L9: Comparing interventions using g computation

Jess Edwards,

Epidemiology, UNC Chapel Hill

EPID 722

Spring 2021

Flight plan

1. Intro to g-computation
    - Single timepoint
    - Survival analysis
2. Variations
3. Example

Background

The law of total probability

Consider a situation where B can take on 2 values, $B \in (1, 2)$ and we are interested in learning $P(A)$

$$P(A) = P(A|B = 1)P(B = 1) + P(A|B = 2)P(B = 2)$$

$$P(A) = \sum_{b} P(A|B = b)P(B = b)$$

What is our goal today

To estimate causal effects, or to provide our best guess at what will occur under various actions using data.

We will rely on the **g formula** to link potential outcomes to observed data distributions.

We will refer to a specific method we use to estimate causal effects as **g computation**.

Roadmap

Today, we will start by becoming familiar with the g formula for (potential) outcomes at a single time point,

e.g., to estimate $P(Y^a = 1)$

and learning how to do g computation in the single time point setting.

We will move on to rewrite the g formula for survival (potential) outcomes,

e.g., $P(T^a \leq t)$

and learn how to do g computation in the survival setting

Notation

W set of baseline characteristics, with possible values w

A exposure plan (treatment, policy, intervention), with possible values a

Y outcome of interest, with possible values y

n units in the study, indexed by i

Y^a is the *potential* outcome under plan a

Our parameter of interest (for the moment) is $P(Y^a = 1)$ or $E(Y^a)$.

But of course we do not observe Y^a, rather, our data set contains measures of Y, A, and W.[1]

[1] Note that we do not actually "observe" Y, A, or W either, we *measure* them, and our measurements are often error prone. See Edwards JK, Cole SR, Westreich D. All your data are always missing: incorporating bias due to measurement error into the potential outcomes framework. International journal of epidemiology. 2015 Aug 1;44(4):1452-9.

We will use **causal consistency** to link the potential outcomes to the observed data:

> If $A = a$ then $Y^a = Y$

But potential outcomes are still missing if $A \neq a$. But we can make guesses about Y^a when $A \neq a$ by assuming **exposure plan exchangeability**:

> $A \perp\!\!\!\perp Y^a$, which implies $E(Y^a|A) = E(Y^a)$

We often have reason to believe that A is associated with Y^a. In these settings, we can relax this assumption by assuming exposure plan exchangeability **conditional on** W:

> $A \perp\!\!\!\perp Y^a|W$, which implies $E(Y^a|A, W) = E(Y^a|W)$

The g formula

Scenario 1: A is randomized

$E(Y^a) =$

$E(Y^a) = E(Y^a|A = a)$ exchangeability

$E(Y^a) = E(Y|A = a)$ counterfactual consistency

Scenario 2: A is observed

$E(Y^a) =$

$E(Y^a) = \sum_w E(Y^a|W = w)P(W = w)$ Law of total probability

$E(Y^a) = \sum_w E(Y^a|W = w, A = a)P(W = w)$ exchangeability (+ positivity)

$E(Y^a) = \sum_w E(Y|W = w, A = a)P(W = w)$ counterfactual consistency

What are we hiding?

Throughout this course, we will assume **no interference**, or that Y_i^a does not depend on A_j for $i \neq j$, or, in other words, a participants potential outcome does not depend on another participant's exposure.

We will also asssume **treatment version irrelevance**, or that Y^a will be the same regardless of the version of treatment a received. In other words, if there are versions of treatment $V(a)$, then $Y^a = Y^{a,v(a)}$.

Finally, until week 11, will assume **no measurement error** of treatments, outcome, or covariates.

The "fundamental problem"

Recall that the fundamental problem of causal inference is that Y^a is missing when $A \neq a$. (Why?)

When we use IP-weighting, we upweight those with $A = a$ to have the same distribution of W as the entire study sample, just as we would ues IP-weighting to handle missing (factual) data.

When we use g-formula, we *impute* Y^a for those with $A \neq a$ after "training" our imputation model to predict $E(Y^a|W)$ in the subset of data where $A = a$.

"By hand" example

W	A	Y	n
0	0	0	20
0	0	1	5
0	1	0	10
0	1	1	30
1	0	0	20
1	0	1	20
1	1	0	5
1	1	1	10

W	A	Y	n	Y1	Y0
0	0	0	20	?	0
0	0	1	5	?	1
0	1	0	10	0	?
0	1	1	30	1	?
1	0	0	20	?	0
1	0	1	20	?	1
1	1	0	5	0	?
1	1	1	10	1	?

