Part VII:

Time-dependent covariates

Blood glucose maintenance in the ICU

- When patients are admitted to an intensive care unit (ICU), an important goal is to restore physiological systems to their normal function
- Tight control of blood glucose (BG) concentration, for example, may prevent life-threatening adverse events
- One lever that nurses and doctors have at their disposal for controlling BG is the *insulin dose rate* (IDR)
  - ⋆ insulin is administered intravenously
  - ⋆ the rate at which it is administered can be modified
- Suppose, therefore, that interest lies in understanding the causal relationship between IDR and BG concentration
  - ⋆ knowing this would (presumably) be useful for future decision-making

- Towards investigating this relationship, the `glucose` dataset consists of K=345 patients from the Vanderbilt University trauma ICU
  - ⋆ admitted between 05/2004 and 12/2005
  - ⋆ focus on first 24 hours of their hospitalization
  - ⋆ n_k ranges from 3 to 14

- Variable key:

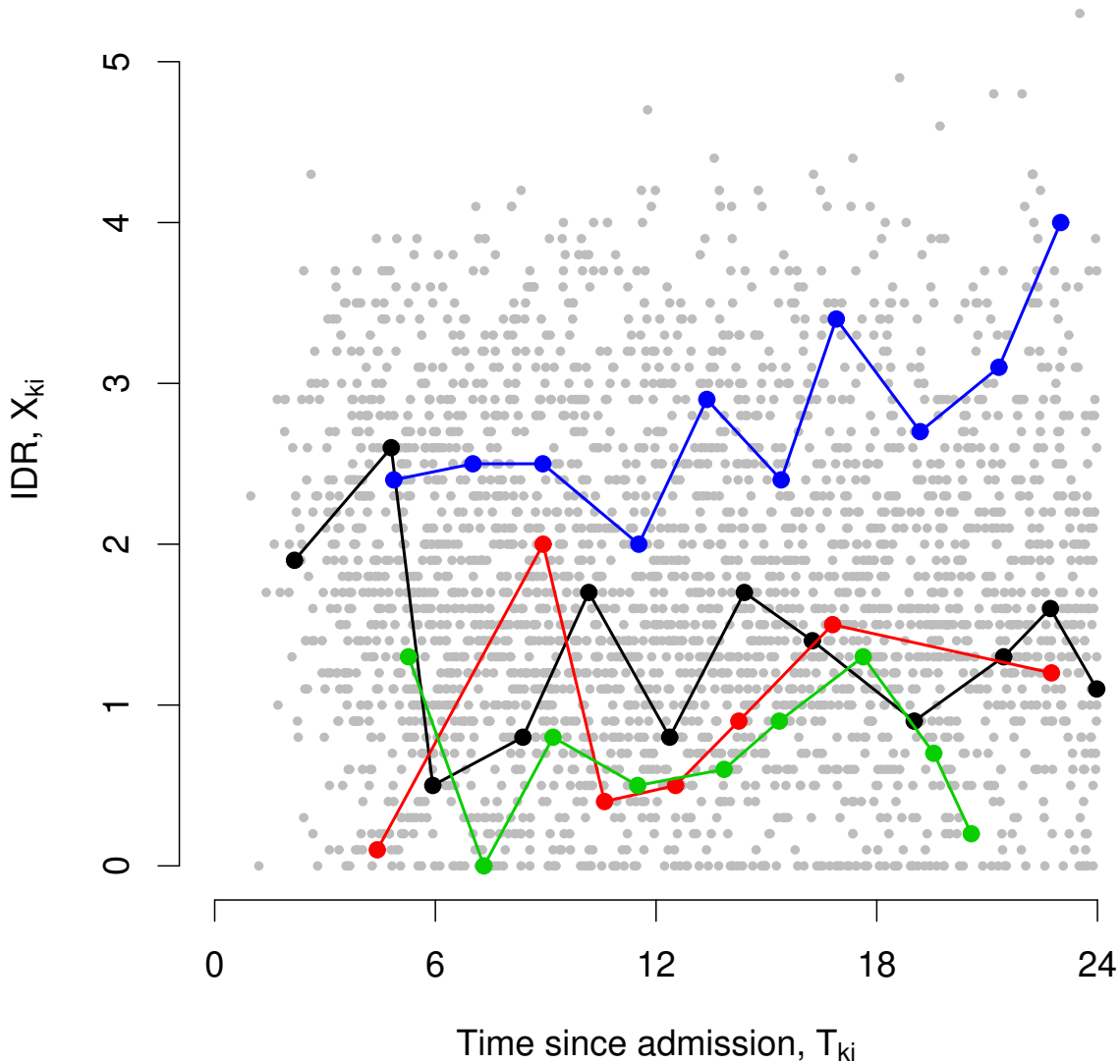
`id`	unique patient identifier
`Tki`	time since admission to the ICU
`Xki`	IDR
`Yki`	BG concentration

- Note, BG level is measured after the administration of insulin, so that X_{ki} precedes Y_{ki} in time

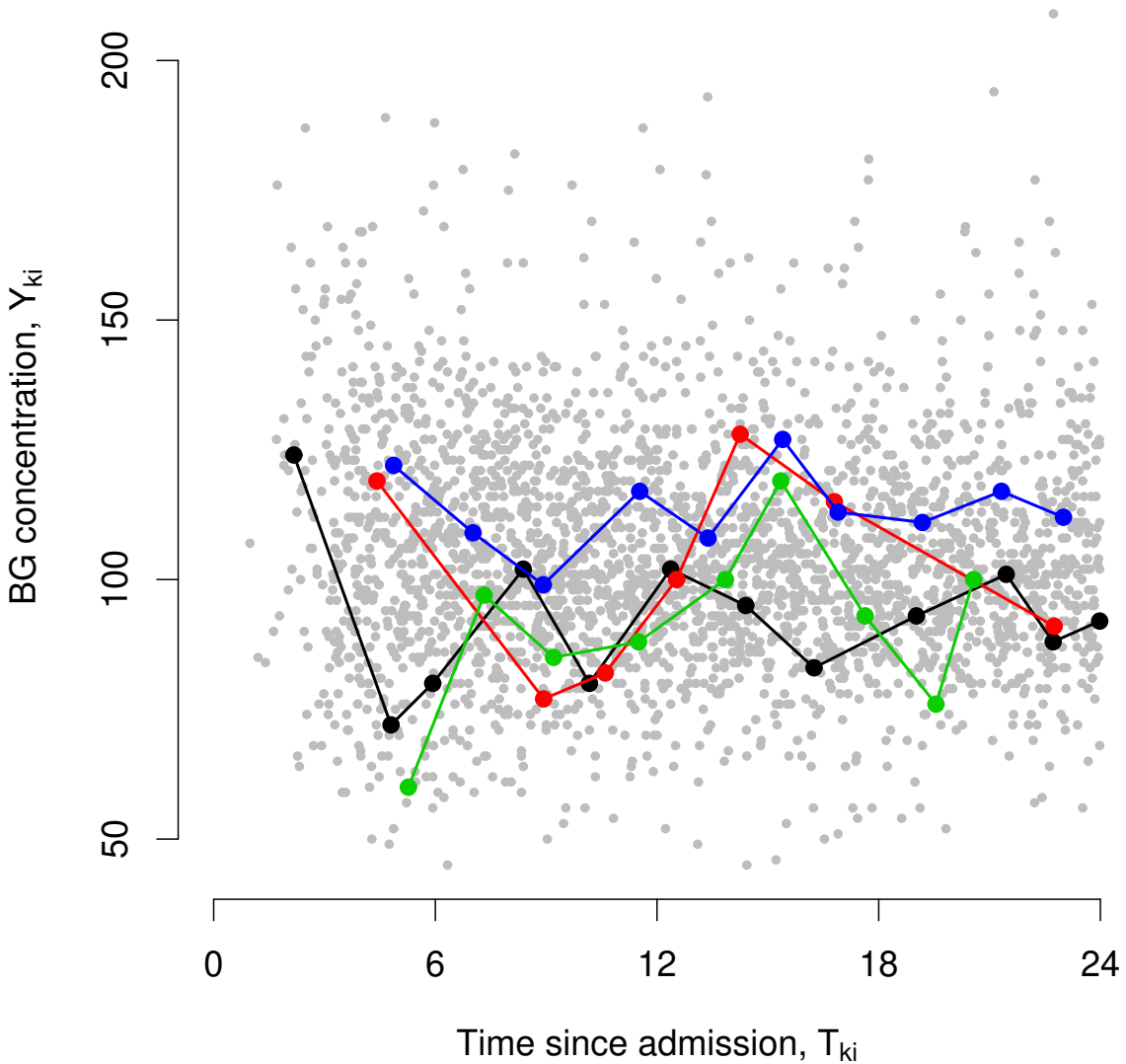
```
> ##
> dim(glucose)
[1] 2858    9
>
> head(glucose)
   id   Tki Xki Yki
1   1  1.90 2.9 122
2   1  7.65 2.5  82
3   1  8.33 0.0  66
4   1 10.83 0.0  95
5   1 11.93 0.7 106
15  2  3.90 2.1  76
>
> ##
> length(unique(glucose$id))
[1] 345
> table(table(glucose$id))

 3  4  5  6  7  8  9 10 11 12 13 14
23 24 16 21 28 34 58 72 55 11  2  1
```

```
> ## Timing of measurements
> ##
> round(summary(glucose$Tki), 2)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
    0.1     7.7    12.5    12.8    17.9    24.0
>
> ## Insulin dose rates (IDR)
> ##
> round(summary(glucose$Xki), 2)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   0.00    1.00    1.60    1.74    2.40    5.30
>
> ## Blood glucose (BG) concentration levels
> ##
> round(summary(glucose$Yki), 2)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   45.0    91.0   102.0   104.5   117.0   209.0
```

