

Chapter 21: G-Methods for Time-Varying Treatments

Contents

1	21.1 The G-Formula for Time-Varying Treatments (p. 277)	1
1.1	Implementation via Parametric Models	2
1.2	G-Formula for Dynamic Strategies	2
2	21.2 IP Weighting for Time-Varying Treatments (p. 282)	2
2.1	Marginal Structural Models	3
2.2	Censoring Weights	3
3	21.3 A Doubly Robust Estimator for Time-Varying Treatments (p. 286)	4
4	21.4 G-Estimation for Time-Varying Treatments (p. 289)	4
4.1	G-Estimation Algorithm	5
4.2	Rank-Preserving Structural Failure Time Models	6
5	21.5 Censoring Is a Time-Varying Treatment (p. 297)	6
5.1	Handling Censoring via IP Weighting	6
5.2	G-Formula with Censoring	6
6	21.6 The Big G-Formula (p. 300)	7
6.1	Practical Implementation	7
7	Summary	7
8	References	8

Chapters 12–14 introduced three families of methods — the parametric g-formula, IP weighting for marginal structural models, and g-estimation of structural nested models — in the context of a single, time-fixed treatment. Chapter 20 showed that these same methods are needed in the longitudinal setting because standard regression fails when treatment-confounder feedback is present. This chapter extends all three g-methods to **time-varying treatments**, deriving their estimands, identifying assumptions, and practical implementations.

This chapter is based on Hernán and Robins (2020, chap. 21, pp. 277–302).

Roadmap: All three g-methods share the same identifying assumptions (sequential exchangeability, positivity, consistency) but differ in what they model (the outcome distribution, the treatment mechanism, or a structural model for the counterfactual). Understanding their relationships and trade-offs is essential for choosing the right tool in any given applied problem.

1 21.1 The G-Formula for Time-Varying Treatments (p. 277)

The **g-computation formula** (or simply the **g-formula**) is the natural extension of standardization to the longitudinal setting. It expresses the counterfactual mean $E[Y^{\bar{a}\kappa}]$ as a function of the observed data by averaging the conditional expectation of Y over the joint distribution of the time-varying covariates under the counterfactual treatment history.

Definition 1.1 (The G-Formula for Time-Varying Treatments). Under sequential exchangeability, positivity, and consistency,

$$E[Y^{\bar{a}_K}] = \sum_{\bar{l}_K} E[Y \mid \bar{A}_K = \bar{a}_K, \bar{L}_K = \bar{l}_K] \prod_{k=0}^K f(l_k \mid \bar{A}_{k-1} = \bar{a}_{k-1}, \bar{L}_{k-1} = \bar{l}_{k-1}),$$

where the sum is over all possible covariate histories \bar{l}_K , and $f(l_k \mid \cdot)$ is the conditional density (or probability mass) of L_k given past treatment and covariate history.

The product $\prod_{k=0}^K f(l_k \mid \bar{a}_{k-1}, \bar{l}_{k-1})$ is the joint density of the covariate history \bar{L}_K evaluated under the counterfactual treatment \bar{a}_K . In other words, we average the conditional outcome expectation over the distribution that the covariates would have if treatment were set to \bar{a}_K throughout.

1.1 Implementation via Parametric Models

In practice, the g-formula is implemented by fitting parametric models for the outcome and each time-varying covariate, then computing expectations by Monte Carlo simulation:

1. **Fit a model** for $E[Y \mid \bar{A}_K = \bar{a}_K, \bar{L}_K = \bar{l}_K]$.
2. **Fit models** for $f(l_k \mid \bar{a}_{k-1}, \bar{l}_{k-1})$ for each $k = 0, \dots, K$.
3. **Simulate** a large pseudo-population by drawing covariate histories under the counterfactual treatment \bar{a}_K , using the fitted covariate models.
4. **Predict** the outcome for each simulated individual using the fitted outcome model.
5. **Average** the predicted outcomes to estimate $E[Y^{\bar{a}_K}]$.

Repeat steps 3–5 for each comparison strategy to obtain the causal contrast.

Why Monte Carlo? The sum in the g-formula is over all possible covariate histories \bar{l}_K . With K time points and even moderately many covariates at each time point, the number of possible histories grows exponentially, making direct summation infeasible. Monte Carlo simulation sidesteps this by sampling from the covariate models.