How do we fill in Y^1 when $A = 0$ and Y^0 when $A = 1$?

Substitute in $E(Y^1|A = 1, W = w)$ for Y^1 when $A = 0$ and $W = w$.

Example

```
eya1 <- dat %>%
  group_by(A,W) %>%
  summarize(py = sum(n*Y)/sum(n))

eya1
```

```
## # A tibble: 4 x 3
## # Groups:   A [2]
##       A     W    py
##   <dbl> <dbl> <dbl>
## 1     0     0 0.2
## 2     0     1 0.5
## 3     1     0 0.75
## 4     1     1 0.667
```

W	A	Y	n	Y1	Y0
0	0	0	20		0
0	0	1	5		1
0	1	0	10	0	
0	1	1	30	1	
1	0	0	20		0
1	0	1	20		1
1	1	0	5	0	
1	1	1	10	1	

But toy examples only get us so far.

We usually have >1 covariate in W, and covariates may be continuous. --> simple tabular approaches will soon run into the curse of dimensionality.

However, we can use parametric models to estimate $E(Y|A,W)$

Alternative approach to example

Fit model for

$$P(Y = 1|A, W) = \text{expit}(\beta_0 + \beta_1 A + \beta_2 W + \beta_3 AW)$$

```
mod <- glm(Y~A + W + A*W,
           weights = n,
           data = dat,
           family = "binomial"(link = "logit"))
summary(mod)$coefficients[,c(1:2)]
```

```
##               Estimate Std. Error
## (Intercept) -1.386294  0.4999566
## A            2.484907  0.6190954
## W            1.386294  0.5915713
## A:W         -1.791759  0.8850304
```

$$P(Y^1 = 1|W) = \text{expit}(\hat{\beta}_0 + \hat{\beta}_1(1) + \hat{\beta}_2 W + \hat{\beta}_3(1)W) = \text{expit}[-1.38 + 2.48(1) + 1.39W + (-1.79)(1)W]$$
$$P(Y^0 = 1|W) = \text{expit}(\hat{\beta}_0 + \hat{\beta}_1(0) + \hat{\beta}_2 W + \hat{\beta}_3(0)W) = \text{expit}[-1.38 + 2.48(0) + 1.39W + (-1.79)(0)W]`$$

We can also ask statistical software for model predictions

```
y1 = predict(mod, newdata = dat %>% mutate(A = 1), type = "response")
y0 = predict(mod, newdata = dat %>% mutate(A = 0), type = "response")
dat2 <- data.frame(dat, Y1 = y1, Y0 = y0)
```

W	A	Y	n	Y1	Y0
0	0	0	20	0.7500000	0.2
0	0	1	5	0.7500000	0.2
0	1	0	10	0.7500000	0.2
0	1	1	30	0.7500000	0.2
1	0	0	20	0.6666667	0.5
1	0	1	20	0.6666667	0.5
1	1	0	5	0.6666667	0.5
1	1	1	10	0.6666667	0.5

$RD = E(Y^1 - Y^0)$ = 0.36

Model specification

Model specification is important.

- Correct functional forms
- Interaction terms

What if we misspecify the model for $P(Y = 1|A, W)$?

```
mod_ms <- glm(Y ~ A + W,
              weights = n,
              data = dat,
              family = "binomial"(li
summary(mod_ms)$coefficients[,c(1
```

```
##                Estimate Std. Error
## (Intercept) -0.8969213  0.3877288
## A            1.7155621  0.4455658
## W            0.6717659  0.4401585
```

W	A	Y	n	Y1	Y0
0	0	0	20	0.6939478	0.2896836
0	0	1	5	0.6939478	0.2896836
0	1	0	10	0.6939478	0.2896836
0	1	1	30	0.6939478	0.2896836
1	0	0	20	0.8161393	0.4439478
1	0	1	20	0.8161393	0.4439478
1	1	0	5	0.8161393	0.4439478
1	1	1	10	0.8161393	0.4439478

$RD = E(Y^1 - Y^0)$ = 0.39

An Algorithm

1. Specify parameter of interest (e.g., $RD = E(Y^1 - Y^0)$)
2. Identify all relevant confounders W
3. Fit regression model for $E(Y|A, W)$, save regression coefficients $\hat{\beta}$
    - Check model fit by predicting $E_{AW}(\hat{E}(Y|A, W))$ using $\hat{\beta}$ and the observed distribution of A and W. Do you recover $E(Y)$?
4. To estimate $E(Y^1)$, set $a = 1$ and predict $E(Y|a = 1, W)$ using $\hat{\beta}$ and the observed distribution of W
5. To estimate $E(Y^0)$, set $a = 0$ and predict $E(Y|a = 0, W)$ using $\hat{\beta}$ and the observed distribution of W

Note that we need not fit a parametric model for $P(W = w)$. Instead, the empirical distribution of the observed participants provides a nonparametric estimate of $P(W)$, regardless of the dimension of W.

