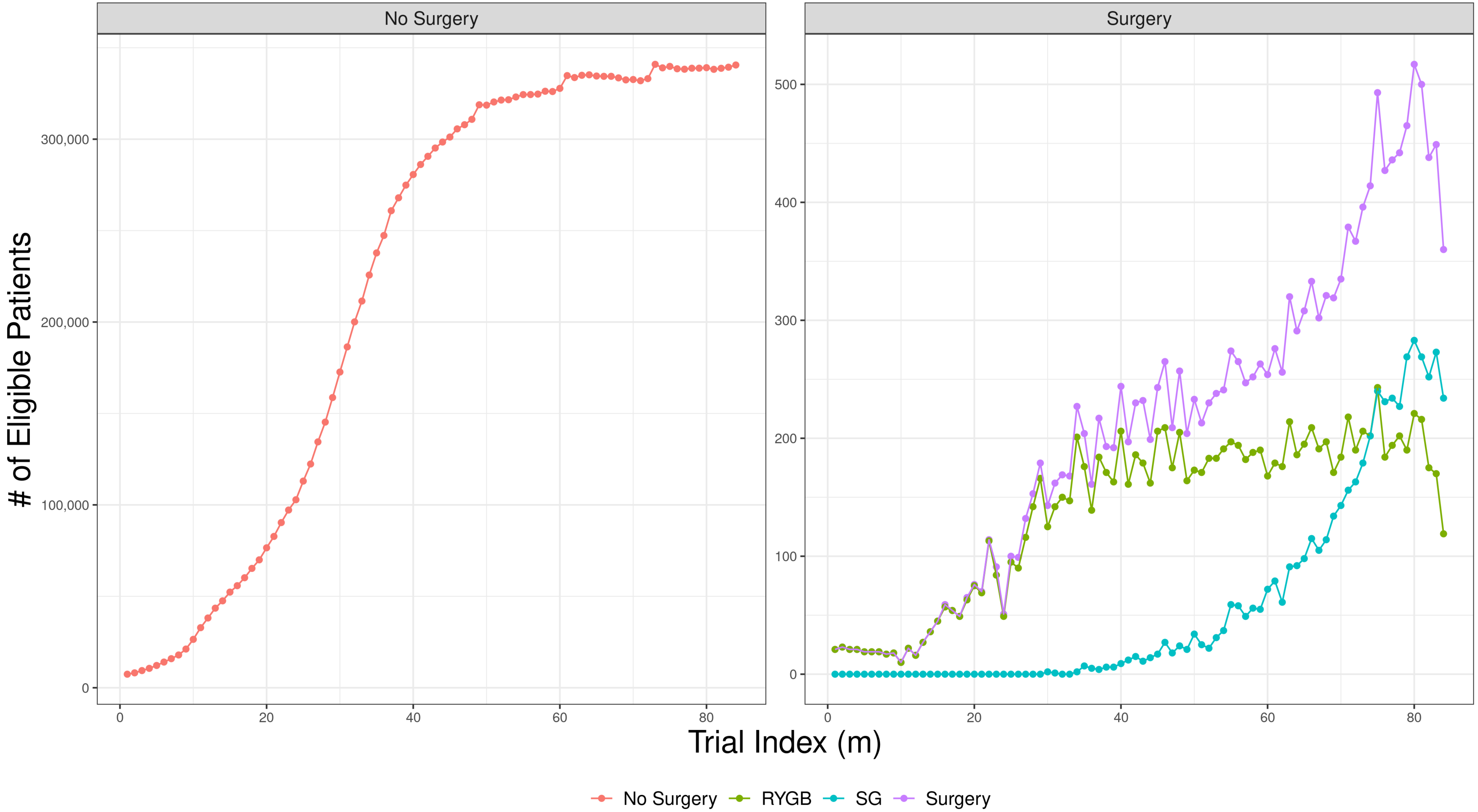
$$\dot{\chi}^*_m(O;P) = \frac{\mathbf{1}(E_m = 1)}{P(E_m = 1)} \left\{ \mu_m(1,\boldsymbol{L}_m) - \mu_m(0,\boldsymbol{L}_m) - \chi_m(P) + \left(\frac{A_m}{\pi_m(\boldsymbol{L}_m)} - \frac{1 - A_m}{1 - \pi_m(\boldsymbol{L}_m)} \right) \Big(Y_m - \mu_m(A_m, \boldsymbol{L}_m) \Big) \right\}$$

$$\mu_m(a_m, \boldsymbol{L}_m) = \mathbb{E}[Y_m \mid A_m = a_m,\ \boldsymbol{L}_m,\ E_m = 1]$$

$$\pi_m(\boldsymbol{L}_m) = P(A_m = 1 \mid \boldsymbol{L}_m)$$

But...how much data do we have at trial *m*?

Trial Effective Sample Sizes



Marginal Structural Model + Projection Parameter Approach

- For a fixed m, insufficient for **ML-based** nuisance models
- Ultimately, our interest is in the **trend** in $\chi_m(P)$ across calendar time (m)
- Can we pool **across trials** without imposing strong parametric assumptions?
- Consider the marginal structural model (MSM)

$$\mathbb{E}[Y_m(a_m = 1) - Y_m(a_m = 0) \mid E_m = 1] \approx \psi(m; \boldsymbol{\beta})$$

- When MSM is saturated, model is correct
- Otherwise, we adopt **assumption-lean** projection approach:

$$\boldsymbol{\beta}(P) = \underset{\boldsymbol{\beta} \in \mathbb{R}^k}{\operatorname{argmin}} \sum_{m=1}^{M} w(m) P(E_m = 1)\Big(\chi_m(P) - \psi(m; \boldsymbol{\beta})\Big)^2$$

Benefits of Projection Parameter Approach

- Projection approach (Neugebaur & van Der Laan, 2007) is inherently model-agnostic i.e., valid even if MSM not correctly specified
  - Greater transparency about how information is shared across time
  - We combine with **pooled** but **flexible** nuisance models

- Directly feeds into model selection strategy for comparing candidate MSMs
  - Identify functional form of **how** treatment effects vary over calendar time
  - Including the possibility of a constant effect