Bias from model misspecification: The g-formula is consistent only if all component models are correctly specified. With many time points and covariates, this is a strong requirement. Doubly robust estimators (Section 21.3) and targeted learning approaches can provide some protection against misspecification.

Confidence intervals: Bootstrap resampling is the standard approach for obtaining confidence intervals from the parametric g-formula. Each bootstrap replicate re-estimates all models and re-simulates the pseudo-population.

1.2 G-Formula for Dynamic Strategies

For a dynamic strategy g , the g-formula takes the same form but the treatment history \bar{a}_K is replaced by the counterfactual treatment history $\bar{g}(\bar{l}_K)$, which depends on the (simulated) covariate history. In step 3 of the algorithm above, at each time k the treatment is set to $g_k(\bar{a}_{k-1}, \bar{l}_k)$ based on the simulated covariates, not an externally specified value.

2 21.2 IP Weighting for Time-Varying Treatments (p. 282)

Inverse probability weighting for time-varying treatments creates a pseudo-population in which the time-varying covariates are independent of treatment at every time point. In this pseudo-population, treatment-confounder feedback is eliminated and the weighted outcome mean consistently estimates the counterfactual mean.

Definition 2.1 (IP Weights for Time-Varying Treatments). The unstabilized IP weight for treatment history $\bar{A}_K = \bar{a}_K$ is

$$W^A = \prod_{k=0}^K \frac{1}{f[A_k | \bar{A}_{k-1}, \bar{L}_k]},$$

and the **stabilized IP weight** is

$$SW^A = \prod_{k=0}^K \frac{f[A_k | \bar{A}_{k-1}]}{f[A_k | \bar{A}_{k-1}, \bar{L}_k]}.$$

Each factor in the product is the ratio of the marginal treatment probability (numerator) to the full conditional treatment probability (denominator). The denominator uses all available covariate information to assign weights that balance the treated and untreated over time. The numerator stabilizes the weights by using the marginal probability, which reduces variance without introducing additional bias under sequential exchangeability.

2.1 Marginal Structural Models

IP weighting for time-varying treatments is typically embedded in a **marginal structural model** (MSM) for the counterfactual mean:

$$E[Y^{\bar{a}_K}] = m(\bar{a}_K; \beta),$$

where m is a parametric model (e.g., logistic or linear) indexed by β . The MSM parameters are estimated by solving weighted estimating equations:

$$\sum_i SW_i^A \cdot \frac{\partial}{\partial \beta} m(\bar{A}_{K,i}; \beta) = 0.$$

Standard software for linear or logistic regression can be used simply by supplying SW^A as an observation weight.

Why stabilized weights:

Unstabilized weights W^A have expectation 1 by construction, but their variance can be very large when some covariate histories are rare. Stabilized weights SW^A also have expectation 1 but have substantially lower variance because they incorporate the marginal probability of the observed treatment history. Extreme weights — weights far from 1 — are a diagnostic warning sign: they indicate violations of positivity or model misspecification.

Weight truncation: A pragmatic approach to handle extreme weights is to truncate (or “trim”) them at a high percentile (e.g., 99th percentile). Truncation introduces a small bias but can substantially reduce variance. The analyst should report both truncated and untruncated results.

Model for the treatment mechanism: The denominator of SW^A is estimated by fitting a model for A_k given $(\bar{A}_{k-1}, \bar{L}_k)$ at each time k (e.g., logistic regression). The numerator is estimated by fitting a model for A_k given only \bar{A}_{k-1} .

2.2 Censoring Weights

When follow-up is incomplete due to loss to follow-up or other forms of censoring, a second set of weights is needed to correct for selection bias introduced by censoring.

Definition 2.2 (Stabilized Censoring IP Weights). Let C_k be an indicator of censoring at time k (1 = censored, 0 = not censored). The stabilized censoring IP weight is

$$SW^C = \prod_{k=0}^K \frac{\Pr[C_k = 0 \mid \bar{C}_{k-1} = 0, \bar{A}_{k-1}]}{\Pr[C_k = 0 \mid \bar{C}_{k-1} = 0, \bar{A}_{k-1}, \bar{L}_k]}.$$

The combined weight $SW^{A,C} = SW^A \cdot SW^C$ simultaneously adjusts for time-varying confounding and informative censoring. This combined weight is applied to the complete-case analysis (restricting to uncensored individuals, then weighting them to represent the full cohort).