G formula for survival analysis

Theory

So far we have focused on estimating $P(Y^a = 1)$.

But in a survival analysis framework we may instead be interested in estimating $P(T^a \leq t)$

$P(T^a \leq t) = \sum_w P(T^a \leq t | W = w) P(W = w)$ Law of total probability

$P(T^a \leq t) = \sum_w P(T^a \leq t | W = w, A = a) P(W = w)$ exchangeability (+ positivity)

$P(T^a \leq t) = \sum_w P(T \leq t | W = w, A = a) P(W = w)$ counterfactual consistency

Method 1: Discrete time

As in the single time point analysis, we first need a parametric model for $P(T \leq t|W = w, A = a)$.

To simplify the modeling approach, we can coarsen (or discretize) time. For example, if we coarsen time into J intervals (e.g., weeks), we can rewrite $P(T \leq t|W = w, A = a)$ as

$$P(T \leq t|W = w, A = a) = \prod_{j=1}^{J}[1 - P(Y_j = 1|Y_{j-1} = 0, W = w, A = a)]$$

We then fit a parametric model in the person-week data set for

$$P(Y_j = 1|Y_{j-1} = 0, W = w, A = a) = \text{expit}[\beta_0 + \beta_1 g(j) + \beta_2 W + \beta_3 A + \beta_4 AW]$$

We use estimates of $\hat{\beta}$ to predict the probability of the outcome (we will call this predicted probability $\eta_{ij}(a)$) for each participant at each time point j under each exposure plan (here $a = 1$ and $a = 0$), e.g.,

$$\eta_{ij}(1) = \text{expit}[\hat{\beta}_0 + \hat{\beta}_1 g(j) + \hat{\beta}_2 W_i + \hat{\beta}_3(1) + \hat{\beta}_4(1) W_i]$$

Note that we predict $\hat{\eta}_{ij}(1)$ and $\hat{\eta}_{ij}(0)$ **only** for time points j in which the participant had not yet had the outcome by time $j - 1$.

Note that we **do** predict $\hat{\eta}_{ij}(1)$ and $\hat{\eta}_{ij}(0)$ for time points that occur after a participant was censored in the observed data.

With $\hat{\eta}_{ij}(1)$ and $\hat{\eta}_{ij}(0)$ in hand, we estimate risk under each plan as

$$P(T^1 \leq t) = \prod_{j=1}^{J} [1 - \hat{\eta}_{ij}(1)]$$

$$P(T^0 \leq t) = \prod_{j=1}^{J} [1 - \hat{\eta}_{ij}(0)]$$

Method 2: Breslow estimator

Rather than estimate $P(Y_j = 1|Y_{j-1} = 0, W = w, A = a)$ using pooled logistic regression (which requires discretizing time), we could estimate $P(T \leq t|A = a, W)$ using the Breslow estimator.

We will not go into depth on the details of the Breslow estimator here, but if you want to learn more, the paper by Lin is straightforward and helpful

> Lin DY. On the Breslow estimator. Lifetime data analysis. 2007 Dec 1;13(4):4

Overview

Basic idea is to use the Breslow estimator to estimate the hazard function that would have been observed had the entire study sample received intervention a, and then use this hazard function to compute risk

- recall, $F(t) = 1 - \exp(-H(t))$

Breslow estimator algorithm

Step 1: fit a Cox model among those with $A = a$, conditional on covariates W, save coefficients $\hat{\alpha}$

Step 2: count up the number of events and number at risk at each event time in group with $A = a$

Step 3: among those with $A = a$, compute baseline hazard function at each event time using the Breslow estimator

$$h_0(k|A = a) = \frac{d_k}{\sum_{i=1}^{n} R_{ik} \exp(\hat{\alpha} W)}$$

where d_k is the number of events at event time k and R_{ik} is an indicator that person i is in the risk set at event time k.

