Comparative effectiveness of dynamic treatment regimes

An application of the parametric g-formula

Miguel Hernán
Departments of Epidemiology and Biostatistics
Harvard School of Public Health
www.hsph.harvard.edu/causal
What is comparative effectiveness research (CER)?

- “The generation and synthesis of evidence that compares the benefits and harms of alternative methods to prevent, diagnose, treat, and monitor a clinical condition or to improve the delivery of care”
  - IOM 2009

- Comparative effectiveness or comparative safety
What is the purpose of CER?

- “To assist consumers, clinicians, purchasers, and policy makers to make informed decisions that will improve health care at both the individual and population levels”
  - IOM 2009

- To help decision-makers
CER is about causal inference

- “The generation and synthesis of evidence that compares the benefits and harms…”
- “to make informed decisions that will improve health…”

☐ In other words, CER is about comparing the effects of well-defined interventions or strategies to answer policy, public health, and clinical questions.
3 requirements for CER

1. A relevant, well-defined question
   - A key prerequisite

2. High-quality data
   - Appropriate for the question above

3. Valid statistical methods
   - To be applied to the high-quality data to answer the well-defined question

8 June 2012 G-formula for CER
CER questions are becoming increasingly complex

- En epidemiology, clinical medicine, comparative effectiveness research...
- Often not about the average treatment effect of a non-time-varying treatment
- Often involving dynamic strategies or regimes
  - i.e., treatment strategies that depend on the individual’s evolving covariates
Example: antiretroviral therapy (ART) for HIV-infected patients

- RCTs have been successful in showing that assignment to combined ART is best
- However, key questions still unanswered
  - what to start
  - when to start
  - when to switch
  - when to monitor
- Dozens of relevant strategies with potentially very wide variation in effectiveness
Examples of complex questions involving dynamic regimes

☐ When to start treatment?
  ■ What would be the 5-year mortality risk when cART is initiated within 6 months of either an AIDS diagnosis or CD4 cell count first dropping below 500 cells/mm$^3$?
  ■ Below 350?
  ■ Within 3 months?
  ■ Within 6 months of CD4 first dropping below 500 and viral load above 1000?
Examples of complex questions involving dynamic regimes

- When to change treatment?
  - What would be the 5-year mortality risk when the cART regime is switched within 6 months of immunologic failure?
  - What if we wait until viral load is greater than 1000? Until CD4<350?

- When to monitor treated patients?
  - Should we measure viral load every 3, 6, 12 months? Before/After treatment?
CER questions are complex because clinical decisions are

- Classical RCTs, however, provide **simple** information
  - Typically an intention-to-treat analysis will say that assignment to A is better than assignment to B
- That’s fine for one-time treatments
  - e.g., surgery, a one-dose vaccination, CABG
- But what about long-term clinical strategies?
  - e.g., therapeutics for chronic diseases with no cure (diabetes, HIV disease, chronic kidney disease, cardiovascular disease... )
One option: RCTs

- Design RCTs that compare the effectiveness of explicitly defined clinical strategies on clinically relevant outcomes over a clinically relevant (i.e., long) period

- An example of strategy would be:
  - start treatment A within 3 months of the CD4 cell count first dropping below 500 or of an AIDS diagnosis, whichever occurs first; switch to treatment B within 3 months of the viral load going over 1000 or immediately if toxicity occurs, unless the patient is pregnant or has a liver disorder, in which case switch to treatment C; if C fails then...
CER and personalized medicine

- Complex clinical strategies need to be compared if we get serious about comparative effectiveness research for individualized treatment decisions.

- But once we get this serious, RCTs suddenly look less attractive...
Less attractive RCTs

- Adherence to complex strategies over a long period may be low
  - intention-to-treat analysis → a black box if common deviations from protocol
  - as treated analyses: time-varying confounding like in observational studies

- Long follow-up → drop-out
  - an intention-to-treat analysis not even possible
  - selection bias like in observational studies

- For a discussion on RCTs for CER, see Hernán and Hernández-Díaz. *Clinical Trials* 2012.
The next best option

- Use observational data to emulate an RCT

- Need prospective collection of data on time-varying data on treatments, outcomes, and confounders

- Not all data sources include this information
  - Therefore not all data sources can be used
Enter observational registries and electronic medical records

- These data sources are our best chance to compare effectiveness of clinical strategies in a timely way in unselected populations
- They include prospective data on time-varying treatments, outcomes, and confounders
  - RCTs do not typically collect this information
- They can be used to emulate trials without baseline randomization
Observational data for CER are becoming increasingly complex

- Increasingly high-dimensional
- Longitudinal studies with time-varying treatments and confounders
- Measurement of treatment and confounders may occur at subject-specific (random) times
Examples of complex longitudinal data

- Large populations followed for long periods with frequent measurement of treatments, outcomes, and confounders
  - At prespecified intervals: e.g., the Nurses’ Health Study, MACS/WIHS
  - At random intervals: e.g., clinical cohorts like the HIV-CAUSAL Collaboration, USRDS claims, electronic medical records
Many methods originally developed for less complex questions

- Conventional regression
  - Bias if confounders affected by previous treatment

- Propensity score matching
  - For time-varying treatments?

