

# Introduction to g-methods NorPEN webinar

Jessica Young

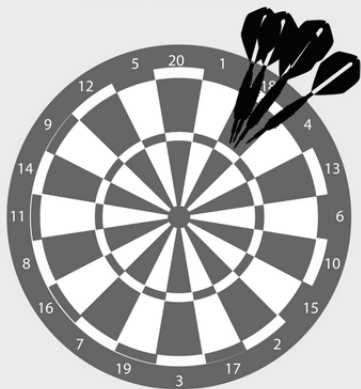
November 11, 2020

# What are “g-methods”?

- G-methods are nothing more than examples of *statistics* – computational algorithms applied to a set of measurements (data)
  - ▶ Other examples: sample average, sample variance, sample average difference comparing two groups
- We care about certain properties of any statistic: the dart board visual aid
  - ▶ want the darts to scatter equally around the center (unbiased)
  - ▶ want tight scatter around the center (low variance)

Can think of any dart as a statistic computed from our data in one sample

**High Bias**  
Low Variance



**High Variance**  
Low Bias



# Center of the dart board

The center represents the quantity we want the statistic (the data) to tell us something about: the question we want to answer

- **However** tools for constructing statistics limited to questions about population features we can, at least in principle, observe (or measure)
- Call these “statistical parameters”

Examples of “statistical parameters”:

- Among population diagnosed with high blood pressure, difference in mean blood pressure one year post-diagnosis among those who initiated and adhered to daily treatment upon diagnosis **versus** among those who never took that treatment
- Overall mean blood pressure

**Problem:** many research questions are about causal effects which are not statistical parameters.

# What is a causal effect?

Contrast (e.g. difference) of outcomes in the **same individuals** but under implementation of **different treatment assignment rules (interventions)**.

- Ex: Mean blood pressure one year later had, since baseline, everyone taken treatment daily **versus** had instead they never taken the treatment

Quantifies *counterfactual (potential)* features of a population.

- For any individual, blood pressure can only be observed under one of these treatment scenarios (and may not be observed under either scenario)

Rather than statistical parameters, causal effects are counterfactual parameters.

# Causal inference versus Statistical inference

When interest in a causal effect, extra steps needed to choose a statistic (approach to analyzing our data)

- 1 Articulate the causal effect we want
- 2 Consider subject matter assumptions that let us equate this effect to some statistical parameter (function of only measured characteristics)
  - ▶ *identifying assumptions* – ex: “no unmeasured confounding”
- 3 Finally we can choose a statistic for that statistical parameter and understand its “dart board” properties. Different statistics depend on additional assumptions
  - ▶ *statistical assumptions* – ex: blood pressure is normally distributed given treatment

Step 3 constitutes *statistical inference*. Adding steps 1 and 2 constitutes *causal inference*.

# The $g$ -formula: a foundational statistical parameter in causal inference

In 1986, Jamie Robins showed that, under a general set of identifying assumptions, the population mean of an outcome had we implemented some treatment assignment rule equals the  *$g$ -computation algorithm formula* indexed by that rule – a function of measured features of the population (hopefully features you have measured in your study)

- Thus, a contrast (e.g. difference) in this function indexed by different rules equals the causal effect

Often shortened to the  *$g$ -formula*.

# The g-formula: a foundational statistical parameter in causal inference

Why was this a big deal?

- The treatment rules (**the causal questions**) that this result applies to are extremely general and include, possibly complex, treatment rules that vary over time.
- The **identifying assumptions** that give this result are also extremely general in that they allow complex longitudinal data structures such that
  - ① characteristics associated with future outcomes that affect treatment decisions may change over time
  - ② these characteristics may also, themselves, be affected by treatment at an earlier time
- “time-varying confounding affected by past treatment”



# Back to g-methods

Generally, the g-formula is a complex, high-dimensional function of variables measured in the study.

- Different g-methods compute different statistics for the g-formula indexed by a treatment rule (or contrasts indexed by different rules).
- When targeting the same g-formula, they differ only by statistical assumptions – (“the dart board properties”).

The now enormous literature on g-methods has come out of the fact that there are many types of rules we can consider and many types of statistical assumptions we can make to estimate the g-formula indexed by one of those rules.