Step 4: compute the hazard function that would have been seen had the full study sample received exposure a by multiplying $h_0(k|A=a)$ by the linear predictor from cox model for each person in the FULL sample, (not limited to those with $A=a$)

$$h^a(k) = h_0(k|A=a) \times \exp(\hat{\alpha} W_i)$$

Step 5: Estimate $\mu_i(t,a) = P(T_i^a \leq t | W_i; \hat{\alpha})$ as $\mu_i(t,a) = 1 - \exp(-H^a(t))$, where $H^a(t) = \int_0^t h^a(t) dt$.

Step 6: Average across units to estimate risk.

$$F(t) = n^{-1} \sum_{i=1}^{n} \mu_i(t,a)$$

Variations

Changing the referent group

Changing the referent is much easier using g computation than when using standard regression. Why?

Standard regression

$$\log(P(Y=1)) = \beta_0 + \beta_1 A + \beta_2 W$$

$RR = \exp(\hat{\beta}_1)$ <- compares $A = 1$ to $A = 0$

g Computation

$$\text{logit}(P(Y=1)) = \beta_0 + \beta_1 A + \beta_2 W$$

$P(Y^1 = 1) = \text{expit}(\hat{\beta}_0 + \hat{\beta}_1(1) + \hat{\beta}_2 W)$ <- risk under exposure

$P(Y^0 = 1) = \text{expit}(\hat{\beta}_0 + \hat{\beta}_1(0) + \hat{\beta}_2 W)$ <- risk under no exposure

$P(Y = 1) = \text{expit}(\hat{\beta}_0 + \hat{\beta}_1(A) + \hat{\beta}_2 W)$ <- risk under the "natural course"

More interesting interventions

Estimating the effects of more interesting interventions is much easier using g computation than when using standard regression. Why?

Say A is a continuous exposure and we want to estimate the effect of limiting exposure to A to 5 units. Let's call this intervention lim.

g Computation

$$\text{logit}(P(Y=1)) = \beta_0 + \beta_1 g(A) + \beta_2 W$$

Now create new exposure under intervention:

if $A > 5$ then $A^{lim} = 5$ <-- everyone exposed to more than 5 units now gets 5 units

if $A \leq 5$ then set $A^{lim} = A$ <-- those exposed to fewer than 5 units receive their original exposure

$P(Y^{lim} = 1) = \text{expit}(\hat{\beta}_0 + \hat{\beta}_1 g(A^{int}) + \hat{\beta}_2 W)$ <- risk under intervention lim

Targeted interventions

Say we are interested in an intervention to treat everyone with covariate $W = 1$. Specifically, we want to know the difference in risk between treating only those with $W = 1$ and treating the entire study population. Let's call this intervention tar.

g Computation

$$\text{logit}(P(Y = 1)) = \beta_0 + \beta_1 A + \beta_2 W$$

Create new exposure variable A^{tar} with values $A^{tar} = 1$ if $W = 1$ and $A^{tar} = 0$ if $W = 0$.

$P(Y^{tar} = 1) = \text{expit}(\hat{\beta}_0 + \hat{\beta}_1(A^{tar}) + \hat{\beta}_2 W)$ <- risk under intervention tar

$P(Y^{1} = 1) = \text{expit}(\hat{\beta}_0 + \hat{\beta}_1(1) + \hat{\beta}_2 W)$ <- risk under full exposure

$P(Y = 1) = \text{expit}(\hat{\beta}_0 + \hat{\beta}_1(A) + \hat{\beta}_2 W)$ <- risk under the "natural course"

Example

Context

Say we have a study of 1000 patients receiving routine physical exams at UNC on October 1, 2021.

Your goal is to follow them all for 1 year to assess flu incidence. We will use various techniques to estimate flu incidence under various strategies that involve providing flu vaccines on the date of the routine physical exam.

All patients have agreed to contact the study team if they get the flu during the risk period.

Assume no patients had received a flu shot prior to their exam date and that patient outcomes are independent.