- Instrumental variable estimation
  - Who are the compliers when treatment is given every month? And the never takers?

- Principal stratification
  - How are principal strata defined?
G-methods

- Robins, 1986 and onwards
- “G” stands for “Generalized” causal comparisons
  - Including the contrast of dynamic regimes in the presence of time-varying confounders

- G-formula (Robins 1986)
- g-estimation of structural nested models (Robins 1989)
- inverse probability weighting of marginal structural models (Robins 1998)
Aside: Time-varying confounders “affected” by exposure

$A_t$: Antiretroviral therapy (0: no, 1: yes) at time $t$

$Y$: Viral load (1 if detectable, 0: otherwise)

$L$: CD4 count (0: high, 1: low)

$U$: True immunosupression level

(Unknown to data analyst: No effect of $A_t$ on $Y$)
Aside: Stratification to compute the causal effect of $A$

- Is the conditional risk ratio equal to the causal risk ratio (i.e., one)?
  - $\Pr[Y=1|A=2, L_1=l] / \Pr[Y=1|A=0, L_1=l]$

- NO

- Conditioning on $L_1$
  - eliminates confounding (blocks the back-door path) for one component of $A$, i.e., $A_1$
  - creates selection bias for the other component of $A$, i.e., $A_0$
  - As long as one component of $A$ is associated with $Y$, $A$ is associated with $Y$
Aside:

To stratify or not to stratify...

- Not stratifying is bad because there is confounding

- Stratifying is bad because stratification eliminates confounding at the cost of introducing selection bias

- Because the confounder for part of the exposure is affected by another part of the exposure
Aside:
More generally

- Bias if either the confounder is affected by the exposure or shares a common cause with it
- Bias even if the confounder is not on the causal pathway from exposure to outcome
What if the confounder is on the causal pathway?

- Conditioning on the confounder not only creates selection bias but also prevents identification of the total effect of exposure.
  - In general, stratification by intermediate variables to identify direct effects is dangerous.
In summary

- Methods that estimate association measure ignoring data on $L_1$
  - Association measure does not have a causal interpretation if there is confounding by $L_1$

- Methods that estimate association measure within levels of $L_1$
  - Association measure does not have a causal interpretation if $L_1$ affected by exposure (or a cause of the exposure)

- Need for other methods
G-formula, IP weighting, G-estimation

- Appropriately adjust for confounding when time-dependent confounders are affected by exposure (or by causes of exposure)

- For example, in IP weighting adjustment is achieved by eliminating the arrow from confounder to subsequent exposure (in the pseudo-population)
  - Not by conditioning on the confounder
Identifying assumptions

- Exchangeability
  - Sequential ignorability, no unmeasured confounding
- Positivity
  - Experimental treatment assumption
- Well-defined interventions

✓ See Robins and Hernán (2008)
The g-formula

- aka the g-computation algorithm formula (Robins 1986)
- Later rediscovered by computer scientists (Pearl et al)

- Under its assumptions, the g-formula can be used to estimate any average causal effect
- Not a causal method but the causal method
The g-formula

- Solution to causal inference from complex longitudinal data?
- Well, there is no “solution” to causal inference from observational data
  - Unmeasured confounding always a possibility
    - Lack of exchangeability given the measured covariates
  - The g-formula adjusts appropriately for time-varying measured confounding
    - Like IP weighting, unlike conventional methods
The g-formula

- Nonparametric method!!
  - Need to estimate the conditional distribution of outcome, confounders...
  - For longitudinal data and/or continuous covariates it requires essentially infinite data and computing time
  - Not a trivial problem...
    - More later

- Let’s describe the g-formula for the “when to start” question
Data and notation (I)

- Time $k$ measured in months
  - $k = 0, 1, 2, \ldots, 150$
- $A_k$
  - indicator of treatment initiation before end of month $k$
- $V$
  - vector of variables measured at or before $k=0$
    - sex, geographic origin, mode of transmission, race, cohort, calendar year years since HIV diagnosis, age
Data and notation (II)

- \( L_k \)
  - vector of time-varying variables during period \( k \)
  - indicators for measurement of viral load (\( L_{1,k} \)) and CD4 (\( L_{2,k} \)), most recently measured viral load (\( L_{3,k} \)) and CD4 (\( L_{4,k} \)), AIDS diagnosis (\( L_{5,k} \))

- \( Y_{k+1} \)
  - indicator of death by end of month \( k+1 \)