# Examples of treatment assignment rules: static rules

Consider population at time diagnosed with high blood pressure (“time 0” or baseline)

- Take any blood pressure medication every day for the next year
- Never take medication for the next year

These are examples of *static* rules: treatment at each follow-up is completely determined without reference to individual patient characteristics

# Examples of treatment assignment rules: dynamic rules

Consider same population and the following rule applied on each day  $t$  of one year follow-up

- Take blood pressure medication on day  $t$  if no contraindication has developed by  $t$ , otherwise do not take treatment.

Or consider a healthy population and the rule on each day  $t$  of one year follow-up

- Initiate blood pressure medication on day  $t$  if systolic blood pressure has risen above cutoff  $x$  within the last week, otherwise do not initiate on day  $t$ . We might consider a number of cutoffs  $x$ .

These are examples of *dynamic* rules: treatment at each follow-up is determined by time-evolving individual patient characteristics.

# Counterfactual notation for defining causal effects

Suppose interest is in the causal effect of following one versus another time-varying treatment rule on a later outcome (say blood pressure at one year follow-up) in the study population. We can use a notation to write this effect (counterfactual parameter) as

$$E[Y^{g_1}] - E[Y^{g_2}] = E[Y^{g_1} - Y^{g_2}]$$

where  $g$  denotes a rule of interest  $Y^g$  is the outcome for an individual in the population had, possibly contrary to fact (counterfactual outcome under  $g$ ) and  $E$  denotes expectation (study population mean).

# Idealized study

Imagine we were able to conduct a randomized trial where individuals meeting eligibility for the study population are enrolled and then randomized to say one of two study arms where the protocol is to follow a rule of interest ( $g_1$  or  $g_2$ ). Also assume

- 1 Everyone adheres to the protocol for the next year
- 2 The outcome is measured at the end of that year for all individuals originally randomized.

# Randomization and identification

Let  $Z$  represent the result of a coin toss (if you get heads you are assigned arm 1, if you get tails arm 2). Because the value of  $Z$  is determined only by the coin flip, we have that

$$Y^g \perp\!\!\!\perp Z$$

where  $\perp\!\!\!\perp$  denotes independence for either  $g = g_1$  or  $g = g_2$ . In other words, the value of  $Z$  an individual gets is not associated with their counterfactual outcomes.

- This independence assumption is an identifying assumption which is guaranteed by the definition of  $Z$ .

# Idealized study

Design, which guarantees this independence assumption, coupled with perfect adherence and complete outcome measurement, guarantees counterfactual parameter  $E[Y^{g_1}] - E[Y^{g_2}]$  equals the statistical parameter

$$E[Y|Z = 1] - E[Y|Z = 0]$$

Alternatively we say the effect we want is *identified* by this statistical parameter. Many available statistics for the conditional mean difference (last step).

# Real world study

Real world studies are rarely “ideal” particularly for causal effects of time-varying treatments.

- People don't follow the protocol
- People drop out of the study
- A trial may not be feasible or timely, we may have data such that nothing is randomized (investigator has no control at any time in determining what treatment people might take, there is no coin flipping, no  $Z$ ) – observational study



# Real world study

We still may have the same causal question! We just have a less ideal study to answer it. That is

- To equate our causal effect to a statistical parameter, we'll have to make assumptions that are not guaranteed.
- Further the statistical parameter may be more complex than a simple comparison of outcome means.

# Observational study

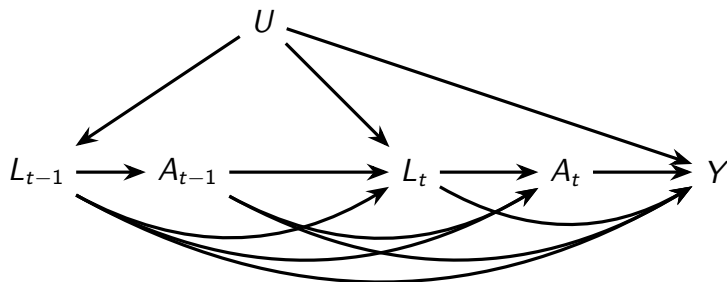
Suppose we had an observational study where data was collected (say through EHR) on  $n$  individuals meeting criteria for the study population of interest at time 0. Notation for what we measure.