3 21.3 A Doubly Robust Estimator for Time-Varying Treatments (p. 286)

The g-formula and IP weighting each depend on a single set of correctly specified models: the g-formula requires correct models for the outcome and each covariate; IP weighting requires correct models for the treatment (and censoring) mechanism at each time. **Doubly robust estimators** combine both approaches and are consistent if *at least one* of the two sets of models is correctly specified.

Definition 3.1 (Doubly Robust Estimator (Augmented IPW)). The augmented inverse probability weighted (AIPW) estimator for the counterfactual mean $E[Y^{\bar{a}_K}]$ takes the form

$$\hat{E}[Y^{\bar{a}_K}]_{\text{DR}} = \frac{1}{n} \sum_{i=1}^n \left[\hat{m}(\bar{A}_{K,i}) + \frac{\mathbf{1}(\bar{A}_{K,i} = \bar{a}_K)}{W_i^A} (Y_i - \hat{m}(\bar{A}_{K,i})) \right],$$

where $\hat{m}(\bar{A}_{K,i})$ is the g-formula prediction and W_i^A is the IP weight for individual i .

The double robustness property follows from the structure of this estimator:

- If the outcome model \hat{m} is correct but the weights W^A are misspecified, the weighted correction term has mean zero, so the estimator reduces to the g-formula (consistent).
- If the weight model is correct but the outcome model is misspecified, the weights rebalance the data so that the weighted residuals are mean zero, and the estimator is still consistent.
- If both models are correctly specified, the estimator is **efficient** (achieves the semiparametric efficiency bound).

Practical considerations:

In the time-varying setting, “doubly robust” means that the outcome model must correctly specify the joint conditional distribution $E[Y \mid \bar{A}_K, \bar{L}_K]$, and the treatment model must correctly specify $f[A_k \mid \bar{A}_{k-1}, \bar{L}_k]$ for every k . There are $K + 1$ treatment models and one outcome model; getting at least one *set* correct is less stringent than getting both right, but still demanding in practice.

Machine learning integration: The AIPW estimator is compatible with flexible machine learning methods for estimating the nuisance functions (outcome model and treatment model), provided cross-fitting is used to avoid overfitting bias. Targeted minimum loss-based estimation (TMLE) is a related approach that incorporates outcome model targeting for improved finite-sample performance.

4 21.4 G-Estimation for Time-Varying Treatments (p. 289)

G-estimation — the third g-method — focuses on directly modeling the **structural nested mean model** (SNMM), which describes how treatment at each time k modifies the counterfactual mean outcome, conditional on the history up to k .

Definition 4.1 (Structural Nested Mean Model (SNMM)). A structural nested mean model specifies the **blip-down function** $\gamma_k(\bar{a}_k, \bar{l}_k; \psi)$, which represents the effect on the outcome of setting treatment at time k from a_k to 0, given the past history $(\bar{a}_{k-1}, \bar{l}_k)$:

$$\mathbb{E}\left[Y^{\bar{a}_k, \bar{0}_{k+1}} - Y^{\bar{a}_{k-1}, \bar{0}_k} \mid \bar{A}_{k-1} = \bar{a}_{k-1}, \bar{L}_k = \bar{l}_k\right] = \gamma_k(\bar{a}_k, \bar{l}_k; \psi),$$

where $\bar{0}_j$ denotes no treatment from time j onward.

The blip function γ_k captures the **incremental causal effect** of treatment at time k on the final outcome, nesting within the effect of earlier treatment decisions. A common parametric specification is the **linear SNMM**:

$$\gamma_k(\bar{a}_k, \bar{l}_k; \psi) = \psi_1 a_k + \psi_2 a_k \cdot (\text{modifier}),$$

where the modifier might be a baseline or time-varying covariate.

4.1 G-Estimation Algorithm

The key insight behind g-estimation is that under sequential exchangeability, the **blipped-down outcome**

$$H_k(\psi) = Y - \sum_{j=k}^K \gamma_j(\bar{A}_j, \bar{L}_j; \psi)$$

(the outcome with the effect of treatment from time k onward “removed”) must be mean-independent of A_k conditional on $(\bar{A}_{k-1}, \bar{L}_k)$ when ψ equals the true parameter. G-estimation exploits this to estimate ψ by searching for the value that makes $A_k \perp\!\!\!\perp H_k(\psi) \mid \bar{A}_{k-1}, \bar{L}_k$ for all k .