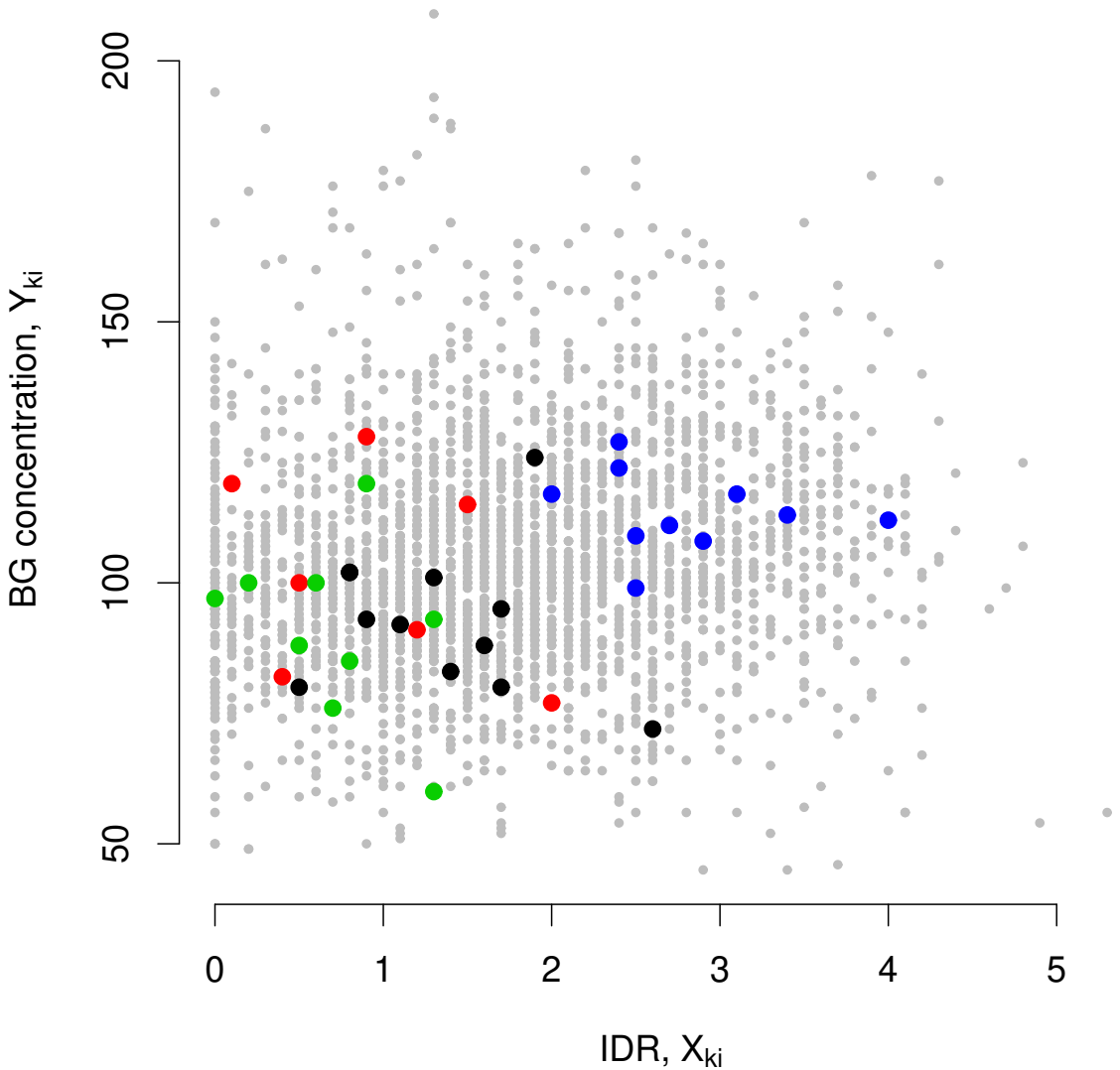
- IDR during first 24 hours:



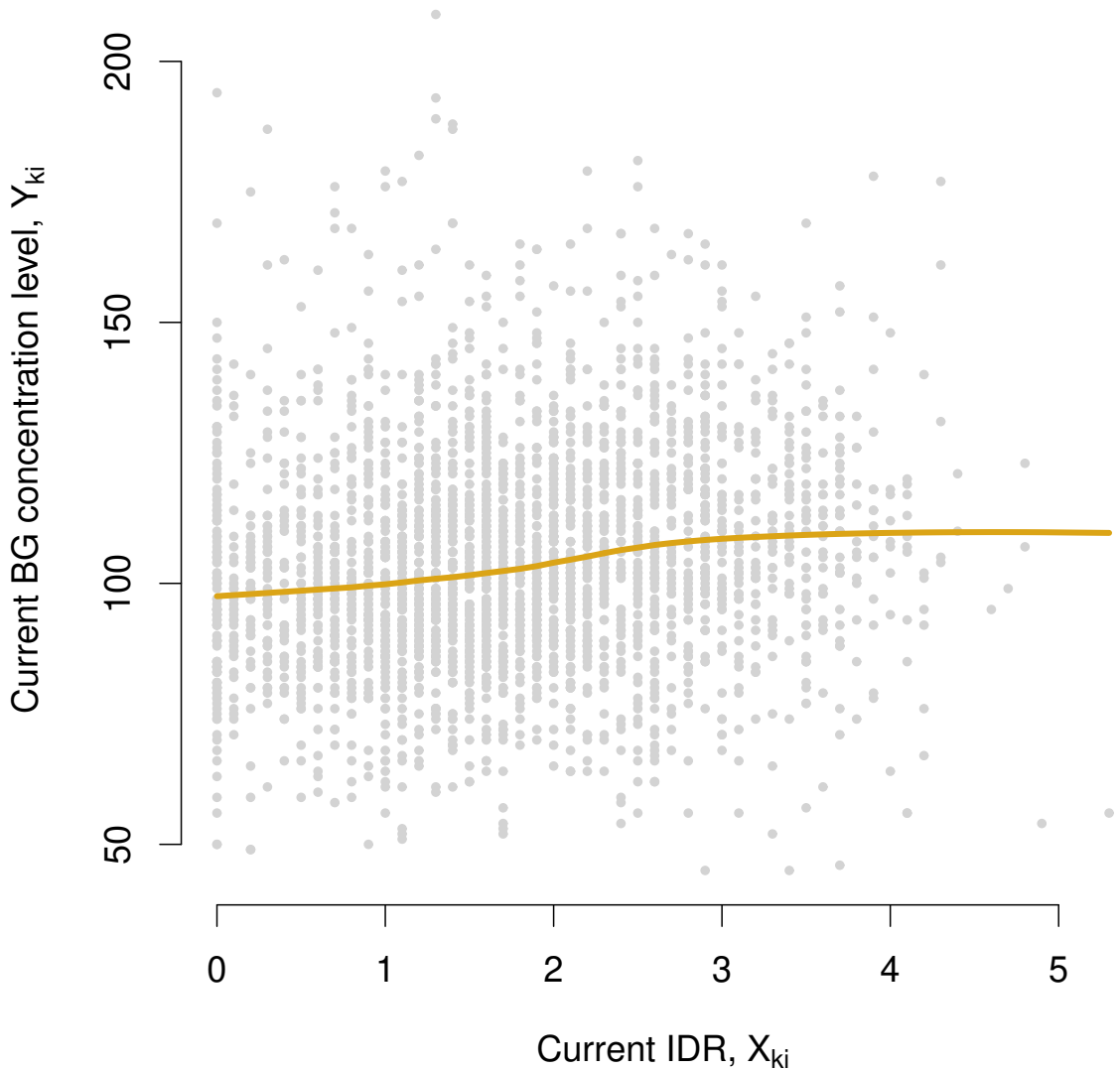
- BG concentration during first 24 hours:



- IDR vs BG concentration during the first 24 hours:



- Applying a lowess smoother suggests that there is a small positive association between IDR and BG concentration:



Targets of inference

- Consider using these data to answer the question:

  Q: *What is the effect of insulin dose rate on blood glucose concentration?*

- As written, this question is vaguely formulated
  - ⋆ i.e. doesn't provide a precise framework for conducting analyses

- As such, an analysis could proceed in a number of ways, depending on a series of decisions made by the investigative team

- This is often the case in real collaborative settings, especially if the data were not collected for the purposes of some specific question
  - ⋆ e.g. EMR data or as in the big Harvard cohorts

- In terms of specific decisions that need to be made, clearly a critical decision is how to formulate the response

- Recall that a key benefit of longitudinal data is that one can examine the impact of covariates, or changes in covariates, on the change in the response from timepoint-to-timepoint

- For the IDR-BG association, therefore, one could take the response to be:

  (1) BG concentration level at a particular time point
  - ∗ i.e. Y_{ki}

  (2) BG concentration change at a particular time point
  - ∗ i.e. $Z_{ki} = Y_{ki} - Y_{k,i-1}$

- In addition, one might consider whether or not to specify some threshold and use a dichotomized version of the response
  - ⋆ e.g. whether the change in BG is greater than 25%

- A series of decisions also have to be made as to how the primary exposure of interest is to be included, as well as any adjustment variables

Adjust for confounding
Increase efficiency

- Collectively all of these decisions result in a model in which key parameter(s) can be identified as being central to the question
  - ⋆ i.e. the more precisely formulated question
  - ⋆ refer to these parameters as *targets of inference*

- In the next few slides we'll consider six specific models
  - ⋆ differ in the response
  - ⋆ differ in the specification of the linear predictor
  - ⋆ throughout $\boldsymbol{X}_{ki}^{*}$ is a vector of adjustment variables

- Note, although a common notation is used across the six models, the regression parameters are distinct and have different interpretations
  - ⋆ see Schildcrout et al (2011) for connections between the models

Cross-sectional mean models

Adjustment variable

$$Y_{ki} = \beta_0 + \boldsymbol{\beta}_* \boldsymbol{X}^*_{ki} + \beta_x X_{ki} + \epsilon_{ki} \quad \text{(1-CS)}$$

$$Z_{ki} = \beta_0 + \boldsymbol{\beta}_* \boldsymbol{X}^*_{ki} + \beta_x X_{ki} + \epsilon_{ki} \quad \text{(2-CS)}$$

- Marginal models that relate current IDR levels to current BG concentration levels/change

Distributed lag mean models

$$Y_{ki} = \beta_0 + \boldsymbol{\beta}_* \boldsymbol{X}^*_{ki} + \beta_x X_{ki} + \beta_{x,l} X_{k,i-1} + \epsilon_{ki} \quad \text{(1-DL)}$$

$$Z_{ki} = \beta_0 + \boldsymbol{\beta}_* \boldsymbol{X}^*_{ki} + \beta_x X_{ki} + \beta_{x,l} X_{k,i-1} + \epsilon_{ki} \quad \text{(2-DL)}$$

- Marginal models that permits characterization of the joint impact of previous and current IDR levels on the current response
  - ⋆ via the inclusion of time-dependent exposure at the previous time point
  - ⋆ could include multiple lags

Transition mean models

$$Y_{ki} = \beta_0 + \boldsymbol{\beta}_* \boldsymbol{X}^*_{ki} + \beta_x X_{ki} + \beta_{y,l} Y_{k,i-1} + \epsilon_{ki} \qquad \text{(1-TN)}$$

$$Z_{ki} = \beta_0 + \boldsymbol{\beta}_* \boldsymbol{X}^*_{ki} + \beta_x X_{ki} + \beta_{y,l} Y_{k,i-1} + \epsilon_{ki} \qquad \text{(2-TN)}$$

- Additional adjustment for the previous response levels:
  - ⋆ may be of intrinsic scientific interest
  - ⋆ may be included to improve model fit
  - ⋆ may be included if previous response levels are potential confounders
    - ∗ more on this later
- Note, transition models are not marginal with respect to cluster membership
  - ⋆ marginally, the dependence structure is dictated, in part, by the mean model
  - ⋆ similar to a mixed effects model

Summary

Cross-sectional mean models:

$$Y_{ki} = \beta_0 + \boldsymbol{\beta}_* \boldsymbol{X}^*_{ki} + \beta_x X_{ki} + \epsilon_{ki} \quad \text{(1-CS)}$$

$$Z_{ki} = \beta_0 + \boldsymbol{\beta}_* \boldsymbol{X}^*_{ki} + \beta_x X_{ki} + \epsilon_{ki} \quad \text{(2-CS)}$$