- Outcome of interest (e.g. blood pressure measured at end of year):  $Y$
- Treatment status on each day of the year follow-up:  $A_t$ ,  $t = 0, \dots, 365$
- Covariates on each day  $t$ :  $L_t$  (lab measures, new diagnoses, patient visit)

Use overbar to denote the whole history of a covariate

- $\bar{A}_K = (A_0, A_1, \dots, A_K)$  and  $\bar{L}_K = (L_0, L_1, \dots, L_K)$
- Allow that  $L_0$  contains not just day 0 measurements of t-v covariates but also time fixed characteristics (e.g. biological sex)

# Causal Directed Acyclic Graph (DAG)



Allows that measured time-varying covariates are affected by past treatment and also that there are unmeasured common causes of these measured covariates and the outcome.

# Task

Now consider assumptions that let us link our causal effect of interest  $E[Y^{g_1}] - E[Y^{g_2}]$  to a statistical parameter, that is some function of the measured variables:

$$O = (\bar{L}_K, \bar{A}_K, Y)$$

# Exchangeability

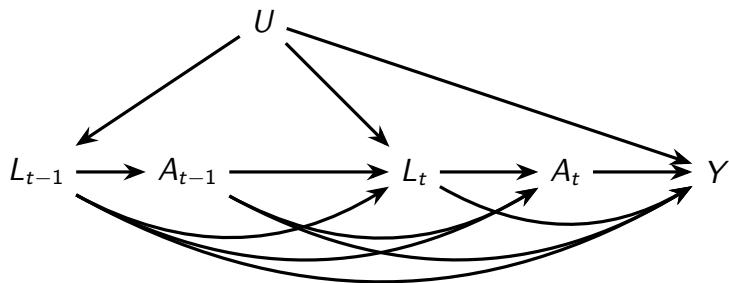
Consider an assumption very similar to the counterfactual independence that was guaranteed in the ideal trial but modified:

$$Y^g \perp\!\!\!\perp A_t \mid \bar{L}_t, \bar{A}_{t-1} = \bar{a}_{t-1}^g$$

This states that the counterfactual outcome associated under rule  $g$  is independent of the treatment actually received on a given day  $t$  within levels of measured covariate history.

- Many names for this assumption including *exchangeability* and *no unmeasured confounding*
- Can refer to the covariate history  $\bar{L}_t$  as the *measured confounder history*
- This independence guaranteed in a *sequentially randomized trial* where  $A_t$  assigned by weighted coin, at most dependent on measured past (not guaranteed in observational study)

# Causal graphs and exchangeability



Exchangeability can be evaluated with respect to assumptions encoded in the causal DAG if it is assumed to represent an underlying counterfactual causal model (NPSEM).

# Consistency

**Consistency:** If individual has treatment history consistent with  $g$  then future covariates and outcome are the values they would take under  $g$

- If  $\bar{A}_t = \bar{a}_t^g$  then  $\bar{L}_{t+1}^g = \bar{L}_{t+1}$  and  $Y^g = Y$ .
- Requires that there are not “multiple versions of treatment” and “no interference” (other people’s treatments do not affect my counterfactual outcomes) – Stable Unit Treatment Value Assumption (SUTVA)

# Positivity

**Positivity:** For any measured covariate and treatment history plausible in the observational study and consistent with  $g$  prior to time  $t$ , it must be possible to observe a value of treatment consistent with  $g$  at time  $t$ , for all times  $t$ .

- Depends on choice of  $g$ : if  $g$  is “always treat” and past measured history includes contraindication for treatment, will be violated. Could rectify this by changing to a dynamic rule that accommodates this reality.



# The g-formula

Robins (1986,1987) proved that, given these assumptions, can identify  $E[Y^g]$  by:

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

This is the g-formula indexed by rule  $g$ . Depends on  $g$  because each term restricted to treatment histories that are consistent with the rule  $g$ . Contrasts indexed by different choices of rule equal causal effects.