Practically, this is often implemented via:

1. Fit a model for $A_k \mid \bar{A}_{k-1}, \bar{L}_k$ to obtain predicted probabilities \hat{p}_k .
2. Construct residuals $R_k = A_k - \hat{p}_k$ (the “de-confounded” treatment residual).
3. Estimate ψ by solving the estimating equation

$$\sum_i \sum_k R_{k,i} \cdot \frac{\partial}{\partial \psi} H_k(\psi_i) = 0.$$

Advantages of g-estimation:

- G-estimation only requires a model for the treatment mechanism, not the outcome distribution. This can be advantageous when the outcome model is hard to specify correctly.
- G-estimation of SNMMs produces effect estimates that are **doubly robust** in the sense that they are consistent if either the treatment model or a correctly specified “nuisance” component of the outcome model is correct.
- SNMMs allow for **effect modification**: the effect of treatment at time k can vary as a function of the individual’s history up to time k .

Limitations:

- SNMMs require specifying the structural model for the blip functions, which may be difficult or subject to misspecification.
- Unlike the g-formula, g-estimation does not directly produce predictions of the marginal distribution of Y^g ; it estimates incremental effect parameters that must be inverted to recover counterfactual means.
- The connection between SNMM parameters and policy-relevant estimands (e.g., the mean outcome under “always treat” vs. “never treat”) requires additional computation.

4.2 Rank-Preserving Structural Failure Time Models

A related family of models, **rank-preserving structural failure time models** (RP-SFTMs), extends g-estimation to survival outcomes. The RP-SFTM posits that treatment a_k accelerates or decelerates the failure time by a multiplicable factor $\exp(\psi a_k)$, and g-estimation proceeds by finding the ψ that makes the “counterfactual failure time under no treatment” independent of A_k given past history.

5 21.5 Censoring Is a Time-Varying Treatment (p. 297)

A conceptually important insight is that **informative censoring** can be analyzed using the same framework as time-varying treatments. Censoring at time k is a binary time-varying variable C_k with a similar recursive structure to treatment:

$$L_0 \rightarrow A_0 \rightarrow L_1 \rightarrow C_1 \rightarrow A_1 \rightarrow L_2 \rightarrow C_2 \rightarrow \dots \rightarrow Y.$$

When censoring depends on the time-varying covariate history \bar{L}_k (informative censoring), the observed sample of uncensored individuals is not representative of the full population. Restricting analysis to the uncensored introduces selection bias.

5.1 Handling Censoring via IP Weighting

The censoring IP weight SW^C (defined in Section 21.2) addresses this by upweighting uncensored individuals whose counterparts in the same covariate stratum were censored. An individual who is uncensored despite having covariate values that typically lead to censoring receives a high weight, making them represent themselves and those who were censored.

Under **sequential censoring exchangeability**:

$$C_k \perp\!\!\!\perp Y^{\bar{a}_k, \bar{c}=\bar{0}} \mid \bar{C}_{k-1} = \bar{0}, \bar{A}_{k-1}, \bar{L}_k,$$

the combined weight $SW^{A,C} = SW^A \cdot SW^C$ consistently estimates the counterfactual mean in the uncensored pseudo-population.

Why censoring is “just” a time-varying treatment:

Both treatment A_k and censoring C_k are time-varying binary variables whose assignment at each time depends on past history. Both can be confounded by time-varying covariates. The same sequential exchangeability and positivity conditions are needed for identification in both cases. The IP weight framework handles both simultaneously through a product of time-specific weights, one for each treatment and censoring event.

Competing events: When death is the competing event for a non-fatal outcome, it can similarly be viewed as a time-varying treatment (or censoring, depending on the estimand). See Fine Point 17.1 in Chapter 17.

5.2 G-Formula with Censoring

The g-formula for time-varying treatments also extends naturally to handle censoring. The product of covariate densities in the g-formula must include a model for $\Pr[C_k = 0 \mid \bar{C}_{k-1} = \bar{0}, \bar{A}_{k-1}, \bar{L}_k]$, and the Monte Carlo simulation sets $C_k = 0$ at every time point (i.e., we intervene to eliminate censoring). This gives the counterfactual mean under the joint intervention: “follow the treatment strategy g and do not censor anyone.”