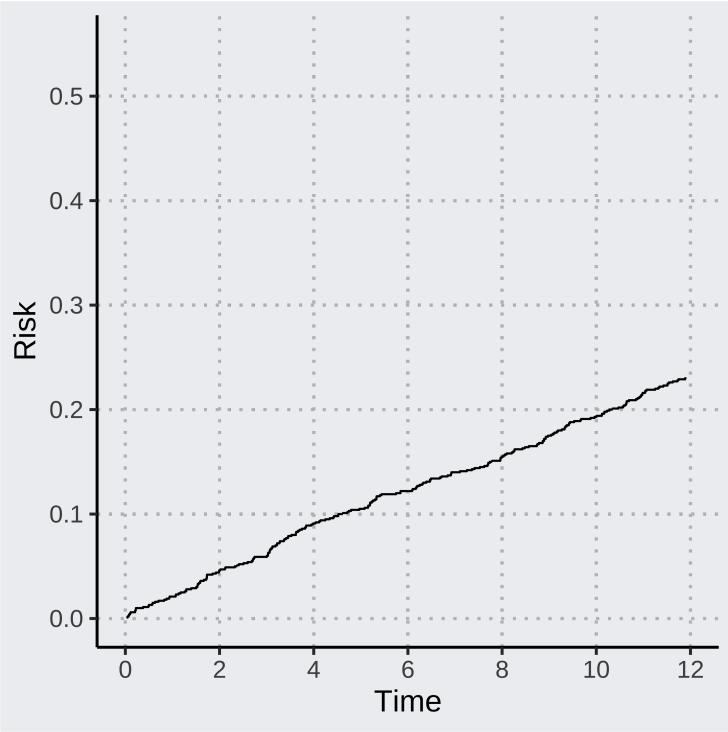
Natural course

Standard Kaplan-Meier

$$F(t) = 1 - \prod_{k \in R_k \leq t} (1 - \frac{d_k}{n_k})$$

```
nckm <- survfit(Surv(t, delta)
                ~ 1,
                data = mydat)
```

```
proc phreg data = mydat noprint;
    model t*delta(0) = ;
    baseline out = nckm
    survival = s;
```



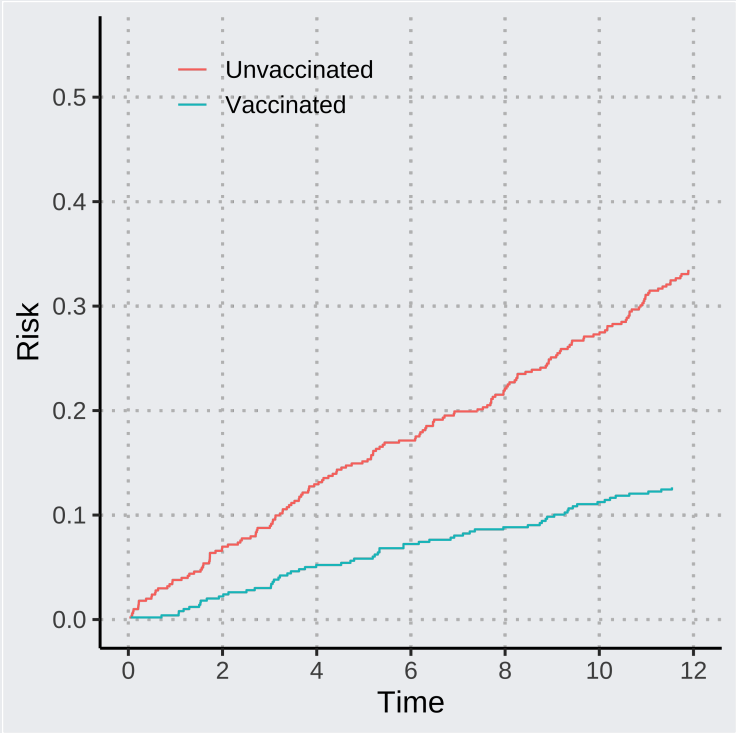
Crude

Stratified KM

$$F(t|A=a) = 1 - \prod_{k \in R_k \leq t, A=a} (1 - \frac{d_k}{n_k})$$

```
crudekm <- survfit(Surv(t, delta)
                   ~ a,
                   data = mydat)
```

```
proc phreg data = mydat noprint;
  strata a;
    model t*delta(0) = ;
    baseline out = nckm
    survival = s;
run;
```