Distributed lag models:

$$Y_{ki} = \beta_0 + \boldsymbol{\beta}_* \boldsymbol{X}^*_{ki} + \beta_x X_{ki} + \beta_{x,l} X_{k,i-1} + \epsilon_{ki} \quad \text{(1-DL)}$$

$$Z_{ki} = \beta_0 + \boldsymbol{\beta}_* \boldsymbol{X}^*_{ki} + \beta_x X_{ki} + \beta_{x,l} X_{k,i-1} + \epsilon_{ki} \quad \text{(2-DL)}$$

Transition models:

$$Y_{ki} = \beta_0 + \boldsymbol{\beta}_* \boldsymbol{X}^*_{ki} + \beta_x X_{ki} + \beta_{y,l} Y_{k,i-1} + \epsilon_{ki} \quad \text{(1-TN)}$$

$$Z_{ki} = \beta_0 + \boldsymbol{\beta}_* \boldsymbol{X}^*_{ki} + \beta_x X_{ki} + \beta_{y,l} Y_{k,i-1} + \epsilon_{ki} \quad \text{(2-TN)}$$

Preliminary results

- Point estimates based on GEE under three working dependence structures:

Model/covariate	**BG level, Y_{ki}**			**BG change, Z_{ki}**		
	Ind	Exch	AR-1	Ind	Exch	AR-1
Cross-sectional						
IDR, β_x	6.3	0.9	-13.7	-32.6	-51.1	-25.7
Distributed lag						
IDR, β_x	9.7	4.0	-17.7	-42.9	-49.6	-39.4
lagged IDR, $\beta_{x,l}$	-5.3	-7.7	-6.4	16.0	11.6	14.4
Transition						
IDR, β_x	-17.6	-27.7	-16.4	-17.6	-27.7	-16.4
lagged BG, $\beta_{y,l}$	0.6	0.6	0.7	-0.4	-0.4	-0.3

Change of dependence structures makes a big difference

⋆ adjusted for a range of potential confounders

⋆ see Schildcrout et al (2011) for details

Comments

- The choices presented are not meant to be exhaustive
  - ⋆ many ways of formulating models
  - ⋆ e.g. we could have specified additional lags in either $(1\text{-}DL)$ or $(2\text{-}DL)$, or added lagged exposure terms into $(1\text{-}TN)$ or $(2\text{-}TN)$
- That there are such striking differences in the point estimates among the models presented may be concerning
- Results illustrate that as we conduct analyses that involve time-dependent covariates a number of important issues need to be considered:
  - ⋆ the nature/type of the time-dependent covariate
  - ⋆ the choice of working dependence structure
  - ⋆ the control of confounding bias

- As we will see, the extent to which these issues create meaningful problems (i.e. bias) depends on the model type and specification

- Particularly problematic is when the exposure of interest is an *endogenous time-dependent covariate*
  - ⋆ IDR is an example of such a covariate

Types of time-dependent covariates

- Towards distinguishing different types, we'll find it useful to use the following notation:
  - ⋆ $Y_k(t)$ is the value of the response for the k^{th} subject at time t
  - ⋆ $X_k(t)$ is the value of the exposure for the k^{th} subject at time t
  - ⋆ $\mathcal{Y}_k(t)$ is the history of the response process until time t:

$$\mathcal{Y}_k(t) \;=\; \{Y_k(t')|\; t' < t\}$$

  - ⋆ $\mathcal{X}_k(t)$ is the history of the exposure process until time t:

$$\mathcal{X}_k(t) \;=\; \{X_k(t')|\; t' < t\}$$

  - ⋆ $\boldsymbol{X}_k^*$ is a vector of time-invariant (or baseline) covariates

Exogenous time-dependent covariates

- $X_k(t)$ is *exogenous* with respect to the outcome process if the exposure at time t is conditionally independent of the history of the outcome process at time t, given the history of the exposure process at time t:

$$[X_k(t)|\ \mathcal{Y}_k(t), \mathcal{X}_k(t), \boldsymbol{X}_k^*] \ = \ [X_k(t)|\ \mathcal{X}_k(t), \boldsymbol{X}_k^*]$$

Independent
X_k(t) does not depents on past Y

- Intuitively, exposure at time t does not depend on prior response values

- Examples:
  - ⋆ age in the dental growth data
  - ⋆ ambient air pollution measurements

- Note, exposures that are time-invariant are intrinsically exogenous
  - ⋆ e.g. point exposures such as type of bariatric surgery
  - ⋆ e.g. genetic profiles in cancer studies

Endogenous time-dependent covariates

- $X_k(t)$ is *endogenous* with respect to the outcome process if the exposure at time t is conditionally dependent (in some way) on the history of the outcome process at time t, given the history of the exposure process at time t:

$$[X_k(t)|\ \mathcal{Y}_k(t), \mathcal{X}_k(t), \boldsymbol{X}_k^*] \ \neq\ [X_k(t)|\ \mathcal{X}_k(t), \boldsymbol{X}_k^*]$$

- Intuitively, exposure at time t depends (in part) on prior response values

- Potential examples:
  - ⋆ IDR for patients in the ICU may depend on prior BG levels
  - ⋆ personal air pollution exposure may depend on prior symptoms
  - ⋆ ART for patients with HIV may depend on disease progression
  - ⋆ adherence in randomized trials may depend on the patients response to their allocated treatment

- Whether the exposure of interest is exogenous or endogenous dictates, in part, the analysis strategy

- If the exposure is exogenous, the analysis can focus on specifying the lag dependence in models for the mean
  - ⋆ i.e. $X_k(t)$, $X_k(t-1)$, $X_k(t-2)$, ...
  - ⋆ care is needed when, however, choosing the dependence structure (see the next section)

- If the exposure is endogenous, then the potential for time-dependent confounding is a serious problem
  - ⋆ standard methods (i.e. regression adjustment) will generally fail to obtain 'true' causal estimates
  - ⋆ must focus on selecting a meaningful target of inference together with a method for valid estimation/inference

Choice of working dependence structure

- Suppose that interest lies in the following cross-sectional marginal mean model:

$$\mu_{ki} \equiv \mathrm{E}[Y_{ki}|\ \boldsymbol{X}_{ki}] = \boldsymbol{X}_{ki}^T\boldsymbol{\beta}$$

- For a given specification of the working dependence structure, the estimating equations for GEE 1.0 are:

$$\boldsymbol{U}(\boldsymbol{\beta}) = \sum_{k=1}^{K} \boldsymbol{D}_k^T \boldsymbol{V}_k^{-1}(\boldsymbol{Y}_k - \boldsymbol{\mu}_k) = \mathbf{0}$$

- Recall, the asymptotic properties of the resulting estimator rely on the estimating equations being unbiased
  - ⋆ i.e. $\mathrm{E}[\boldsymbol{U}(\boldsymbol{\beta})] = \mathbf{0}$ where the expectation is with respect to the distribution of $\boldsymbol{Y}_k$

- Re-writing the estimating function as:

$$\boldsymbol{U}(\boldsymbol{\beta}) \;=\; \sum_{k=1}^{K} \boldsymbol{U}_k(\boldsymbol{\beta}) \;=\; \sum_{k=1}^{K} \boldsymbol{D}_k^T \boldsymbol{W}_k (\boldsymbol{Y}_k - \boldsymbol{\mu}_k)$$

  one can show that the contribution to $\boldsymbol{U}(\boldsymbol{\beta})$ from the k^{th} subject/cluster is a $p \times 1$ vector is given by:

  dimension of beta

  Unit specific residuals

$$\boldsymbol{U}_k(\boldsymbol{\beta}) \;=\; \sum_{i=1}^{n_k} \left\{ \sum_{j=1}^{n_k} \boldsymbol{D}_{kj} W_{k,ij} \right\} (Y_{ki} - \mu_{ki})$$

  - ⋆ i.e. a weighted average of the study unit-specific 'residuals'

- Inspecting the weights, we see that the contribution to $\boldsymbol{U}_k(\boldsymbol{\beta})$ made by the i^{th} study unit depends on:
  - ⋆ choice of working dependence structure, via $W_{k,ij}$
  - ⋆ totality of the covariates for the k^{th} cluster/subject, via the $\boldsymbol{D}_{kj}$ summed over j

- Now consider the expectation of each summand in $\boldsymbol{U}_k(\boldsymbol{\beta})$:

$$\begin{aligned}
\mathsf{E}_{\boldsymbol{Y}}[\boldsymbol{D}_{kj} W_{k,ij}(Y_{ki} - \mu_{ki})] &= \mathsf{E}_{\boldsymbol{X}}[\mathsf{E}_{\boldsymbol{Y}|\boldsymbol{X}}[\boldsymbol{D}_{kj} W_{k,ij}(Y_{ki} - \mu_{ki})|\ \boldsymbol{X}_k]] \\
&= \mathsf{E}_{\boldsymbol{X}}[\boldsymbol{D}_{kj} W_{k,ij} \mathsf{E}_{\boldsymbol{Y}|\boldsymbol{X}}[Y_{ki} - \mu_{ki}|\ \boldsymbol{X}_k]] \\
&= \mathsf{E}_{\boldsymbol{X}}[\boldsymbol{D}_{kj} W_{k,ij}(\mathsf{E}_{\boldsymbol{Y}|\boldsymbol{X}}[Y_{ki}|\ \boldsymbol{X}_k] - \mu_{ki})]
\end{aligned}$$

  - ⋆ notice that in order to perform the second of these operations we have to condition on the entire vector $\boldsymbol{X}_k$

- If the estimation equation is to be unbiased we must therefore have:

$$\mathsf{E}[Y_{ki}|\ \boldsymbol{X}_k] = \mu_{ki}$$

  or equivalently that

$$\mathsf{E}[Y_{ki}|\ \boldsymbol{X}_{ki}] = \mathsf{E}[Y_{ki}|\ \boldsymbol{X}_{k1}, \ldots, \boldsymbol{X}_{kn_k}] \quad \forall\ i = 1, \ldots, n_k$$

  - ⋆ *full covariate conditional mean* assumption
  - ⋆ Pepe and Anderson (1994)

- Note the assumption is required in addition to the correct specification of the marginal mean model, $\mathrm{E}[Y_{ki}|\ \boldsymbol{X}_{ki}]$

- In some settings, the assumption will be trivially satisfied:
  - ⋆ if all of the components of $\boldsymbol{X}_{ki}$ are time-invariant
  - ⋆ if all lagged covariates that are predictive of the outcome are included in the model

- If the assumption is not satisfied, however, all is not lost and one can obtain consistent estimates of $\boldsymbol{\beta}$ by adopting a working independence correlation structure
  - ⋆ under working independence, $W_{k,ij} = 0\ \forall\ i \neq j$ so that the contribution to $\boldsymbol{U}(\boldsymbol{\beta})$ from the k^{th} subject/cluster simplifies to:

$$\boldsymbol{U}_k(\boldsymbol{\beta}) \ = \ \sum_{i=1}^{n_k} \boldsymbol{D}_{ki} W_{k,ii} (Y_{ki} - \mu_{ki})$$