Estimation Strategy

- Estimation/inference on trend still require estimates of trial-specific effects
  - Use pooled models for nuisance functions $\tilde{\mu}(A_m, \boldsymbol{L}_m, m)$ and $\tilde{\pi}(\boldsymbol{L}_m, m)$ estimated on pooled dataset

$$\mathscr{D} = \bigcup_{m=1}^{M} \left\{ (\boldsymbol{L}_{m,i}, A_{m,i}, Y_{m,i}, m) \right\}_{i:E_{m,i}=1}$$

- Use nonparametric/flexible machine learning models for pooled nuisance functions for influence-function based estimator

$$\widehat{\chi}_m = \mathbb{P}_n \left[\frac{\mathbf{1}(E_m = 1)}{\mathbb{P}_n[E_m = 1]} \left\{ \widehat{\mu}_m(1, \boldsymbol{L}_m) - \widehat{\mu}_m(0, \boldsymbol{L}_m) - \left(\frac{A_m}{\widehat{\pi}_m(\boldsymbol{L}_m)} - \frac{1 - A_m}{1 - \widehat{\pi}_m(\boldsymbol{L}_m)} \right) \left(Y_m - \widehat{\mu}_m(A_m, \boldsymbol{L}_m) \right) \right\} \right]$$

Model Selection Among Candidate MSMs

- Let $\{\hat{\psi}_1, \ldots, \hat{\psi}_K\}$ denote a set of K candidate MSMs
  - Constant $\psi(m; \boldsymbol{\beta}) = \beta$;
  - Linear $\psi(m; \boldsymbol{\beta}) = \beta_0 + \beta_1 m$;
  - Higher order polynomials and splines
- Loss Function:

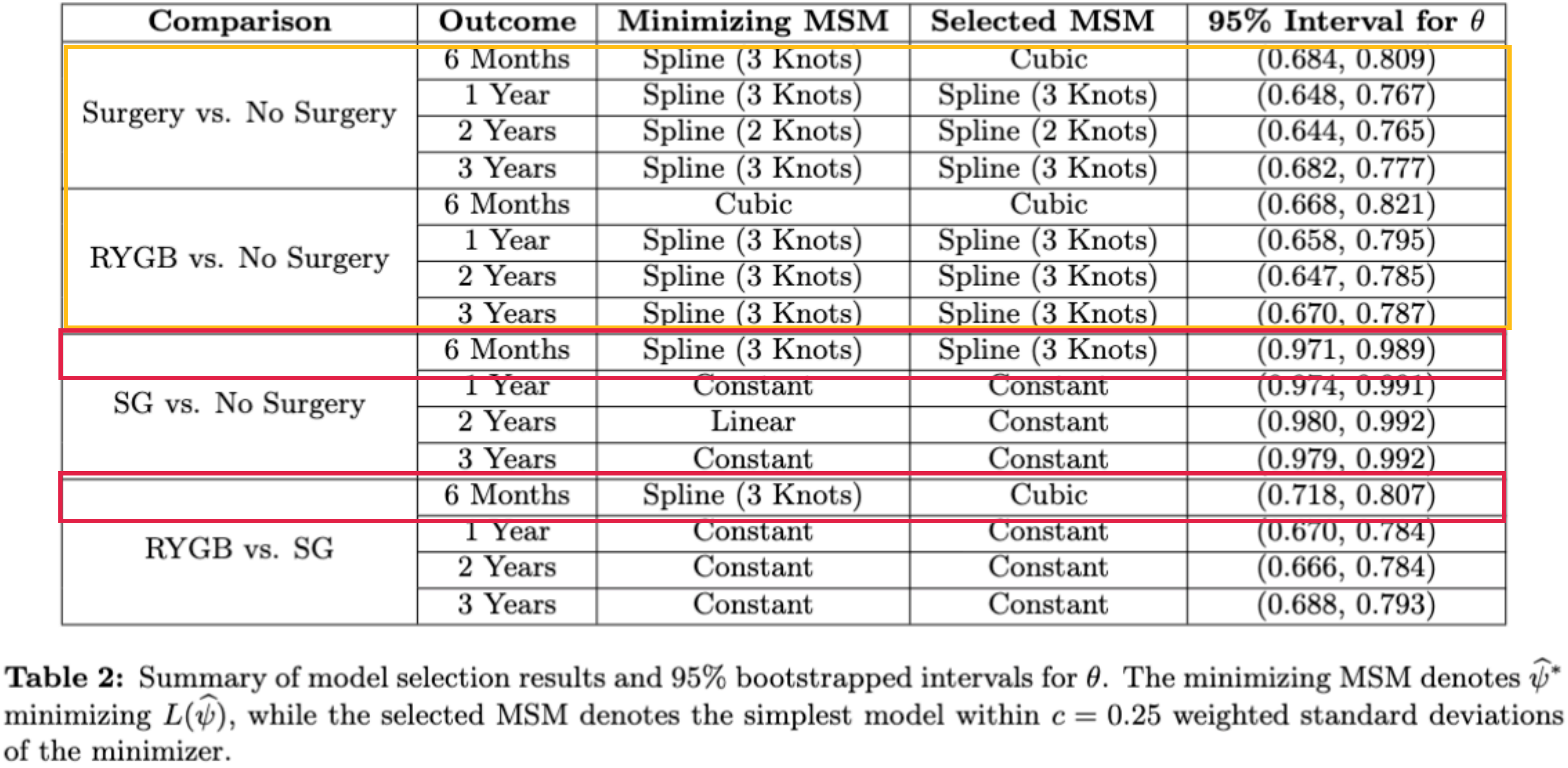
$$L(\hat{\psi}_k) = \mathbb{P}_n \left[\sum_{m=1}^{M} w(m) \left\{ \psi_k(m; \widehat{\boldsymbol{\beta}}_k)^2 - 2\psi_k(m; \widehat{\boldsymbol{\beta}}_k) \dot{\chi}_m(O; \hat{P}) \right\} \right]$$

- Reasonable loss function to work with because it reflects estimating equation-based estimator of (pseudo)risk if one treats $\hat{\psi}_k$ as fixed