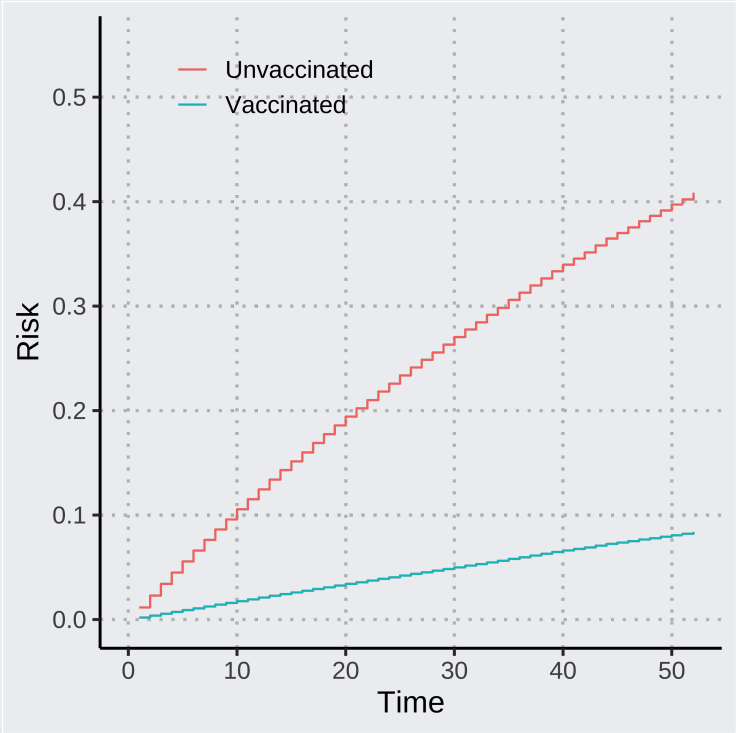


g computation, pooled logistic

1. Expand dataset to have 1 record per person-week
2. Fit pooled logistic regression model to person-week dataset

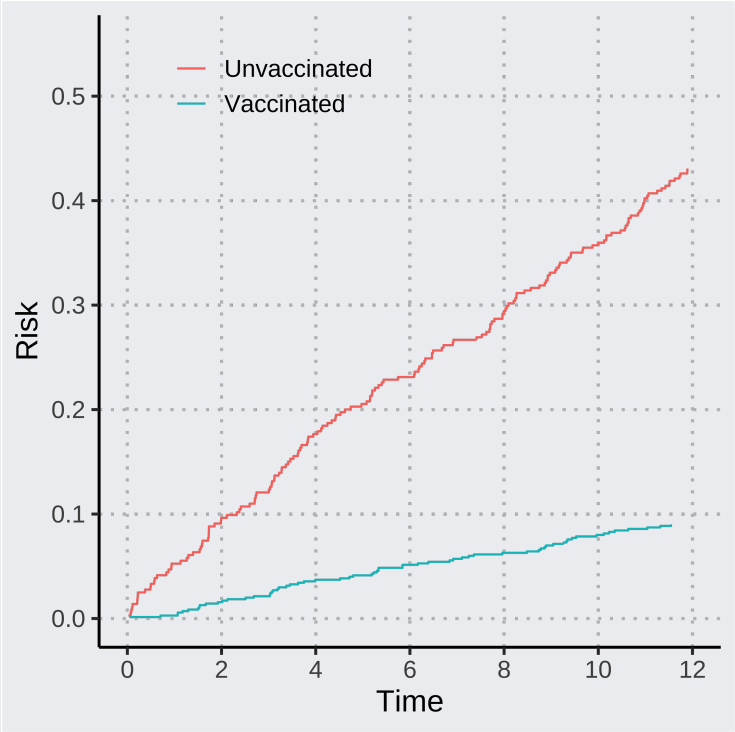
$$P(Y = 1|A, W, j) = \text{expit}(\beta_0 + \beta_1 A + \beta_2 W + \beta_3 AZ + \beta_4 g(j))$$

1. Use estimates from model above to predict $P(Y = 1|a, W, j) = \eta(j, a)$, where j indexes weeks
2. Compute $P(T^a \leq j|W_i)$ for each person $P(T^a \leq t|W_i) = 1 - \prod_{j=1}^{t}(1 - \eta(j, a))$



g computation, Breslow

1. Return to person-level dataset
2. Fit Breslow estimator to compute $\mu_i(t, a) = P(T^a \leq t | W_i)$ for each person at each event time t
3. Summarize across people



APPENDIX

More interesting interventions (part 2)

Estimating the effects of more interesting interventions is much easier using g computation than when using standard regression. Why?

Say we want to estimate the effect of increasing the proportion treated with drug a from 50% to 75%. Let's call this increase intervention g.

g Computation

$$\text{logit}(P(Y=1)) = \beta_0 + \beta_1 A + \beta_2 W$$

Now create new treatment under intervention:

if $A = 1$ then $A^g = 1$ <-- everyone already exposed stays exposed

if $A = 0$ then set A^g to 1 with probability 0.5 <-- half of those unexposed become exposed under intervention

$P(Y^g = 1) = \text{expit}(\hat{\beta}_0 + \hat{\beta}_1(A^g) + \hat{\beta}_2 W)$ <- risk under intervention g