  - ⋆ going through the operations to show that the corresponding estimating equation is unbiased only requires conditioning on $\boldsymbol{X}_{ki}$

Simple illustration

- To illustrate the potential for bias when the analysis involves a time-dependent exposure, consider the following data generating mechanism:

$$Y_k(t) \;=\; \theta_0 \;+\; \theta_1 X_k(t) \;+\; \theta_2 X_k(t-1) \;+\; \gamma_{0k} \;+\; \epsilon_k(t)$$

  where $X_k(t) = \rho_x X_k(t-1) \;+\; \mathrm{e}_k(t)$

  - ⋆ a distributed lag random intercepts model, with an exogenous time-dependent covariate

- Given this specification, and letting $\overline{\boldsymbol{X}}_k(T_k) = \{\boldsymbol{X}_k(1), \boldsymbol{X}_k(2), \ldots, \boldsymbol{X}_k(T_k)\}$, the induced full conditional mean and cross-sectional mean models are:

$$\begin{aligned} \mathrm{E}[Y_k(t) |\; \overline{\boldsymbol{X}}_k(T_k)] \;&=\; \theta_0 \;+\; \theta_1 X_k(t) \;+\; \theta_2 X_k(t-1) \\ \mathrm{E}[Y_k(t) |\; \boldsymbol{X}_k(t)] \;&=\; \beta_0 \;+\; \beta_1 X_k(t) \end{aligned}$$

  where $\beta_0 = \theta_0$ and $\beta_1 = \theta_1 + \rho_x \theta_2$

- Consequently, the full covariate conditional mean assumption is not satisfied so that even if the cross-sectional mean model is correctly specified bias may arise

- To see the potential for bias a little more explicitly, let $t' = t - 1$ and consider the expectation of the contribution to $\boldsymbol{U}_k(\boldsymbol{\beta})$ made by the i^{th} study unit:

$$\mathrm{E}[\boldsymbol{D}_{kt'} W_{k,tt'} (Y_{kt} - \mu_{kt}) | \, \overline{\boldsymbol{X}}_k(T_k)] \;=\; W_{k,tt'} \;\times\; \theta_2 \;\times\; (1 - \rho_x^2)$$

  ⋆ see Diggle et al (2002) Chapter 12

- Consequently, bias will arise when: As shown in the above formula

  (1) the exposure is time-dependent (i.e. $\rho_x \neq 1$)

  (2) there is a lag relationship between the exposure and the response (i.e. $\theta_2 \neq 0$)

  (3) covariance weighting is used (i.e. $W_{k,tt'} \neq 0$)

- Simulation study with $\boldsymbol{\theta} = (0,\ 1,\ 1)$ and

$$\begin{aligned} \gamma_{0k} &\sim \text{Normal}(0,\ 1) \\ \epsilon_k(t) &\sim \text{Normal}(0,\ 1) \\ \mathsf{e}_k(t) &\sim \text{Normal}(0,\ 1-\rho_X^2) \\ T_k &\sim \text{Uniform}\{2,\ldots,10\} \end{aligned}$$

Covariate correlation, ρ_X	**0.90**	**0.70**	**0.50**	**0.30**	**0.10**
Induced β_1	1.90	1.70	1.50	1.30	1.10
Mean $\widehat{\beta}_1$					
Working independence	1.90	1.70	1.50	1.30	1.10
Working exchangeable	1.73	1.51	1.33	1.19	1.01
Working auto-regressive	1.73	1.36	1.11	0.89	0.74
Percent bias					
Working independence	0%	0%	0%	0%	0%
Working exchangeable	-9%	-11%	-11%	-8%	-8%
Working auto-regressive	-9%	-20%	-26%	-32%	-33%

- Schildcrout and Heagerty (2005) provide general bias calculations for continuous response data with an AR(1) time-dependent exogenous covariate
  - ⋆ bias in $\widehat{\beta}_1$ from GEE with a working exchangeable dependence structure:

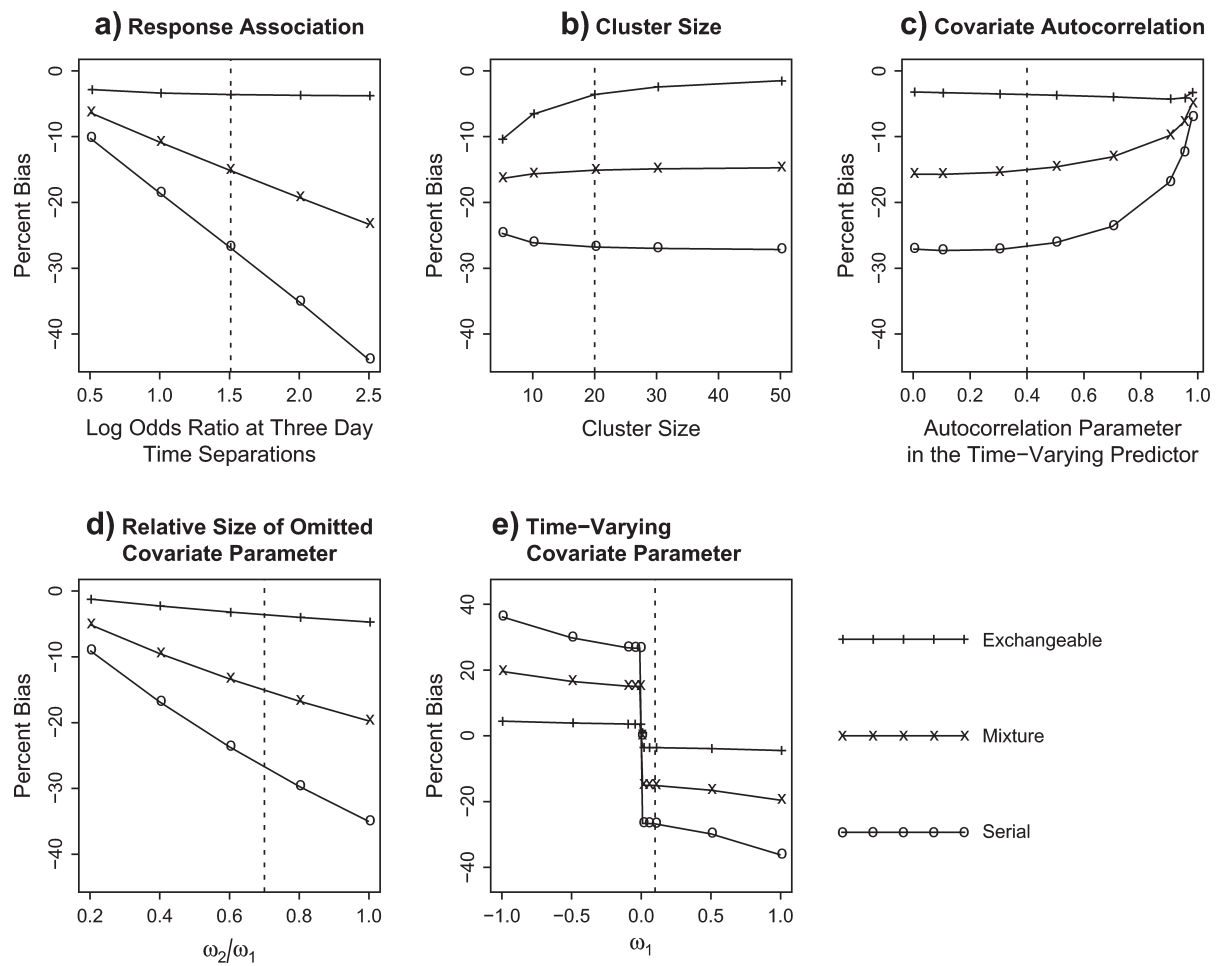
$$-\frac{\theta_2 \rho_{\mathsf{E}}(1-\rho_X^2)[n(1-\rho_{\mathsf{X}})-1+\rho_X^n]}{\sigma_{\mathsf{E}}^2[n^2\rho_{\mathsf{E}}(1-\rho_X)^2-n(2\rho_{\mathsf{E}}-1+\rho_X)(1-\rho)-2\rho_{\mathsf{E}}\rho_X(\rho_X^n-1)]}$$

  - ⋆ bias in $\widehat{\beta}_1$ from GEE with a working auto-regressive dependence structure:

$$-\frac{\theta_2 \rho_{\mathsf{AR1}}(1-\rho_X^2)(n-1)}{\sigma_{\mathsf{AR1}}^2[n(\rho_{\mathsf{AR1}}^2-2\rho_X\rho_{\mathsf{AR1}}+1)+2\rho_{\mathsf{AR1}}(\rho_X-\rho_{\mathsf{AR1}})]}$$

- Notes:
  - ⋆ if θ_1 and θ_2 have the same sign, the bias will likely be an attrition
  - ⋆ magnitude of bias grows with ρ_{E} and ρ_{AR1}
  - ⋆ bias for GEE-Ex is inversely related to the cluster size, n larger number of repeated measures, less bias
  - ⋆ no bias if no response dependence, ρ_X=1, or θ_2=0

- Simulation-based asymptotic bias calculations for binary response data, as a function of various components of the specification:



Summary

- If any component of $\boldsymbol{X}_{ki}$ is an exogenous time-dependent covariate, interest lies in the marginal mean, and estimation/inference is to be conducted via GEE then either:

  (1) the full covariate conditional mean assumption for the adopted model

  $$\mu_{ki} \;=\; \mathsf{E}[Y_{ki}|\; \boldsymbol{X}_{k1}, \ldots, \boldsymbol{X}_{kn_k}] \quad \forall\; i = 1, \ldots, n_k$$

  must hold if covariance weighting is to be used, or

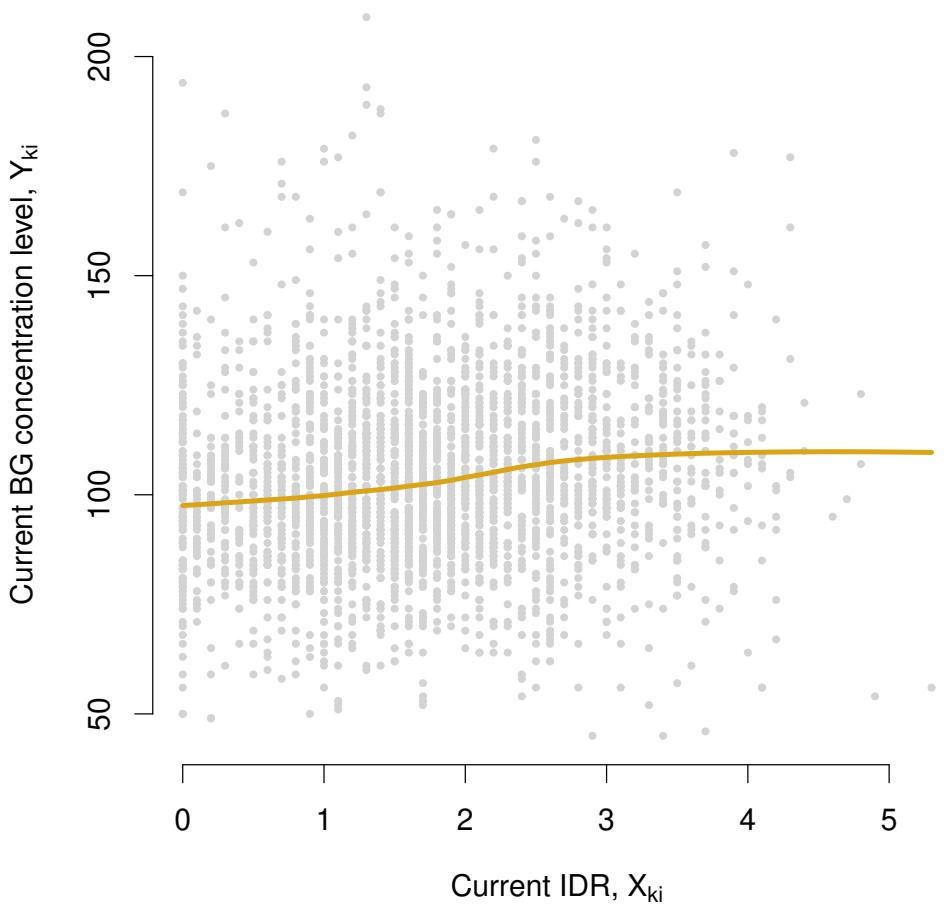
  (2) an independence working structure must be adopted