Calendar Time–Specific Treatment Effect Estimates

6 Months
1 Year
2 Years
3 Years
Surgery vs. No Surgery
RYGB vs. No Surgery
SG vs. No Surgery
RYGB vs. SG
Difference in Mean %–Weight Change
−20%
−24%
−28%
−32%
−20%
−25%
−30%
−14%
−16%
−18%
−20%
−22%
−4%
−6%
−8%
−10%
0
20
40
60
80
Trial Index (m)
ψ(m;β)
Constant
Linear
Cubic
Spline (2 Knots)
Spline (3 Knots)

Results

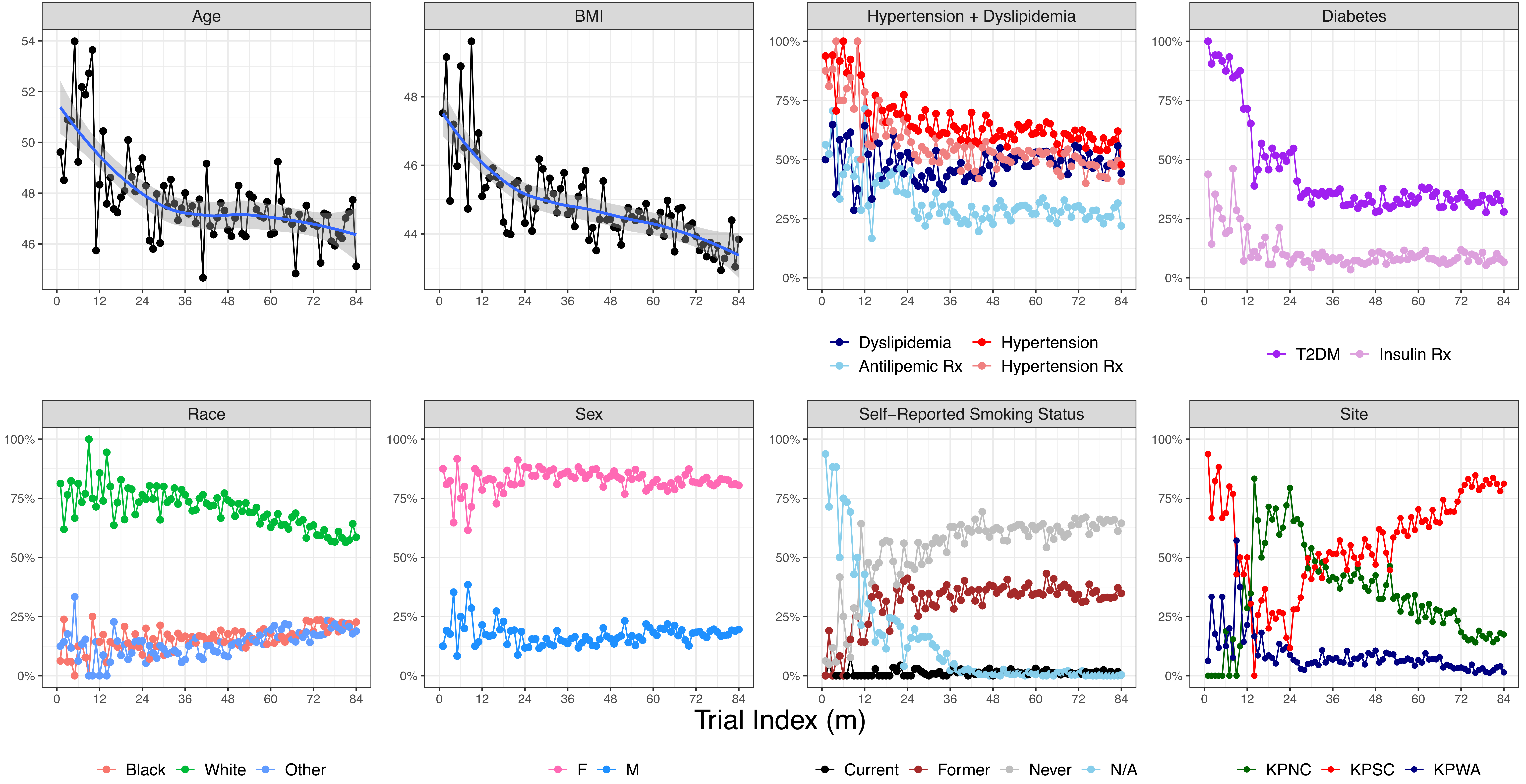


Comparison	Outcome	Minimizing MSM	Selected MSM	95% Interval for θ
Surgery vs. No Surgery	6 Months	Spline (3 Knots)	Cubic	(0.684, 0.809)
	1 Year	Spline (3 Knots)	Spline (3 Knots)	(0.648, 0.767)
	2 Years	Spline (2 Knots)	Spline (2 Knots)	(0.644, 0.765)
	3 Years	Spline (3 Knots)	Spline (3 Knots)	(0.682, 0.777)
RYGB vs. No Surgery	6 Months	Cubic	Cubic	(0.668, 0.821)
	1 Year	Spline (3 Knots)	Spline (3 Knots)	(0.658, 0.795)
	2 Years	Spline (3 Knots)	Spline (3 Knots)	(0.647, 0.785)
	3 Years	Spline (3 Knots)	Spline (3 Knots)	(0.670, 0.787)
SG vs. No Surgery	6 Months	Spline (3 Knots)	Spline (3 Knots)	(0.971, 0.989)
	1 Year	Constant	Constant	(0.974, 0.991)
	2 Years	Linear	Constant	(0.980, 0.992)
	3 Years	Constant	Constant	(0.979, 0.992)
RYGB vs. SG	6 Months	Spline (3 Knots)	Cubic	(0.718, 0.807)
	1 Year	Constant	Constant	(0.670, 0.784)
	2 Years	Constant	Constant	(0.666, 0.784)
	3 Years	Constant	Constant	(0.688, 0.793)

Table 2: Summary of model selection results and 95% bootstrapped intervals for θ. The minimizing MSM denotes $\widehat{\psi}^*$ minimizing $L(\widehat{\psi})$, while the selected MSM denotes the simplest model within $c = 0.25$ weighted standard deviations of the minimizer.