- \( C_{k+1} \)
  - indicator of censoring by end of month \( k+1 \)
Dynamic regimes of interest

- "start cART when CD4 cell count first drops below x or there is a diagnosis of an AIDS-defining illness, whichever happens first"

- Regimes indexed by x

- Formalization:

  Let \( g_k(\bar{a}_{k-1}, \bar{l}_k) = 1 \) if either

  1. \( g_{k-1}(\bar{a}_{k-2}, \bar{l}_{k-1}) = 0 \) and \( l_{4,k} < x \) or \( l_{5,k} = 1 \) or
  2. \( g_{k-1}(\bar{a}_{k-2}, \bar{l}_{k-1}) = 1 \)

  and let \( g_k(\bar{a}_{k-1}, \bar{l}_k) = 0 \) otherwise for \( k \geq 0 \).
Estimand of interest

\[ \Pr[ Y_{k+1}^x = 1 ] \]

- counterfactual risk of death by month \( k+1 \) for \( k=0,...,60 \) under treatment regime \( x \)
  - and with censoring by loss to follow-up abolished

- The g-formula expresses this counterfactual risk in terms of only the observed data distribution
  - Under the identifying assumptions
The g-formula for
\[ \Pr[Y_{k+1}^x = 1] \]

\[ \sum_{l_k} \sum_{j=0}^{k} \Pr(Y_{j+1} = 1|L_j = l_j, A_j = a_j, Y_j = C_{j+1} = 0) \times \]

\[ \prod_{s=0}^{j} \{ \Pr(Y_s = 0|L_{s-1} = l_{s-1}, A_{s-1} = a_{s-1}, Y_{s-1} = C_s = 0) \times \]

\[ f(l_s|l_{s-1}, a_{s-1}, Y_s = C_s = 0) \} \]

- General form of epidemiologic standardization for time-varying treatments and confounders
Two nontrivial problems

1. Need to estimate the entire likelihood
   - Nonparametric estimation is out of the question
     - except marginal distribution of baseline covariates $V$ can be empirically estimated
   - Parametric estimation perceived as too dangerous

2. Need to compute a huge integral
   - Approximated via Monte Carlo simulation
Should we consider parametric estimation of the g-formula?

- A big chunk of Robins et al’s careers devoted to semiparametric methods
  - IP weighting of MSMs, g-estimation of SNMs
- Model misspecification viewed as particularly bad for longitudinal data
  - Can lead to propagation of errors
- Now we are talking about going fully parametric? Really?
  - Well, you never know until you try
Parametric vs. semiparametric

- IP weighting of dynamic MSMs
  - Model for treatment initiation
  - Structural model to smooth over regimes

- Parametric g-formula
  - Parametric models for every single density in the likelihood of the observed data
Implementation of parametric g-formula

- 3 steps
  1. Parametric modeling to estimate factors of the g-formula
  2. Monte Carlo simulation to approximate the integral
  3. Computation of risk
     - Nonparametric bootstrap for variance estimation
     - Technical details, including on random dynamic regimes, in Young et al. *Stat Biosci* 2011

- Publicly available SAS macro
  - [http://www.hsph.harvard.edu/ causal](http://www.hsph.harvard.edu/causal)
Step 1: Parametric modeling of conditional densities

1. Fit models for the conditional densities of the covariates
   \[ f(l_s|\overline{l}_{s-1}, \overline{a}_{s-1}, \overline{Y}_s = \overline{C}_s = 0) \]
   - e.g., logistic for discrete, log-linear for continuous
   - Use empirical distribution for baseline covariates

2. Fit a logistic model for the conditional density of the outcome
   \[ \Pr[Y_{k+1} = 1|\overline{L}_k = \overline{l}_k, \overline{A}_k = \overline{a}_k, \overline{Y}_k = \overline{C}_{k+1} = 0] \]
Step 2: Monte Carlo simulation under regime $x$

For $k = 0, \ldots, K$ and $v = 1, \ldots, n$:

1. If $k = 0$ set $l_{0,v}$ to the observed values for subject $v$. Otherwise if $k > 0$
   (a) Set $l_{1,k,v} = 1$ if $\sum_{s=1}^{12} l_{1,k-s,v} = 0$ and $k \geq 12$. Otherwise, draw $l_{1,k,v}$ from the
       probability function estimated in step I.1.a based on previously drawn covariates $\bar{l}_{k-1,v}$ and assigned treatment $\bar{a}_{k-1,v}$ under $x$ (see step II.2).
   (b) Set $l_{2,k,v} = l_{2,k-1,v}$ if $l_{1,k,v} = 0$. Otherwise, draw $l_{2,k,v}$ from the density function estimated in step I.1.b based on previously drawn covariates $\bar{l}_{k-1,v}$ and assigned treatment $\bar{a}_{k-1,v}$ under $x$.
   (c) Set $l_{3,k,v} = 1$ if $\sum_{s=1}^{12} l_{3,k-s,v} = 0$ and $k \geq 12$. Otherwise, draw $l_{3,k,v}$ from the
       probability function estimated in step I.1.c based on previously drawn covariates $l_{2,k,v}, l_{1,k,v},\bar{l}_{k-1,v}$ and assigned treatment $\bar{a}_{k-1,v}$ under $x$. 