- The 'tension' between the assumption and choice of working dependence structure manifests as a bias-variance trade-off

- If any component of $\boldsymbol{X}_{ki}$ is an endogenous time-dependent covariate then time-dependent confounding may be an additional issue that needs to be addressed

Time-dependent confounding

- Recall from the initial EDA of the IDR/BG data from Vanderbilt that there was some evidence of a small positive association:



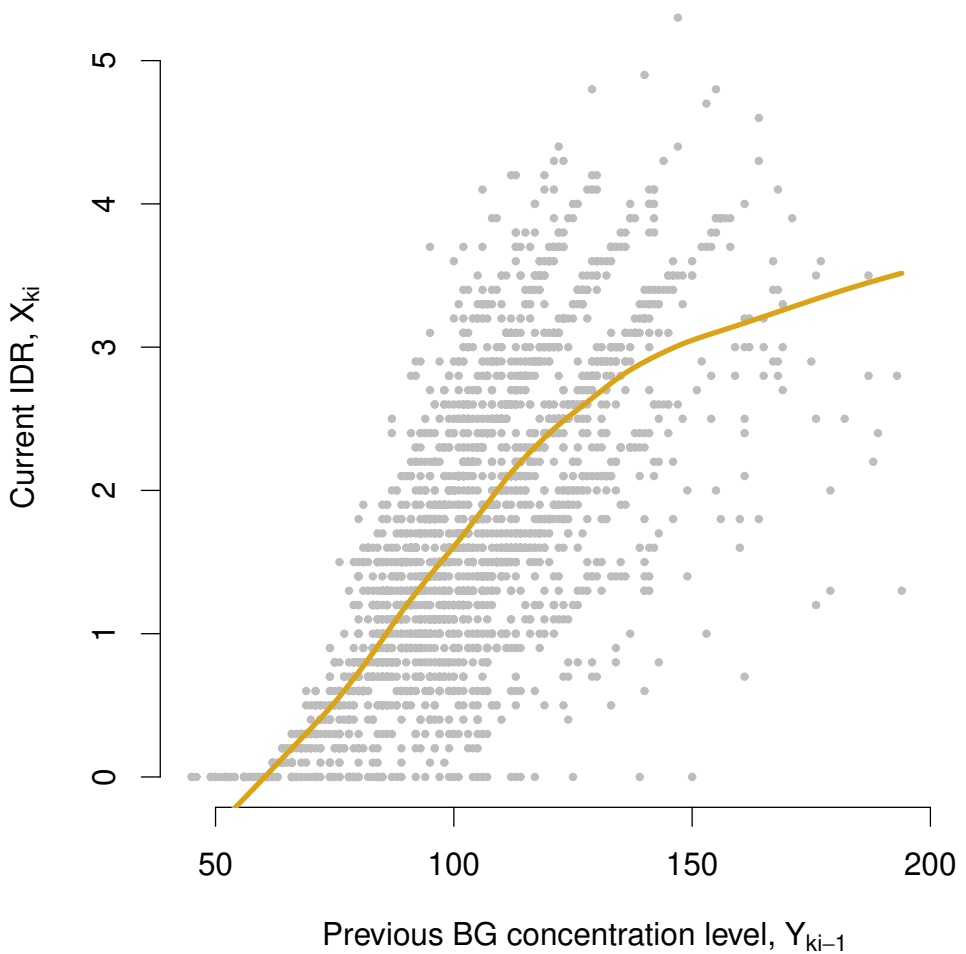
- An interesting feature of Vanderbilt ICU was that it employed a locally-developed computerized decision support systems to help guide decision-making

- Towards improving the control of BG, this involved:
  - ⋆ measuring BG on average every 2 hours
  - ⋆ adjusting IDR in order to keep the glucose concentrations low
    - ∗ most recent BG measurement was used to determine, in part, the next insulin dose

- Consequently there is a feedback mechanism between BG levels and IDR over time
  - ⋆ previous response is a determinant of current exposure

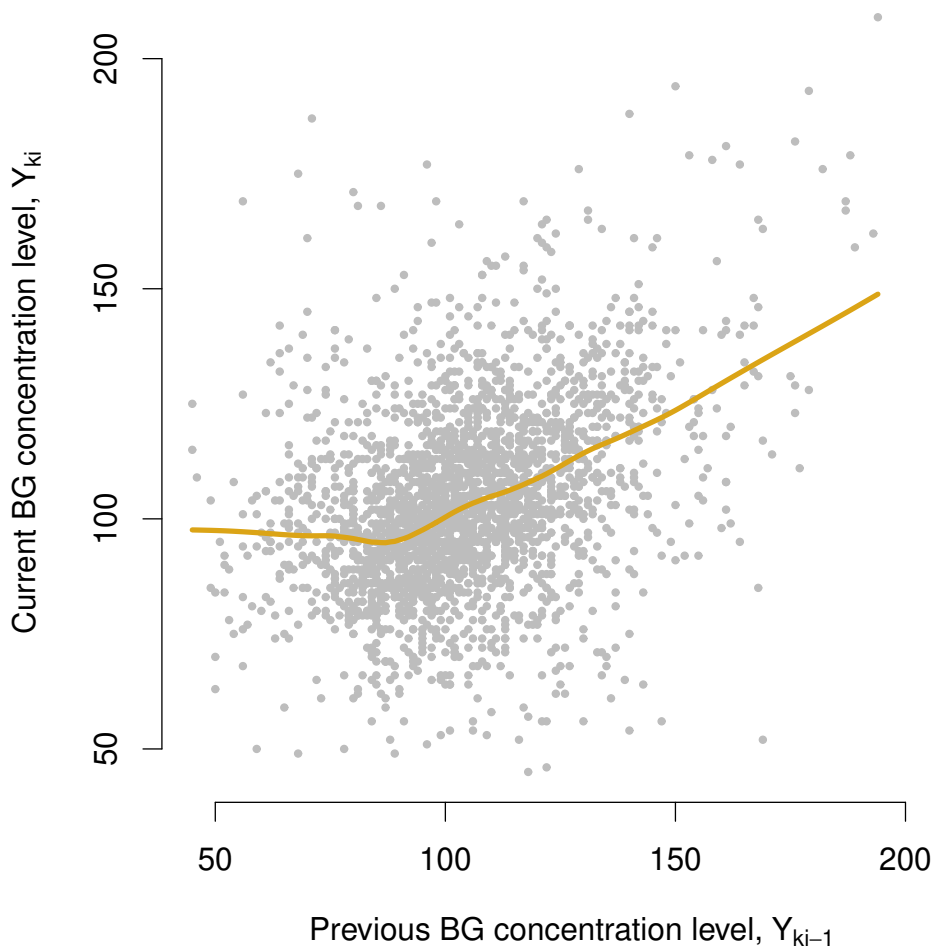
- Intuitively, it also makes sense that previous responses are predictive of current response
  - ⋆ i.e. within-subject correlation in the responses over time

- Graphically represent the relationship between **previous** BG level and **current** IDR:



⋆ suggestive of a strong association

- Graphically represent the relationship between **previous** BG level and **current** BG level:



⋆ also suggestive of a strong association

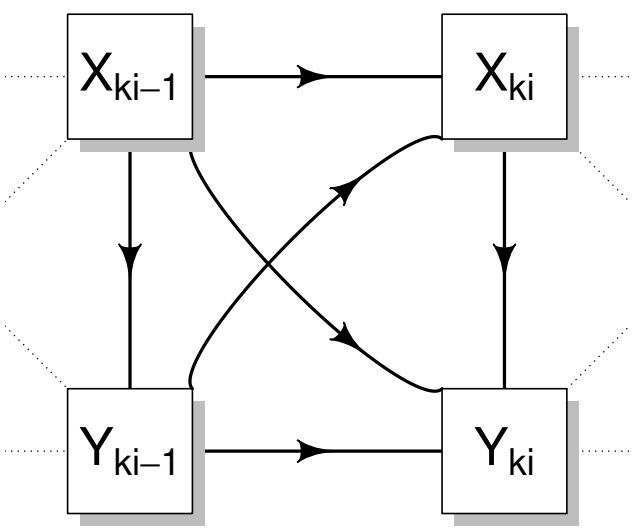
X_k determined by Y_k-1

- The scatterplots provide compelling evidence that IDR is an endogenous time-dependent covariate

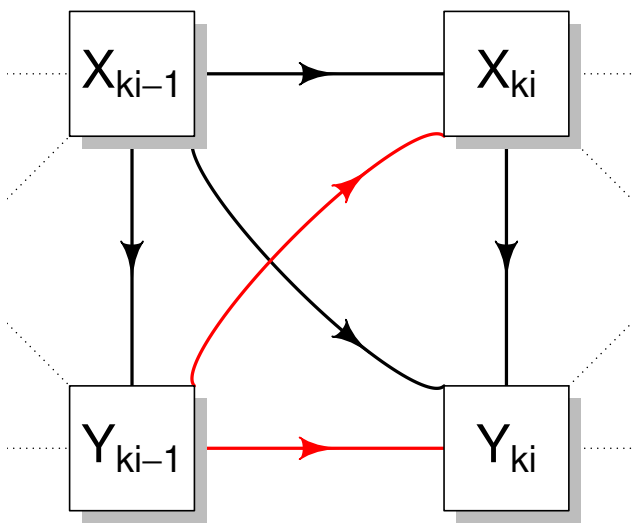
- To help understand the implications of this, suppose interest lies with the distributed lag model:

$$\mathsf{E}[Y_{ki}|\ \boldsymbol{X}_k] \ = \ \beta_0 \ + \ \beta_1 X_{ki} \ + \ \beta_2 X_{k,i-1}$$

and consider the following directed acyclic graph (DAG):