Distribution of Key Covariates in Study Population

Eligible Patients Undergoing Bariatric Surgery



Results Summary

- **Surgery vs. No-Surgery**
    - 3 Years: -27.4% at $m = 1$ vs. -18.3% at $m = 84$ [Change of 9.1%]
    - Constant effect: -19.8%
        - **Misses a lot of the story**
- Not surprising that surgery is “less effective” over time because distribution of procedures is changing
    - Good validation that our method is picking up clinically meaningful changes
    - Many papers use catch-all of “bariatric surgery” → **what that label means is changing**
- Tools for disentangling changes in treatment effect from changes in population receiving treatment → see paper

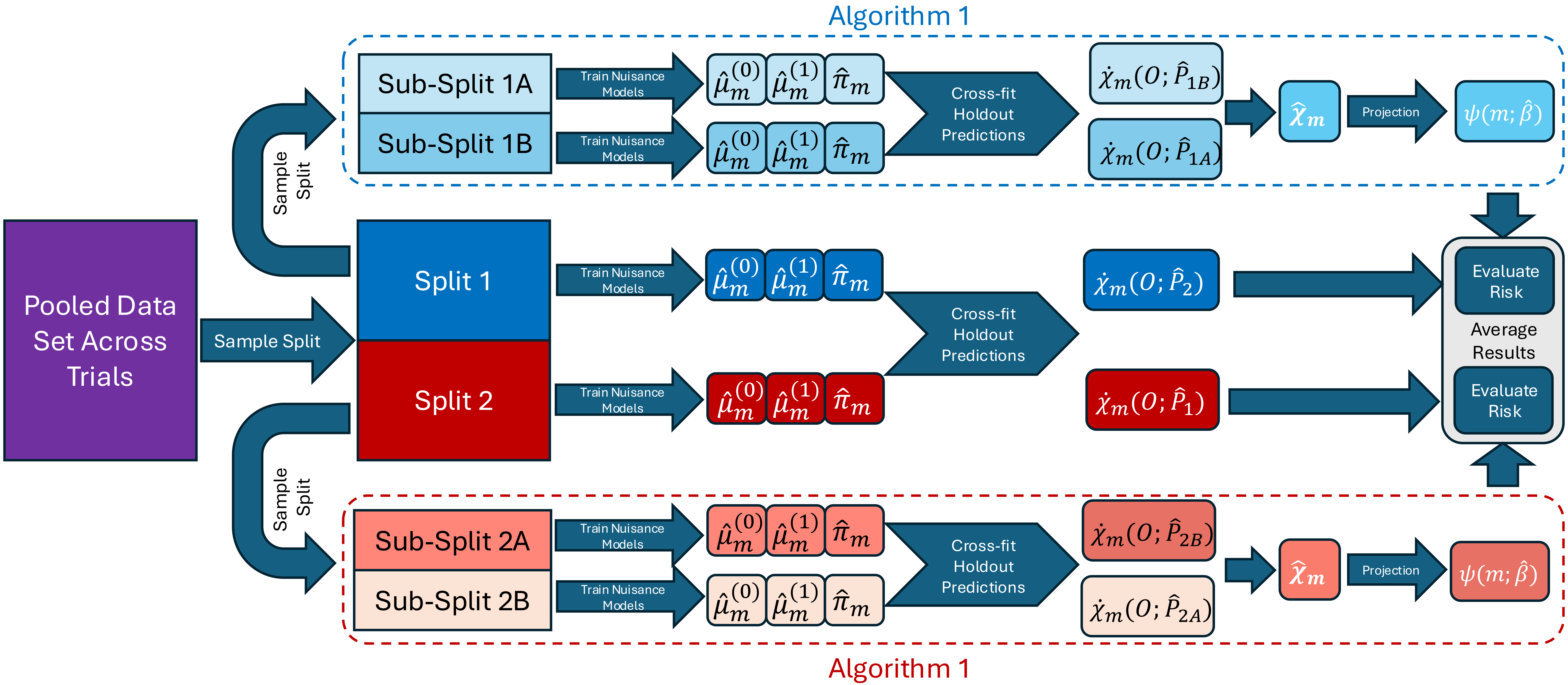
Paper

GitHub

Appendix

HARVARD T.H. CHAN
SCHOOL OF PUBLIC HEALTH

Evaluating $L(\hat{\psi}_k)$


Algorithm 1
Pooled Data Set Across Trials
Sample Split
Sample Split
Sample Split
Split 1
Split 2
Sub-Split 1A
Sub-Split 1B
Sub-Split 2A
Sub-Split 2B
Train Nuisance Models
Train Nuisance Models
Train Nuisance Models
Train Nuisance Models
Train Nuisance Models
Train Nuisance Models
$\hat{\mu}_m^{(0)}$ $\hat{\mu}_m^{(1)}$ $\hat{\pi}_m$
$\hat{\mu}_m^{(0)}$ $\hat{\mu}_m^{(1)}$ $\hat{\pi}_m$
$\hat{\mu}_m^{(0)}$ $\hat{\mu}_m^{(1)}$ $\hat{\pi}_m$
$\hat{\mu}_m^{(0)}$ $\hat{\mu}_m^{(1)}$ $\hat{\pi}_m$
$\hat{\mu}_m^{(0)}$ $\hat{\mu}_m^{(1)}$ $\hat{\pi}_m$
$\hat{\mu}_m^{(0)}$ $\hat{\mu}_m^{(1)}$ $\hat{\pi}_m$
Cross-fit Holdout Predictions
Cross-fit Holdout Predictions
Cross-fit Holdout Predictions
$\dot{\chi}_m(O;\hat{P}_{1B})$
$\dot{\chi}_m(O;\hat{P}_{1A})$
$\hat{\chi}_m$
Projection
$\psi(m;\hat{\beta})$
$\dot{\chi}_m(O;\hat{P}_2)$
$\dot{\chi}_m(O;\hat{P}_1)$
Evaluate Risk
Average Results
Evaluate Risk
$\dot{\chi}_m(O;\hat{P}_{2B})$
$\dot{\chi}_m(O;\hat{P}_{2A})$
$\hat{\chi}_m$
Projection
$\psi(m;\hat{\beta})$
Algorithm 1

Estimation Strategy

- As long as our candidate MSM is twice differentiable in $\boldsymbol{\beta}$, then under A1-A3, $\boldsymbol{\beta}(P)$ is identified as the solution to the estimating equation

$$0 = \sum_{m=1}^{M} w(m)P(E_m = 1)\,\nabla_{\boldsymbol{\beta}}\psi(m;\boldsymbol{\beta})[\chi_m(P) - \psi(m;\boldsymbol{\beta})]$$