6 21.6 The Big G-Formula (p. 300)

The most general form of the g-formula — sometimes called the **big g-formula** — encompasses all time-varying treatments, time-varying covariates, time-varying censoring, and time-varying outcomes within a single unified expression.

Definition 6.1 (The Big G-Formula). Under sequential exchangeability (for both treatment and censoring), positivity, and consistency, the counterfactual mean under joint strategy $(\bar{g}, \bar{c} = \bar{0})$ is

$$E[Y^{g, \bar{c}=\bar{0}}] = \sum_{\bar{l}_K} E[Y | \bar{A}_K = \bar{g}(\bar{l}_K), \bar{C}_K = \bar{0}, \bar{L}_K = \bar{l}_K] \prod_{k=0}^K f(l_k | \bar{A}_{k-1} = \bar{g}_{k-1}, \bar{C}_{k-1} = \bar{0}, \bar{L}_{k-1} = \bar{l}_{k-1}),$$

where the covariate density product is evaluated along the counterfactual treatment path under strategy g with no censoring.

The big g-formula is the workhorse of causal inference in longitudinal data: it provides a general recipe for computing any counterfactual mean, given correct models for the covariate and outcome processes.

6.1 Practical Implementation

In practice, the big g-formula is implemented via a **parametric g-computation algorithm**:

1. Fit outcome model: $E[Y | \bar{A}_K, \bar{C}_K = \bar{0}, \bar{L}_K]$.
2. For each k , fit covariate models: $f(L_k | \bar{A}_{k-1}, \bar{C}_{k-1} = \bar{0}, \bar{L}_{k-1})$.
3. For each k , fit censoring model: $\Pr[C_k = 0 | \bar{C}_{k-1} = \bar{0}, \bar{A}_{k-1}, \bar{L}_k]$.
4. Simulate a large pseudo-population under the intervention $(\bar{g}, C_k = 0)$: at each time k , set treatment to $g_k(\bar{l}_k)$ and set $C_k = 0$; draw L_k from its fitted conditional model.
5. Predict Y for each simulated individual from the outcome model.
6. Average the predictions.

Choice of g-method in practice:

Method	Primary model requirement	Double robustness	Handles dynamic strategies
G-formula	Outcome + covariate models	No	Yes
IP weight-ing (MSM)	Treatment + censoring models	No	With care
AIPW	Both (one can be wrong)	Yes	Yes
G-estimation (SNMM)	Treatment model + SNMM	Partially	Yes

In practice, researchers often apply two or three methods as a sensitivity analysis. Consistency of results across methods that rely on different model sets provides informal evidence that neither set is grossly misspecified.

7 Summary

- The **g-formula** for time-varying treatments standardizes over the counterfactual covariate distribution under the treatment strategy:

$$E[Y^{\bar{a}_K}] = \sum_{\bar{l}_K} E[Y \mid \bar{A}_K = \bar{a}_K, \bar{L}_K = \bar{l}_K] \prod_{k=0}^K f(l_k \mid \bar{a}_{k-1}, \bar{l}_{k-1}).$$

- **IP weights** $SW^A = \prod_{k=0}^K \frac{f[A_k \mid \bar{A}_{k-1}]}{f[A_k \mid \bar{A}_{k-1}, \bar{L}_k]}$ create a pseudo-population free of time-varying confounding, enabling unbiased estimation via weighted outcome regression (MSMs).
- **Censoring weights** SW^C handle informative censoring by the same logic, treating censoring as a time-varying treatment.
- **Doubly robust** (AIPW) estimators combine the g-formula and IP weighting, remaining consistent if at least one set of models is correctly specified.
- **G-estimation** of structural nested mean models (SNMMs) estimates treatment effect parameters by exploiting the conditional independence of blipped-down outcomes and treatment residuals.
- The **big g-formula** provides a unified expression for counterfactual means under joint interventions on treatment and censoring.

8 References

Hernán, Miguel A, and James M Robins. 2020. *Causal Inference: What If*. Chapman & Hall/CRC. <https://miguelhernan.org/whatifbook>.