- From the DAG, we see that $Y_{k,i-1}$ satisfies the 'usual' criteria for being a confounder
  - ⋆ i.e. $Y_{k,i-1}$ is related to both $X_{k,i}$ and $Y_{k,i}$

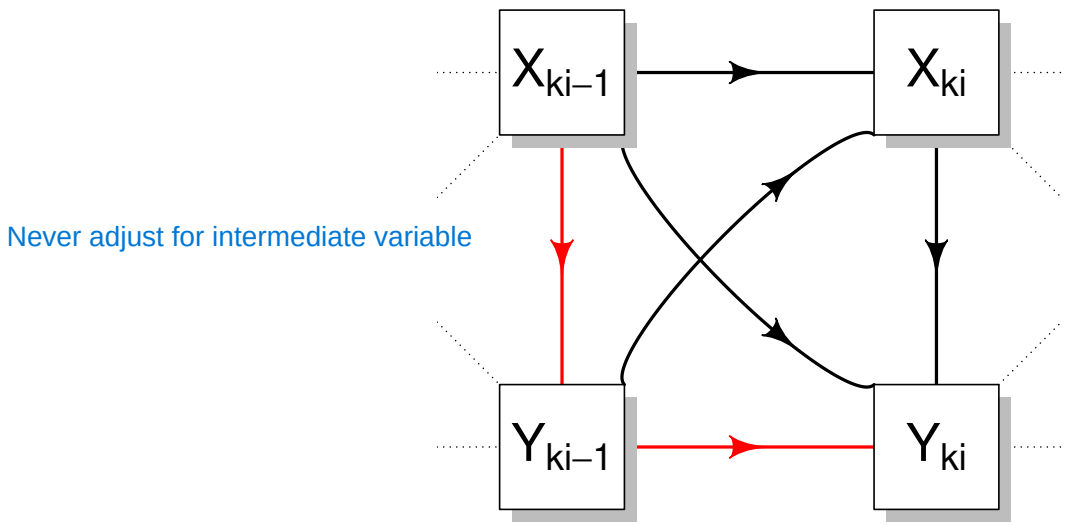


- One might, therefore, reasonably conclude that $Y_{k,i-1}$ should be included in the model
  - ⋆ i.e. to control the 'confounding' effect of $Y_{k,i-1}$, one should fit the model:

WRONG! $\mathsf{E}[Y_{ki}|\ \boldsymbol{X}_k] \ = \ \beta_0 \ + \ \beta_1 X_{ki} \ + \ \beta_2 X_{k,i-1} \ + \ \beta_3 Y_{k,i-1}$

- We also see, however, that $Y_{k,i-1}$ is as an intermediary variable in the relationship between $X_{k,i-1}$ and Y_{ki}

Mediator



- Consequently, including $Y_{k,i-1}$ in the model will partially 'block' the effect of $X_{k,i-1}$ on Y_{ki}
  - ⋆ estimates won't reflect the 'true' nature of the relationship between exposure and the response

- In the presence of time-dependent confounding, therefore, remaining within the 'standard' regression modeling framework will result in bias regardless of whether or not $Y_{k,i-1}$ is included in the model

- To resolve this dilemma, we need to turn to more formal methods in causal inference:
  - ⋆ define an appropriate set of counterfactual outcomes
  - ⋆ select the contrast that is of interest
  - ⋆ use an appropriate method for estimation

- Three general frameworks have been proposed:
  (1) the g-computation algorithm formula
  (2) inverse-probability of treatment weighting (IPTW) for marginal structural models
  (3) g-estimation of structural nested models

- We are going to focus on (1) and (2)

- Moving forward we are also going to focus attention (for the most part) to the relatively simple setting where:
  - ⋆ the data are collected at $K+1$ fixed equally-spaced intervals
  - ⋆ censoring, missed visits and measurement error are absent
  - ⋆ the response is a continuous random variable Y, measured at the end of follow-up
  - ⋆ there is no unmeasured confounding
- Also going to adopt a slightly different notation
  - ⋆ conform to much of the causal inference literature
- Review the methodology for:
  - ⋆ fixed exposures (FE)
  - ⋆ static time-dependent exposures (STDE)
- Won't be covering dynamic time-dependent exposures

- Much of the material is from the book: *Causal Inference* by Hernán MA, Robins JM.
  - ⋆ The pdf version is available online
  - ⋆ substantially more detail
  - ⋆ extensions beyond the simple setting(s) we consider
  - ⋆ comprehensive references

FE: Counterfactuals and causal contrasts

- Suppose interest lies in some fixed exposure A
  - ⋆ i.e. A is time-invariant

- Let $Y_{(a)}$ denote the *counterfactual outcome* corresponding to exposure a
  - ⋆ value of Y had, possibly contrary to fact, exposure been set to $A = a$
  - ⋆ if A is binary then there are two potential outcomes (i.e. $Y_{(0)}$ and $Y_{(1)}$)
  - ⋆ if A is continuous then there are infinitely many potential outcomes

- Given a set of counterfactual outcomes, one can define 'causation' in terms of a comparison between their distributions

- Typically the comparison is on the basis of some some specific feature or functional of the distributions of the counterfactuals

- For example, one could choose to perform the comparison on the basis of the mean, $\mathsf{E}[Y_{(a)}]$
  - ⋆ interpreted as mean value of Y for the (possibly hypothetical) scenario where everyone receives exposure $A = a$
  - ⋆ could, in principle, choose any functional

- A comparison between $\mathsf{E}[Y_{(1)}]$ and $\mathsf{E}[Y_{(0)}]$, therefore, describes a comparison between two (possibly hypothetical) exposure scenarios
  - ⋆ e.g. the causal risk difference: $\text{RD} = \mathsf{E}[Y_{(1)}] - \mathsf{E}[Y_{(0)}]$

- Note, outside of the context of a continuous response, other contrasts may be of interest
  - ⋆ e.g. the causal risk ratio: $\text{RR} = \mathsf{E}[Y_{(1)}]/\mathsf{E}[Y_{(0)}]$
  - ⋆ e.g. the causal odds ratio: $\text{OR} = \dfrac{\mathsf{E}[Y_{(1)}]/(1 - \mathsf{E}[Y_{(1)}])}{\mathsf{E}[Y_{(0)}]/(1 - \mathsf{E}[Y_{(0)}])}$

- Given some contrast(s) of interest, all one has to do is estimate $\mathsf{E}[Y_{(a)}]$

FE: Estimation based on a randomized study

- Focusing on a binary A, suppose $2n$ subjects were into a randomized trial with half allocated to the $A = 0$ arm and half to the $A = 1$ arm:

$A = 0$	$A = 1$
$Y_{0,1}$	$Y_{1,1}$
$\vdots$	$\vdots$
$Y_{0,n}$	$Y_{1,n}$

- Estimate $\mathsf{E}[Y_{(a)}]$ with the empirical mean, $\overline{Y}_a$
  - ⋆ the mean of Y among individuals who *actually* received $A = a$
- Plug-in these values to estimate some causal contrast of interest
  - ⋆ e.g. estimate the causal risk difference by

$$\widehat{\mathsf{RD}} \;=\; \overline{Y}_1 - \overline{Y}_0$$

Q: Under what circumstances can we interpret such estimates as representing causation?

- ⋆ intuitively we appeal to randomization but what is the formal justification for these plug-in estimators?

- Consider the following assumptions:

(i) *Consistency*:

$$\boxed{\text{if } A = a \text{ then } Y_{(a)} = Y}$$

- ∗ guarantees that what we observe when a specific exposure is allocated is the corresponding counterfactual

(ii) *Exchangeability*:

$$\boxed{Y_{(a)} \perp\!\!\!\perp A, \ \forall\ a}$$

- ∗ exposure allocation is independent of the potential outcomes
- ∗ i.e. exposure was not allocated on the basis of what the outcomes might be under various scenarios for treatment

- Note, exchangeability is an assumption regarding the joint distribution of the random variables
  - ⋆ may or may not hold for a particular realization from the joint distribution
  - ⋆ as such, it is typical to check for imbalance across treatment arms that is due to chance

- Under assumptions (i) and (ii) we have that

Mirror histogram: Check covariate balance, asses how good is the randomization in RCT

$$\begin{aligned} \mathsf{E}[Y_{(a)}] &= \mathsf{E}[Y_{(a)} |\ A] \\ &= \mathsf{E}[Y_{(a)} |\ A = a] \\ &= \mathsf{E}[Y |\ A = a] \end{aligned}$$

- Since randomization guarantees exchangeability, assuming consistency holds $\mathsf{E}[Y_{(a)}]$ can be estimated using the observed data by evaluating the empirical mean among those individuals given treatment $A = a$:

$$\widehat{\mathsf{E}[Y_{(a)}]} = \overline{Y}_a$$

FE: Estimation based on observational data

- In the absence of a randomized study, we have to appeal to observational data to answer question(s) of interest
  - ⋆ data collected as part of a specific research study
  - ⋆ data collected for some other (primary) purpose

- The central (additional) challenge is that there may be factors that are simultaneously predictive of the response and the exposure so that the exchangeability assumption is violated:
  - ⋆ i.e. $Y_{(a)} \not\perp\!\!\!\perp A$
  - ⋆ results in *confounding bias*

- To estimate causal associations from observational data we therefore need an extended set of identifying assumptions

- Identifiability assumptions:

(i) *Consistency*:

$$\boxed{\text{if } A = a \text{ then } Y_{(a)} = Y}$$

(ii) *Conditional exchangeability*:

$$\boxed{Y_{(a)} \perp\!\!\!\perp A | \ \boldsymbol{L} = \boldsymbol{l}, \ \forall \ a, \boldsymbol{l}}$$

* i.e. $\exists$ some vector $\boldsymbol{L}$ such that exchangeability holds within each level
* intuitively, exposure allocation is random within levels of $\boldsymbol{L}$
* referred to as the assumption of *no unmeasured confounding*

(iii) *Positivity*:

$$\boxed{\text{if } f_{\boldsymbol{L}}(\boldsymbol{l}) > 0 \text{ then } f_{A|\boldsymbol{L}}(a|\boldsymbol{l}) > 0, \ \forall \ a}$$

⋆ all exposure 'choices' are available to all sub-populations defined by $\boldsymbol{L}$

⋆ exposure allocation is not deterministic for some individuals

FE: g-formula

- Building on the previous development, if assumptions (i)-(iii) hold then:

$$\begin{aligned} \mathsf{E}[Y_{(a)}] \;=\; \mathsf{E}[\mathsf{E}[Y_{(a)}|\boldsymbol{L}]] \;&=\; \int \mathsf{E}[Y_{(a)}|\boldsymbol{L}=\boldsymbol{l}] f_{\boldsymbol{L}}(\boldsymbol{l}) \partial \boldsymbol{l} \\ &= \int \mathsf{E}[Y|A=a,\boldsymbol{L}=\boldsymbol{l}] f_{\boldsymbol{L}}(\boldsymbol{l}) \partial \boldsymbol{l} \end{aligned}$$

  or, if the elements of $\boldsymbol{L}$ are all categorical, then:

$$\mathsf{E}[Y_{(a)}] \;=\; \sum_{\boldsymbol{l}} \mathsf{E}[Y|A=a,\boldsymbol{L}=\boldsymbol{l}] P(\boldsymbol{L}=\boldsymbol{l})$$