- Furthermore, the influence function of $\boldsymbol{\beta}(P)$ (up to proportionality) is given by

$$\dot{\boldsymbol{\beta}}^{*}(O;P) \propto \dot{\boldsymbol{\beta}}^{\dagger}(O;P) = \sum_{m=1}^{M} P(E_m = 1)w(m)\,\nabla_{\boldsymbol{\beta}}\psi(m;\boldsymbol{\beta})\,|_{\boldsymbol{\beta}=\boldsymbol{\beta}(P)}\left(\dot{\chi}_m^{\dagger}(O;P) - \psi\big(m;\boldsymbol{\beta}(P)\big)\right)$$

$$\dot{\chi}_m^{\dagger}(O;P) = \mu_m(1,\boldsymbol{L}_m) - \mu_m(0,\boldsymbol{L}_m) + \left(\frac{A_m}{\pi_m(\boldsymbol{L}_m)} - \frac{1 - A_m}{1 - \pi_m(\boldsymbol{L}_m)}\right)\left(Y_m - \mu_m(A_m, \boldsymbol{L}_m)\right)$$

- Motivates estimator which solves $\mathbb{P}_n[\dot{\boldsymbol{\beta}}^{\dagger}(O;\hat{P})] = 0$, that is

$$0 = \sum_{m=1}^{M} w(m)\mathbb{P}_n(E_m = 1)\,\nabla_{\boldsymbol{\beta}}\psi(m;\boldsymbol{\beta})\,|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}}\left\{\boxed{\hat{\chi}_m} - \psi(m;\hat{\boldsymbol{\beta}})\right\}$$

A Cross-Trial Contrast to Remove Distribution Shift

$$\chi_{j,m}(P) = \int_{\mathscr{L}} \boxed{\mathbb{E}\Big[Y_m(a_m = 1) - Y_m(a_m = 0)} \mid \boxed{\boldsymbol{L}_m = \boldsymbol{\ell}, g(\boldsymbol{L}_m) = 1\Big] dP_{\boldsymbol{L}_j \mid E_j = 1}(\boldsymbol{\ell} \mid E_j = 1)}$$

- Difference in mean counterfactural outcomes at time m
- Re-weight (standardize) the covariate distribution among eligible population at time m to match the covariate distribution of eligible population at time j

- Critical to the definition of $\chi_{j,m}(P)$ is a common population with which to standardize treatment effects to across calendar time
- $m_1 \neq m_2$ differences between $\chi_{j,m_1}(P)$ and $\chi_{j,m_2}(P)$ not be attributable to covariate shift
  - Underlying population on which the causal contrast is defined is the same
  - Will exploit this to quantify the role of effect modification over time

Summarizing Effect Modification

$$\sigma_m^2 = \frac{1}{M-1}\sum_{j=1}^{M}\left(\chi_m(P) - \chi_{m,j}(P)\right)^2$$

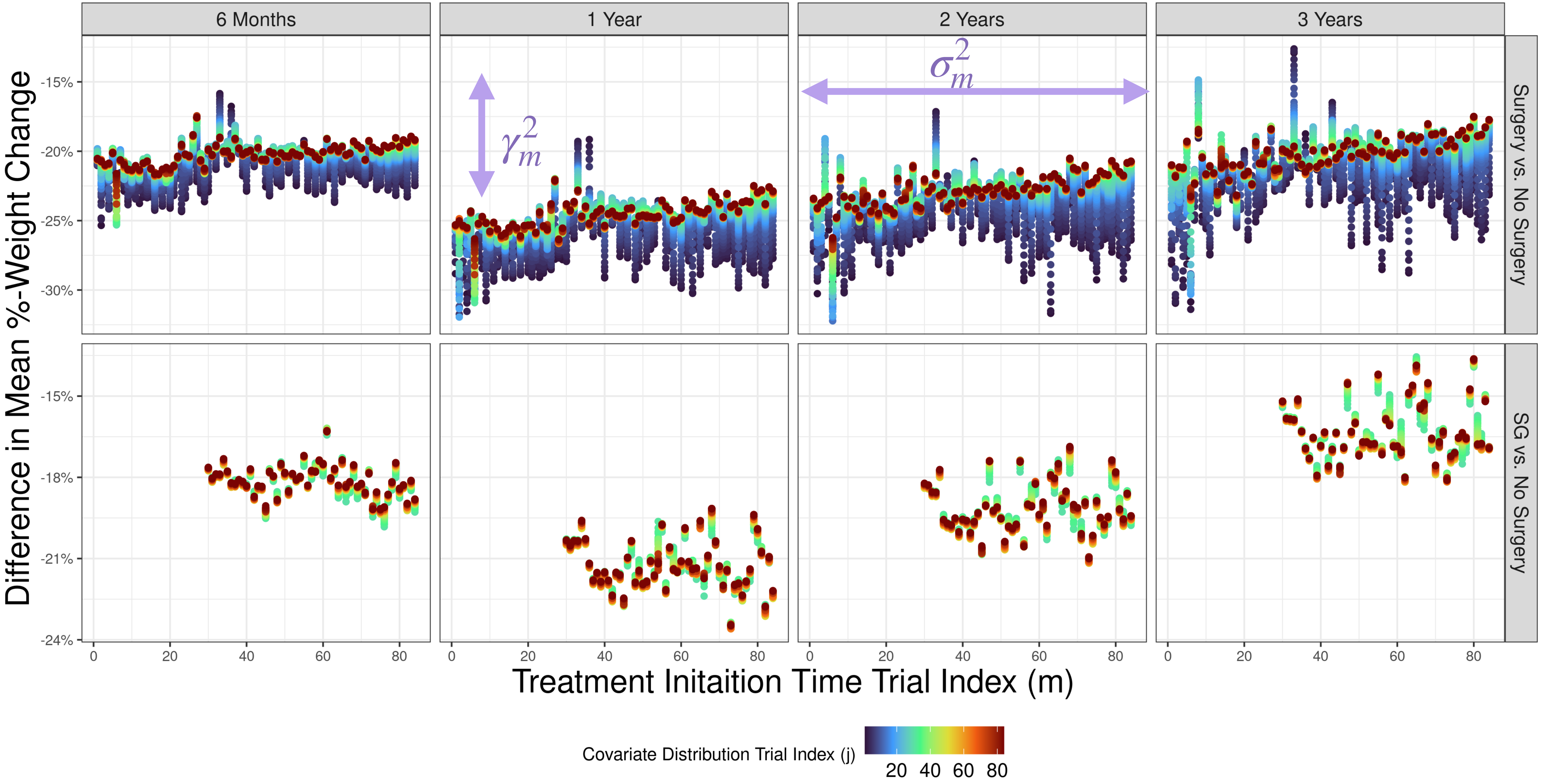
$$\gamma_m^2 = \frac{1}{M-1}\sum_{j=1}^{M}\left(\chi_m(P) - \chi_{j,m}(P)\right)^2$$