8 June 2012 G-formula for CER
Step 2: Monte Carlo simulation under regime $x$

(d) Set $l_{4,k,v} = l_{4,k-1,v}$ if $l_{3,k,v} = 0$. Otherwise, draw $l_{4,k,v}$ from the density function estimated in step I.1.d based on previously drawn covariates $l_{2,k,v}, l_{1,k,v}, \tilde{l}_{k-1,v}$ and assigned treatment $\tilde{a}_{k-1,v}$ under $x$.

(e) Set $l_{5,k,v} = 1$ if $l_{5,k-1,v} = 1$. Otherwise, draw $l_{5,k,v}$ from the probability function estimated in step I.1.e based on previously drawn covariates $\tilde{l}_{4,k,v}, \tilde{l}_{3,k,v}, \tilde{l}_{2,k,v}, \tilde{l}_{1,k,v}$, and assigned treatment $\tilde{a}_{k-1,v}$ under $x$.

2. Assign the treatment $a_{k,v}$ under $x$. Specifically, set $a_{-1,v} = 0$. 
Step 3: Computation of risk under regime $x$

For $k=0, 1, \ldots K$, estimate the g-formula as

$$
\frac{1}{n} \sum_{v=1}^{n} \sum_{j=0}^{k} \hat{\Pr}[Y_{j+1} = 1|L_j = \bar{l}_{j,v}, A_j = \bar{a}_{j,v}, Y_j = \bar{C}_{j+1} = 0] \times
$$

$$
\prod_{s=0}^{j} \{1 - \hat{\Pr}[Y_s = 1|L_{s-1} = \bar{l}_{s-1,v}, A_{s-1} = \bar{a}_{s-1,v}, Y_{s-1} = \bar{C}_s = 0]\}
$$
## Point estimates and 95% CIs

<table>
<thead>
<tr>
<th>x</th>
<th>Risk (%)</th>
<th>Risk Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>500</td>
<td>2.65 (2.15, 3.43)</td>
<td>1.00 (ref)</td>
</tr>
<tr>
<td>450</td>
<td>2.77 (2.28, 3.47)</td>
<td>1.05 (0.99, 1.09)</td>
</tr>
<tr>
<td>400</td>
<td>2.92 (2.43, 3.55)</td>
<td>1.10 (0.99, 1.18)</td>
</tr>
<tr>
<td>350</td>
<td>3.06 (2.59, 3.67)</td>
<td>1.16 (1.00, 1.29)</td>
</tr>
<tr>
<td>300</td>
<td>3.22 (2.73, 3.81)</td>
<td>1.22 (1.02, 1.41)</td>
</tr>
<tr>
<td>250</td>
<td>3.44 (2.87, 4.07)</td>
<td>1.30 (1.03, 1.55)</td>
</tr>
<tr>
<td>200</td>
<td>3.65 (3.01, 4.44)</td>
<td>1.38 (1.05, 1.71)</td>
</tr>
</tbody>
</table>
Survival under 3 dynamic regimes
Stability of results

- Results did not materially change under different modeling strategies
  - functional forms for continuous variables, e.g., polynomial, restricted cubic splines
  - inclusion of "interaction" terms
  - modeling of measurement process

- More formal model selection and goodness of fit tests are ongoing
The “natural course”

- One necessary condition is that our approach predicts the observed data correctly

- That is, we need to be able to predict what would had happened under no treatment intervention
Natural course

- Observed vs.
  - Top: absolute
  - Bottom: difference (95% CI)

8 June 2012

G-formula for CER
Natural course

- Observed vs. predicted
- Top: absolute
- Bottom: difference (95% CI)

8 June 2012  G-formula for CER
Conclusions (I)

- Parametric estimation more efficient than semiparametric estimation, duh!
  - Compared with IP weighted estimates in Cain et al. *Annals of Internal Medicine* 2011

- Parametric g-formula is more flexible than any of the other g-methods
  - and is computationally tractable
Conclusions (II)

- Most shockingly, so far the parametric g-formula yields reasonable/stable estimates
  - More applications coming
  - Weight loss and diabetes, fish and heart disease...
Acknowledgements

□ Co-investigators
  ■ At HSPH: Jessica Young, Lauren Cain, Jamie Robins, Roger Logan
  ■ Members of the HIV-CAUSAL Collaboration

□ Funding Sources
  ■ NIH grants: R01-AI073127 and U10-AA013566