  - ⋆ the average of $\mathsf{E}[Y|A=a,\boldsymbol{L}=\boldsymbol{l}]$ over the distribution of $\boldsymbol{L}$
  - ⋆ referred to as the *g-formula* for $\mathsf{E}[Y_{(a)}]$

- Crucially, $\mathsf{E}[Y|A,\boldsymbol{L}]$ is estimable from the observed data

- In practice, we specify a 'working' outcome regression model
  - ⋆ e.g. linear regression for a continuous response

$$\mathsf{E}[Y|A, \boldsymbol{L}] \;=\; \beta_0 \;+\; \beta_1 A \;+\; \beta_2 L_1 \;+\; \ldots\; \beta_{K+1} L_K$$

  - ⋆ e.g. logistic regression for a binary response

$$\text{logit}\; \mathsf{E}[Y|A, \boldsymbol{L}] \;=\; \beta_0 \;+\; \beta_1 A \;+\; \beta_2 L_1 \;+\; \ldots\; \beta_{K+1} L_K$$

- Whatever the choice of working model, (somehow) obtain estimates $\widehat{\boldsymbol{\beta}}$
  - ⋆ using a random sample $\{(Y_i, A_i, \boldsymbol{L}_i); i = 1, \ldots, n\}$
- At this point we could stop and report $\hat{\beta}_1$, the estimated coefficient associated with A
  - ⋆ the 'unconfounded' conditional association
- The goal, however, is to estimate $\mathsf{E}[Y_{(a)}]$!

- Towards this goal, for each value of $A = a$, calculate the fitted value of the response from the working regression model for the i^{th} individual
  - ⋆ e.g. for the linear working model calculate

$$\widehat{\mu}_i(a) \;=\; \hat{\beta}_0 \;+\; \hat{\beta}_1[A = a] \;+\; \hat{\beta}_2 L_{i,1} \;+\; \ldots\; \hat{\beta}_{K+1} L_{i,K}$$

  - ⋆ this value of a may or may not be what was actually observed for the i^{th} individual

- We can then empirically estimate $\mathsf{E}[Y_{(a)}]$ by

$$\widehat{\mathsf{E}}_{\mathsf{OR}}[Y_{(a)}] \;=\; \frac{1}{n}\sum_{i=1}^{n} \widehat{\mu}_i(a)$$

  - ⋆ sometimes referred to as the *outcome regression estimator*

- Intuitively, the estimator works by asking what we would see if everyone had been exposed to $A = a$
  - ⋆ regardless of what they actually received

FE: IPTW

- A second estimator of $\mathsf{E}[Y_{(a)}]$ is given by

$$\widehat{\mathsf{E}}_{\mathsf{IP}}[Y_{(a)}] \;=\; \frac{\sum_{i=1}^{n} I\{A_i = a\} W_i Y_i}{\sum_{i=1}^{n} I\{A_i = a\} W_i}$$

  where $I\{A_i = a\}$ is an indicator of whether or not the i^{th} individual received treatment a and

$$W_i \;=\; \frac{1}{f(A_i|\boldsymbol{L}_i)}$$

  is the conditional pdf of A given $\boldsymbol{L}$ evaluated at the subjects own values

  - ⋆ i.e. the exposure they actually receive

- $\widehat{\mathsf{E}}_{\mathsf{IP}}[Y_{(a)}]$ is a weighted average of responses among those individuals who actually received $A = a$

- Intuitively, weighting serves to create a pseudo-population within which treatment allocation is independent of $\boldsymbol{L}$

Q: If we have changed the population, however, are we estimating the right quantity?

- Yes, because:

$$\begin{aligned}
\mathsf{E}[I\{A=a\}WY] &= \mathsf{E}[I\{A=a\}WY_{(a)}] \\
&= \mathsf{E}[\mathsf{E}[I\{A=a\}WY_{(a)}|\boldsymbol{L}]] \\
&= \mathsf{E}[\underbrace{\mathsf{E}[I\{A=a\}W|\boldsymbol{L}]}_{=1}\mathsf{E}[Y_{(a)}|\boldsymbol{L}]] \\
&= \mathsf{E}[\mathsf{E}[Y_{(a)}|\boldsymbol{L}]] \\
&= \mathsf{E}[Y_{(a)}]
\end{aligned}$$

- Note, the denominator in the empirical version of the IPTW estimator serves to ensure that the size of the pseudo-population is the same as the size of the original population

- In practice, as with the outcome regression estimator, we need to choose a working model for the weights

- For a binary exposure variable, we usually adopt some logistic regression:

$$\text{logit } P(A = 1|\boldsymbol{L} = \boldsymbol{l}) \;=\; \alpha_0 \;+\; \alpha_1 L_1 \;+\; \ldots \;+\; \alpha_K L_K$$

  ⋆ referred to as the model for the *propensity score*

- Given an estimate of $\boldsymbol{\alpha}$, say $\widehat{\boldsymbol{\alpha}}$, fitted values can then be calculated for each individual depending on the exposure they actually received:

$$\text{if } A_i = 0 \text{ then } W_i^{-1} \;=\; \frac{1}{1 + \exp\{\boldsymbol{L}_i^T \widehat{\boldsymbol{\alpha}}\}}$$

$$\text{if } A_i = 1 \text{ then } W_i^{-1} \;=\; \frac{\exp\{\boldsymbol{L}_i^T \widehat{\boldsymbol{\alpha}}\}}{1 + \exp\{\boldsymbol{L}_i^T \widehat{\boldsymbol{\alpha}}\}}$$

- Finally, note that since $\mathsf{E}[I\{A=a\}W]=1$, we can also write:

$$\frac{\mathsf{E}[I\{A=a\}WY]}{\mathsf{E}[I\{A=a\}W]} \;=\; \mathsf{E}[Y_{(a)}]$$

- Building on this, it turns out that for any function $g(A)$ that is not a function of $\boldsymbol{L}$ one can show that:

$$\frac{\mathsf{E}[I\{A=a\}WYg(A)]}{\mathsf{E}[I\{A=a\}Wg(A)]} \;=\; \mathsf{E}[Y_{(a)}]$$

  - ⋆ see Hernan and Robins (*Journal of Epidemiology and Community Health*, 2006)

- Consequently the introduction of the function $g(A)$ serves to define a class of consistent estimators of $\mathsf{E}[Y_{(a)}]$
  - ⋆ choice of $g(\cdot)$ solely has an impact on efficiency

- If we take $g(A)$ to be the marginal density for A, for example, we obtain the so-called *stabilized IPTW* estimator:

$$\widehat{\mathsf{E}}_{\mathsf{SIP}}[Y_{(a)}] \;=\; \frac{\sum_{i=1}^{n} I\{A_i = a\} SW_i Y_i}{\sum_{i=1}^{n} I\{A_i = a\} SW_i}$$

where

$$SW_i \;=\; \frac{f(A_i)}{f(A_i|\boldsymbol{L}_i)}$$

  - ⋆ efficiency gains are achieved because SW_i will, in general, be far less variable than W_i
  - ⋆ depending on the context, working models might be adopted for both the numerator and denominator of SW_i

STDE: Counterfactuals and causal contrasts

- Moving beyond a fixed exposure, suppose interest lies in some time-dependent binary exposure $A(t)$:
  - $\star$ $A(0)$ represents the baseline measurement
  - $\star$ K follow-up measurements: $A(1), \ldots, A(K)$
- In addition, let $\boldsymbol{L}(t)$ denote a vector of possibly time-dependent covariates
  - $\star$ measured just before exposure at time t is assigned/measured
- Finally, let Y denote a continuous response of interest, measured at the end of follow-up time
- Notationally, it will be convenient to let $\overline{A}(t)$ and $\overline{\boldsymbol{L}}(t)$ denote the history of $A(t)$ and $\boldsymbol{L}(t)$, respectively, up to and including time t

- Towards defining counterfactuals in this setting, let

$$\overline{a} \;=\; \{a(0),\; a(1), \ldots,\; a(K)\}$$

  denote some exposure sequence or regime
  - ⋆ $a(t) = 1$ if the subject is exposed at time t
  - ⋆ 2^{K+1} possible regimes

- Example regimes include:
  - ⋆ continuous exposure: $\{1, 1, 1, \ldots, 1\}$
  - ⋆ no exposure: $\{0, 0, 0, \ldots, 0\}$
  - ⋆ exposure solely during the first two time periods: $\{1, 1, 0, \ldots, 0\}$

- Finally, let $\overline{a}(t)$ represent the exposure history under regime $\overline{a}$ through to time t

- Associated with each $\overline{a}$ is a counterfactual outcome denoted $Y_{(\overline{a})}$
  - ⋆ each subject has 2^{K+1} potential outcomes

- As with fixed exposures, 'causation' can be defined in terms of comparisons between the distributions of $Y_{(\overline{a})}$ under different exposure regimes

- Typically, focus is on the mean so that, in the current context, we'd say that the time-varying exposure $A(t)$ has a causal effect on the average of Y if

$$\mathsf{E}[Y_{(\overline{a})}] \neq \mathsf{E}[Y_{(\overline{a}')}]$$

  for at least two regimes $\overline{a}$ and $\overline{a}'$

- In practice, we are seldom interested in all possible comparisons among the regimes
  - ⋆ only a few will be of interest in the 'real world'

- Whichever regimes are of interest, we need a strategy for estimating $\mathsf{E}[Y_{(\overline{a})}]$ given data from an observational study
  - ⋆ generalized identifiability assumptions
  - ⋆ estimators

STDE: Generalized identifiability assumptions

(i) *Consistency*:

$$\text{if } \overline{A} = \overline{a} \text{ then } Y_{(\overline{a})} = Y$$

(ii) *Conditional exchangeability*:

$$Y_{(\overline{a})} \perp\!\!\!\perp A(t) |\ \overline{A}(t-1) = \overline{a}(t-1), \overline{\boldsymbol{L}}(t) = \overline{\boldsymbol{l}}(t), \qquad \forall\ \overline{a}, \boldsymbol{l}(t)$$

Conditional on everything observed so far. Sequential randomization

(iii) *Positivity*:

$$\text{if}$$
$$f_{\overline{A}(t-1),\, \overline{\boldsymbol{L}}(t)}\{\overline{a}(t-1), \overline{\boldsymbol{l}}(t)\} > 0$$
$$\text{then}$$
$$f_{\overline{A}(t)|\ \overline{A}(t-1),\, \overline{\boldsymbol{L}}(t)}\{a(t)|\ \overline{a}(t-1), \overline{\boldsymbol{l}}(t)\} > 0,\ \forall\ \overline{a}(t)$$