- σ_m^2: How much variation in treatment effects had trial m population received treatment at other points in time (**Change treatment time for fixed population)**
- γ_m^2: characterizes the variability around $\chi_m(P)$ across all trial-eligible populations, had each been treated at time m **(Change population for fixed treatment time)**

$$\theta = \frac{1}{M}\sum_{m=1}^{M}\theta_m = \frac{1}{M}\sum_{m=1}^{M}\frac{\sigma_m^2}{\sigma_m^2 + \gamma_m^2}$$

- Close to 0: variation driven by differences in population (effect modification)
- Close to 1: variation not driven by differences in population (treatment efficacy)

Illustration of $\widehat{\chi}_{j,m}$ for Select Comparisons



Estimation of Cross-Trial Effects

- **Cross-Trial Overlap:** $p(\boldsymbol{\ell} \mid E_j = 1) > 0 \implies p(\boldsymbol{\ell} \mid E_m = 1) > 0$

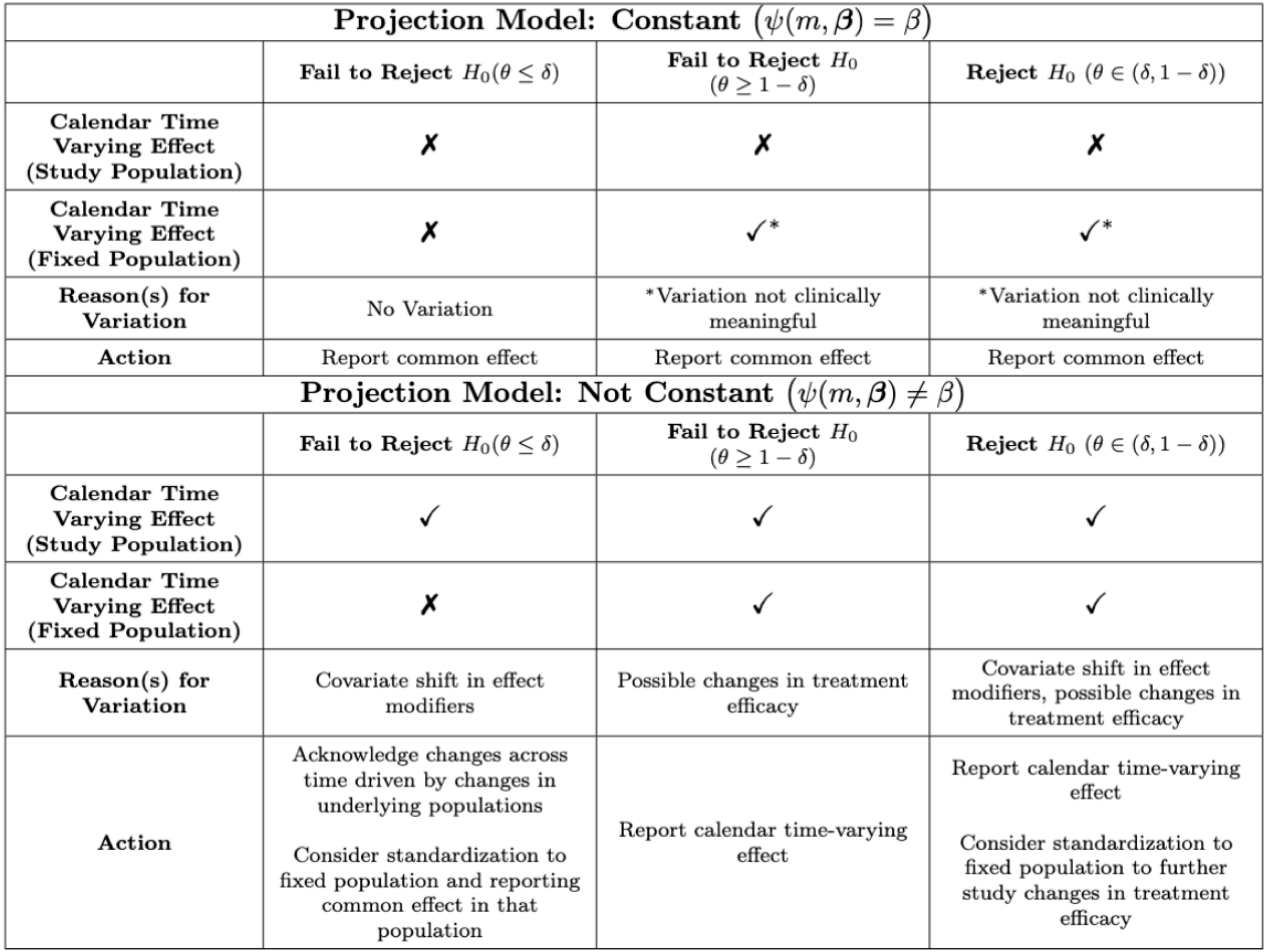
$$\chi_{j,m}(P) = \mathbb{E}[\mu_m(1,\boldsymbol{L}_j) - \mu_m(0,\boldsymbol{L}_j) \mid E_j = 1]$$

$$\dot{\chi}^*_{j,m}(O;P) = \frac{\mathbf{1}(E_j = 1)}{P(E_j = 1)}\left\{\mu_m(1,\boldsymbol{L}_j) - \mu_m(0,\boldsymbol{L}_j) - \chi_{j,m}(P)\right\} + \frac{\mathbf{1}(E_m = 1)}{P(E_m = 1)}\xi_{j,m}(\boldsymbol{L}_m)\left(\frac{A_m}{\pi_m(\boldsymbol{L}_m)} - \frac{1 - A_m}{1 - \pi_m(\boldsymbol{L}_m)}\right)\left(Y_m - \mu_m(A_m, \boldsymbol{L}_m)\right)$$