# Examples

The g-formula indexed by the rule “Always treat” is

$$\sum_{\bar{l}_K} E[Y | \bar{A}_K = \bar{1}_K, \bar{L}_K = \bar{l}_K] \prod_{t=0}^K f(l_t | \bar{1}_{t-1}, \bar{l}_{t-1})$$

When indexed by a dynamic rule,  $a_t^g$  will depend on which confounder history component of the sum we're in.

$$\sum_{\bar{l}_K} E[Y | \bar{A}_K = \bar{a}_{K-1}^g, \bar{L}_K = \bar{l}_K] \prod_{t=0}^K f(l_t | \bar{a}_{t-1}^g, \bar{l}_t)$$

# g-methods

g-methods are different statistical methods for estimating contrasts in the g-formula indexed by different choices of treatment rule  $g$ .

- Now we're at step 3, statistical inference.

Several methods are available that vary in the strength and types of statistical assumptions that they rely on. This has implications for bias, variance and also computational complexity.

# Causal Inference Methods for Patient Centered Outcomes Research (PCORI)

- <http://cimpod.org>
- PCORI funded dissemination project – materials available from two in person conferences covering range of topics in causal inference as well as a series of online workshops.

# 2021 Workshop

- More on causal diagrams
- More detail on the various estimation procedures and variations
- If time, introduction to extensions: data with loss to follow-up, questions about stochastic treatment rules, outcomes subject to competing or truncation events.

Some extra slides with high-level discussion of different g-methods

## parametric g-formula/g-computation

Assume that the terms of g-formula are correctly characterized by parsimonious parametric models:

- A model for  $E[Y|\bar{A}_K, \bar{L}_K]$ . (outcome regression model)
- Models for joint distribution of  $L_t$  given past treatment and confounders for all  $t$  (can be a lot of models).
- Fit the models, and approximate the sum over all levels of confounders by simulating “lots” of treatment and confounder histories consistent with  $g$  using the model parameters.
  - ▶ at each  $t$  simulate confounders from models then set treatment according to  $g$ .
- Use the outcome regression to estimate the outcome mean conditional on each simulated history and then average these estimates.

## parametric g-formula/g-computation

- Advantages: computationally very easy to adapt to any rule, the only step that changes in how you set treatment in simulation. Also relies on familiar parametric modeling approaches so not a huge learning curve. IF the statistical assumptions it requires are correct, the method has low variability (tight scatter around the center of the dart board).
- Disadvantage: these statistical assumptions are generally extremely strong. When they are wrong, you may get tight scatter far from the center.



# Inverse probability weighting

$g$ -formula indexed by rule  $g$  has a weighted representation, can alternatively write it as the weighted outcome mean:

$$E \left( Y \frac{\prod_{t=0}^K I(A_t = a_t^g)}{\prod_{t=0}^K f(A_t | \bar{L}_t, \bar{A}_{t-1})} \right)$$

Weight is 0 for anyone with treatment inconsistent with  $g$  at any follow-up time. Otherwise, it is inverse of the probability of receiving the treatment that person received given their measured past. This representation motivates “inverse probability weighted” estimators. Relies on correctly specified model for weight denominator.

# IP weighting

- Advantage: still computationally straightforward, familiar statistics just with weights. Statistical assumptions are arguably weaker than those required for parametric g-formula (only need to get “treatment model” correct)
- Disadvantage: tends to be highly variable (wide scatter). There are ways to adapt the method to increase precision (stabilized weights). Sometimes people add additional assumption of a marginal structural model (the g-formula itself is a smooth function of  $g$ ).

Many implementations in pharmacoepidemiology that implement “artificial censoring” when data deviates from  $g$  with inverse probability of censoring weighting to correct for this artificial “selection”. This precisely targets the weighted representation of the g-formula that I just showed.

# Doubly-robust/ “state of the art” estimators

Another representation of the g-formula as a series of iterated conditional expectations (means) motivate doubly robust methods – weaken statistical assumptions in that only requires correct specification of one of two sets of parametric models.

- Models for weight denominator
- Models for iterated conditional means

Unlike the “singly robust” methods above, adaptations of these approaches can further weaken statistical assumptions so that all we need are “convergence rate” assumptions on the method of estimating these quantities so that flexible machine learning methods can be used in place of parametric models entirely.

# DR/State of the art methods

Advantages: These state of the art approaches are appealing in that they rely on the weakest statistical assumptions. Disadvantages: They can be computationally intensive which can be a practical limitation. They also add a level of statistical complexity that can create a bigger hurdle for applied audiences – this is changing.