STDE: g-formula

- Under assumptions (i)-(iii), $\mathsf{E}[Y_{(\overline{a})}]$ can be estimated using data from an observational study based on a generalization of the g-formula that accommodates time-dependent confounders
  - ⋆ Robins, Greenland and Hu (JASA, 1999)

- For example, if all elements of $\boldsymbol{L}(t)$ are categorical, we can write:

$$\mathsf{E}[Y_{(\overline{a})}] = \sum_{\overline{\boldsymbol{l}}} \left\{ \overbrace{\mathsf{E}[Y|\overline{A}=\overline{a},\overline{\boldsymbol{L}}=\overline{\boldsymbol{l}}]}^{\text{Regression on history}} \times \overbrace{\prod_{k=0}^{K} f(\boldsymbol{l}(k)|\ \overline{a}(k-1),\ \overline{\boldsymbol{l}}(k-1))}^{\text{Conditional density factorization}} \right\}$$

- In practice, since neither $\mathsf{E}[Y|\overline{A}=\overline{a},\overline{\boldsymbol{L}}=\overline{\boldsymbol{l}}]$ nor $f(\boldsymbol{l}(k)|\ \overline{a}(k-1),\ \overline{\boldsymbol{l}}(k-1))$ are known they must be estimated
  - ⋆ in most settings, this will require fitting some parsimonious parametric working model
  - ⋆ bias will arise, however, if either model is misspecified

STDE: IPTW

- We can also extend the IPTW formula to estimate $\mathsf{E}[Y_{(\overline{a})}]$ as the weighted average of the observed response among those subjects with $\overline{A} = \overline{a}$:

$$\widehat{\mathsf{E}}_{\mathsf{SIP}}[Y_{(\overline{a})}] = \frac{\sum_{i=1}^{n} I\{\overline{A}_i = \overline{a}\} SW_i Y_i}{\sum_{i=1}^{n} I\{\overline{A}_i = \overline{a}\} SW_i}$$

  where the subject-specific weights are:

Compute weight at each time point

$$SW_i = \prod_{k=0}^{K} \frac{f(A_i(k)|\ \overline{A}_i(k-1))}{f(A_i(k)|\overline{A}_i(k-1), \overline{\boldsymbol{L}}_i(k))}$$

- May also consider unstabilized IPTW for which the weights are:

$$W_i = \prod_{k=0}^{K} \frac{1}{f(A_i(k)|\overline{A}_i(k-1), \overline{\boldsymbol{L}}_i(k))}$$

- As in the fixed effects context, a working model will need to be specified for $f(A_i(k)|\overline{A}(k-1), \overline{\boldsymbol{L}}_i(k))$ and, if necessary, $f(A_i(k)|\overline{A}(k-1))$
  - ⋆ e.g., a logistic regression if $A(t)$ is binary
  - ⋆ bias will arise, however, if the model for $f(A_i(k)|\overline{A}(k-1), \overline{\boldsymbol{L}}_i(k))$ is misspecified although not if $f(A_i(k)|\overline{A}(k-1))$ is misspecified

Marginal structural models (MSMs)

- When K is large, it may not be possible to (reasonably) estimate all 2^{K+1} counterfactual means, $\mathsf{E}[Y_{(\overline{a})}]$
  - ⋆ especially when the data are not extensive

- An alternative strategy is to posit a model for how $\mathsf{E}[Y_{(\overline{a})}]$ varies as a function of $\overline{a}$, such as:

$$\mathsf{E}[Y_{(\overline{a})}] \;=\; \beta_0 \;+\; \beta_1 \mathsf{cumsum}(\overline{a})$$

- Referred to as *marginal structural models* because:
  - (i) they model the *marginal* distribution of the counterfactual outcomes
  - (ii) models for counterfactuals are often referred to as *structural*

- Appealing because they can easily be extended to permit flexible functions of $\overline{a}$ and/or effect modification:
  - ⋆ e.g.,

$$\mathsf{E}[Y_{(\overline{a})}] = \beta_0 + \beta_1 \mathsf{cumsum}(\overline{a}) + \beta_2 \mathsf{Gender} + \beta_3 \mathsf{Gender} \times \mathsf{cumsum}(\overline{a})$$

- Estimation of the (causal) parameters in an MSM can proceed via IPTW
  - ⋆ Robins, Hernan and Brumback (*Epidemiology*, 2000)
  - ⋆ use either standard weights or stabilized weights
  - ⋆ consistency requires the denominators in each to be correctly specified

- Finally, it's also worth noting that MSMs are also useful in settings where the exposure is (in principle) continuous
  - ⋆ e.g., if one is interested in a dose-response relationship

Uncertainty

- So far we have only focused on estimating causal parameters
- Depending on the context analytic expressions for (asymptotic) variances can be derived
  - ⋆ valid inference requires consideration of uncertainty due to estimation of the components of the working models
- Most of the time, however, it seems that folks propose the bootstrap as a means to quantifying uncertainty

Vanderbilt ICU data

- Recall the six models we considered for investigating the association between IDR and BG:

Cross-sectional mean models:

$$Y_{ki} = \beta_0 + \boldsymbol{\beta}_* \boldsymbol{X}^*_{ki} + \beta_x X_{ki} + \epsilon_{ki} \quad \text{(1-CS)}$$

$$Z_{ki} = \beta_0 + \boldsymbol{\beta}_* \boldsymbol{X}^*_{ki} + \beta_x X_{ki} + \epsilon_{ki} \quad \text{(2-CS)}$$

Distributed lag models:

$$Y_{ki} = \beta_0 + \boldsymbol{\beta}_* \boldsymbol{X}^*_{ki} + \beta_x X_{ki} + \beta_{x,l} X_{k,i-1} + \epsilon_{ki} \quad \text{(1-DL)}$$

$$Z_{ki} = \beta_0 + \boldsymbol{\beta}_* \boldsymbol{X}^*_{ki} + \beta_x X_{ki} + \beta_{x,l} X_{k,i-1} + \epsilon_{ki} \quad \text{(2-DL)}$$

Transition models:

$$Y_{ki} = \beta_0 + \boldsymbol{\beta}_* \boldsymbol{X}^*_{ki} + \beta_x X_{ki} + \beta_{y,l} Y_{k,i-1} + \epsilon_{ki} \quad \text{(1-TN)}$$

$$Z_{ki} = \beta_0 + \boldsymbol{\beta}_* \boldsymbol{X}^*_{ki} + \beta_x X_{ki} + \beta_{y,l} Y_{k,i-1} + \epsilon_{ki} \quad \text{(2-TN)}$$

- Data analysis taken from Schildcrout et al (*Statistics in Medicine*, 2011)

- Focus on results based on:

  (1) unweighted working independence GEE

  (2) IPTW based on a linear regression for IDR:

$$X_{ki} \sim Y_{k,i-1} + \boldsymbol{W}_{k,i-1}$$

  where $\boldsymbol{W}_{k,i-1}$ is a collection of other potential confounders

  $*$ referred to as the *lagged glucose value strategy* (LVS)

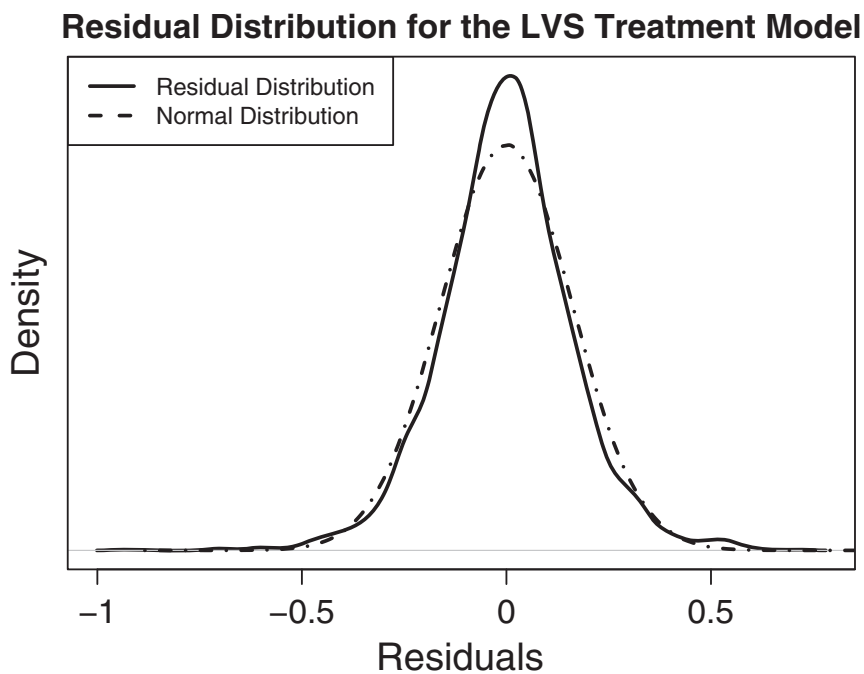
- For (2), a Normal distribution was used as the predictive distribution for $f()$ in the stabilized weights:

$$SW_{ki} = \frac{f(X_{ki}|\ \boldsymbol{W}_{k,i-1})}{f(X_{ki}|\ Y_{k,i-1}, \boldsymbol{W}_{k,i-1})}$$

  $\star$ truncated at 50 to avoid highly inefficient estimates (Cole and Hernan, *AJE*, 2008)

- Since consistency of the IPTW estimator requires the working model in the denominator of SW_{ki} to be correctly specified, it should be checked (to the extent possible)

- Kernel density plot of residuals from the linear regression for IDR:



  ⋆ assuming the mean model is correctly specified, the choice of a Normal distribution seems reasonable

- Point estimates adjusted for a range of potential confounders:

Model/covariate	**BG level, Y_{ki}**		**BG change, Z_{ki}**	
	WI-GEE	IPTW	WI-GEE	IPTW
Cross-sectional				
IDR, β_x	6.3	-23.8	-32.6	-22.7
Distributed lag				
IDR, β_x	9.7	-23.2	-42.9	-20.4
lagged IDR, $\beta_{x,l}$	-5.3	-1.0	16.0	-4.1
Transition				
IDR, β_x	-17.6	-23.1	-17.6	-23.1
lagged BG, $\beta_{y,l}$	0.6	0.7	-0.4	-0.3

  - ⋆ seems clear that time-dependent confounding had a substantial impact
    - ∗ WI-GEE estimates for the cross-sectional and distributed lag models for Y_{ki} are not biologically plausible
  - ⋆ also considerably greater consistency among the IPTW estimates
  - ⋆ IPTW in the transition model 'recovers' some of the blocked path

Summary

- In many longitudinal settings the exposure of interest will be time-dependent
- This gives researchers, for better or worse (!), substantial flexibility in specification of the model
  - ⋆ permits a broad range of scientific questions that one can address
- However, additional issues/complexities, beyond those we typically worry about, arise and have to be considered:

(i) the nature/type of the time-dependent covariate
  - ∗ exogenous vs endogenous

(ii) the choice of working dependence structure
  - ∗ Pepe and Anderson (1994)

(iii) the control of potential time-dependent confounders
  - ∗ motivates the use of formal methods in causal inference