- G-formula using outcome regression at time m on covariate distribution from time j
- AIPW residual from eligible patients in trial m
- Transport ratio required to “make trial m distribution look like that of trial j”

$$\xi_{j,m}(\boldsymbol{\ell}) = \frac{p(\boldsymbol{L}_j = \ell \mid E_j = 1)}{p(\boldsymbol{L}_m = \ell \mid E_m = 1)} = \frac{P(E_m = 1)P(T = j \mid \boldsymbol{L} = \boldsymbol{\ell}, E = 1)}{P(E_j = 1)P(T = m \mid \boldsymbol{L} = \boldsymbol{\ell}, E = 1)}$$

Practical Guidance: How and Why Effects Vary over Calendar Time



Projection Model: Constant $(\psi(m, \boldsymbol{\beta}) = \beta)$

	Fail to Reject $H_0(\theta \leq \delta)$	**Fail to Reject** H_0 $(\theta \geq 1 - \delta)$	**Reject** H_0 $(\theta \in (\delta, 1 - \delta))$
Calendar Time Varying Effect (Study Population)	✗	✗	✗
Calendar Time Varying Effect (Fixed Population)	✗	✓*	✓*
Reason(s) for Variation	No Variation	*Variation not clinically meaningful	*Variation not clinically meaningful
Action	Report common effect	Report common effect	Report common effect

Projection Model: Not Constant $(\psi(m, \boldsymbol{\beta}) \neq \beta)$

	Fail to Reject $H_0(\theta \leq \delta)$	**Fail to Reject** H_0 $(\theta \geq 1 - \delta)$	**Reject** H_0 $(\theta \in (\delta, 1 - \delta))$
Calendar Time Varying Effect (Study Population)	✓	✓	✓
Calendar Time Varying Effect (Fixed Population)	✗	✓	✓
Reason(s) for Variation	Covariate shift in effect modifiers	Possible changes in treatment efficacy	Covariate shift in effect modifiers, possible changes in treatment efficacy
Action	Acknowledge changes across time driven by changes in underlying populations Consider standardization to fixed population and reporting common effect in that population	Report calendar time-varying effect	Report calendar time-varying effect Consider standardization to fixed population to further study changes in treatment efficacy

Hypothesis Testing for θ

$$H_0 : \theta \in \{0,1\} \text{ vs. } H_1 : \theta \in (0,1)$$

- **Challenge 1:** Test is at boundary point(s)
  - Asymptotic distribution of θ under some complex ratio of mixtures of chi-squared distributions
  - **Solution 1:** bootstrap
- **Challenge 2:** Can't do a full nonparametric bootstrap
  - Computationally too intensive due to sample splitting, re-fitting pooled models
  - # of predictions required in cross-trial effects is on the order of $M \times |\mathcal{D}|$
  - **Solution 2:** Asymptotic version of parametric bootstrap
    - $\boldsymbol{S} \in \mathbb{R}^{M \times M}$ standardization matrix with entries $\chi_{j,m}(P)$
    - $\sqrt{n}(\boldsymbol{S} - \widehat{\boldsymbol{S}}) \xrightarrow{d} \mathcal{N}(0,\boldsymbol{\Sigma})$ where $\boldsymbol{\Sigma} \in \mathbb{R}^{M^2 \times M^2}$ is the covariance matrix of cross-trial influence function contributions
    - Resample $\boldsymbol{S}^{(1)}, \ldots, \boldsymbol{S}^{(B)} \sim \mathcal{N}(\widehat{\boldsymbol{S}}, \widehat{\boldsymbol{\Sigma}}_n)$ and compute $\widehat{\theta}^{(1)}, \ldots, \widehat{\theta}^{(B)}$ for B bootstrap replicates

Hypothesis Testing for θ

- **Challenge 3:** Bootstrapped values of $\widehat{\theta}^{(b)}$ will not be 0, 1 in practice
  - Would always reject H_0
  - **Solution 3:** Modify test away from boundary by some small margin δ

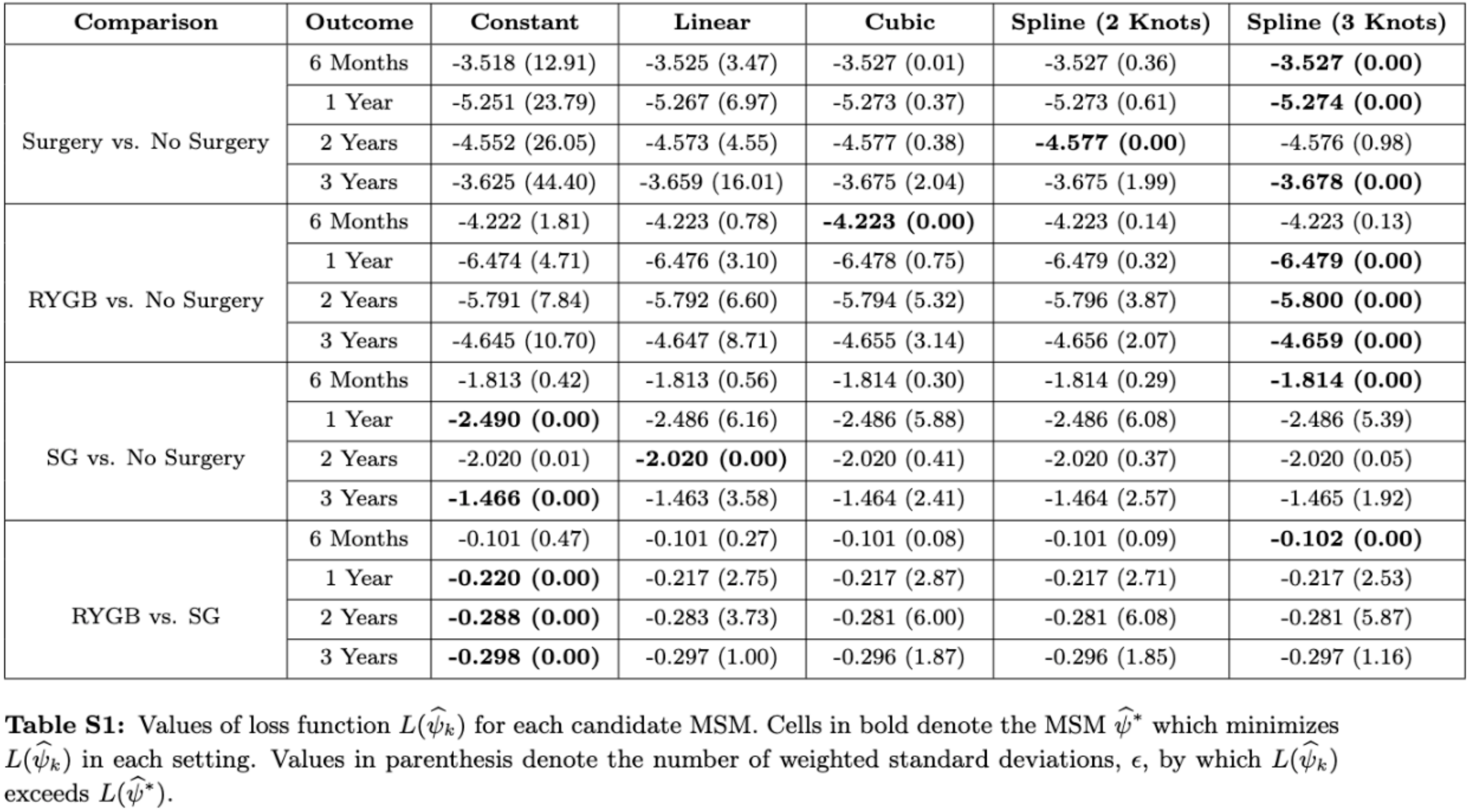
$$H_0 : \theta \in [0,\delta] \cup [1-\delta,1] \text{ vs. } H_1 : \theta \in (\delta,1-\delta)$$

  - One can interpret δ to the fraction of variability in entries in $\boldsymbol{S}$ for which true treatment change is non-negligible in the eyes of the analyst

Why $\widehat{\theta} = 1$ May Not Imply Treatment Efficacy is Changing

- Under assumptions/conditions in this work, changes in $\chi_m(P)$ over time:
  - Underlying treatment efficacy is changing
  - Covariate shift in effect modifier
- Under violations of these assumptions, $\widehat{\theta} > 0$ even when efficacy is unchanged
  - Including m as a covariate in non-parametric pooled models can be a proxy for unmeasured confounders whose distribution varies across time
  - Difference in effect estimates could reflect **both** difference in treatment efficacy and differential bias introduced by unmeasured confounders
  - Similar w/ model misspecification and residual confounding
- If worried about residual confounding in pooled models influencing estimate of trend
  - Set c parameter in MSM selection procedure to be greater
  - Requires changes in $\widehat{\chi}_m$ over time must be greater in order for a non-constant MSM to be selected

$L(\hat{\psi}_k)$ for Candidate MSMs



Comparison	Outcome	Constant	Linear	Cubic	Spline (2 Knots)	Spline (3 Knots)
Surgery vs. No Surgery	6 Months	-3.518 (12.91)	-3.525 (3.47)	-3.527 (0.01)	-3.527 (0.36)	**-3.527 (0.00)**
	1 Year	-5.251 (23.79)	-5.267 (6.97)	-5.273 (0.37)	-5.273 (0.61)	**-5.274 (0.00)**
	2 Years	-4.552 (26.05)	-4.573 (4.55)	-4.577 (0.38)	**-4.577 (0.00)**	-4.576 (0.98)
	3 Years	-3.625 (44.40)	-3.659 (16.01)	-3.675 (2.04)	-3.675 (1.99)	**-3.678 (0.00)**
RYGB vs. No Surgery	6 Months	-4.222 (1.81)	-4.223 (0.78)	**-4.223 (0.00)**	-4.223 (0.14)	-4.223 (0.13)
	1 Year	-6.474 (4.71)	-6.476 (3.10)	-6.478 (0.75)	-6.479 (0.32)	**-6.479 (0.00)**
	2 Years	-5.791 (7.84)	-5.792 (6.60)	-5.794 (5.32)	-5.796 (3.87)	**-5.800 (0.00)**
	3 Years	-4.645 (10.70)	-4.647 (8.71)	-4.655 (3.14)	-4.656 (2.07)	**-4.659 (0.00)**
SG vs. No Surgery	6 Months	-1.813 (0.42)	-1.813 (0.56)	-1.814 (0.30)	-1.814 (0.29)	**-1.814 (0.00)**
	1 Year	**-2.490 (0.00)**	-2.486 (6.16)	-2.486 (5.88)	-2.486 (6.08)	-2.486 (5.39)
	2 Years	-2.020 (0.01)	**-2.020 (0.00)**	-2.020 (0.41)	-2.020 (0.37)	-2.020 (0.05)
	3 Years	**-1.466 (0.00)**	-1.463 (3.58)	-1.464 (2.41)	-1.464 (2.57)	-1.465 (1.92)
RYGB vs. SG	6 Months	-0.101 (0.47)	-0.101 (0.27)	-0.101 (0.08)	-0.101 (0.09)	**-0.102 (0.00)**
	1 Year	**-0.220 (0.00)**	-0.217 (2.75)	-0.217 (2.87)	-0.217 (2.71)	-0.217 (2.53)
	2 Years	**-0.288 (0.00)**	-0.283 (3.73)	-0.281 (6.00)	-0.281 (6.08)	-0.281 (5.87)
	3 Years	**-0.298 (0.00)**	-0.297 (1.00)	-0.296 (1.87)	-0.296 (1.85)	-0.297 (1.16)

Table S1: Values of loss function $L(\widehat{\psi}_k)$ for each candidate MSM. Cells in bold denote the MSM $\widehat{\psi}^*$ which minimizes $L(\widehat{\psi}_k)$ in each setting. Values in parenthesis denote the number of weighted standard deviations, ϵ, by which $L(\widehat{\psi}_k)$ exceeds $L(\widehat{\psi}^*)$.