Approaches to deal with confounding

Alternatives to traditional regression adjustment

ERHS 732

Conditional models

- The simplest regression models we have typically fit for an exposure of interest X and outcome Y have been of the form
$$E[Y|X] = \beta_0 + \beta_1 X$$

Conditional models

- The simplest regression models we have typically fit for an exposure of interest X and outcome Y have been of the form
$$E[Y|X] = \beta_0 + \beta_1 X$$
- However we typically want to adjust for potential confounding, so we typically include more variables in the model

$$E[Y|X, C_i] = \beta_0 + \beta_1 X + \gamma' C_i$$

Conditional models

- The simplest regression models we have typically fit for an exposure of interest X and outcome Y have been of the form
$$E[Y|X] = \beta_0 + \beta_1 X$$
- However we typically want to adjust for potential confounding, so we typically include more variables in the model

$$E[Y|X, C_i] = \beta_0 + \beta_1 X + \gamma' C_i$$

  - where γ' represents a vector of parameter coefficients for each of our covariates C_i

Conditional effect estimates

- Including these additional variables in the model is often sufficient to control for potential confounding by them

Conditional effect estimates

- Including these additional variables in the model is often sufficient to control for potential confounding by them
  - If the variables are predictors of the outcome (they should be) they will also typically improve model fit and help the model explain more of the variance in the outcome

Conditional effect estimates

- Including these additional variables in the model is often sufficient to control for potential confounding by them
  - If the variables are predictors of the outcome (they should be) they will also typically improve model fit and help the model explain more of the variance in the outcome
- The interpretation of the parameter coefficient for X from a model of $E[Y|X, C_i]$ is the effect estimate for a unit increase in X **conditional** on C_i

Conditional effect estimates

- Including these additional variables in the model is often sufficient to control for potential confounding by them
  - If the variables are predictors of the outcome (they should be) they will also typically improve model fit and help the model explain more of the variance in the outcome
- The interpretation of the parameter coefficient for X from a model of $E[Y|X, C_i]$ is the effect estimate for a unit increase in X **conditional** on C_i
  - In other words this is an estimate of the effect assuming all of C_i remain constant, or within levels of C_i

Conditional effect estimates

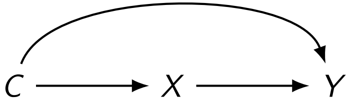
- Including these additional variables in the model is often sufficient to control for potential confounding by them
  - If the variables are predictors of the outcome (they should be) they will also typically improve model fit and help the model explain more of the variance in the outcome
- The interpretation of the parameter coefficient for X from a model of $E[Y|X, C_i]$ is the effect estimate for a unit increase in X **conditional** on C_i
  - In other words this is an estimate of the effect assuming all of C_i remain constant, or within levels of C_i
  - We could look at effect estimates in strata defined by C_i and take the weighted average across the strata

Conditioning on a covariate/confounder

- Including these additional variables in the model is often sufficient to control for potential confounding by them

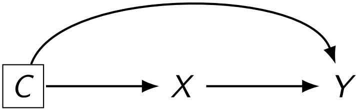
Conditioning on a covariate/confounder

- Including these additional variables in the model is often sufficient to control for potential confounding by them
- Including a variable in a model or stratifying on it essentially conditions on it ($E[Y|X, C]$)



Conditioning on a covariate/confounder

- Including these additional variables in the model is often sufficient to control for potential confounding by them
- Including a variable in a model or stratifying on it essentially conditions on it ($E[Y|X, C]$)



Conditional effect estimates

- Including several variables in the model however comes at a cost as they increase in number

Conditional effect estimates

- Including several variables in the model however comes at a cost as they increase in number
  - Each additional parameter will use a degree of freedom (or more), which maybe a problem if we have limited power

Conditional effect estimates

- Including several variables in the model however comes at a cost as they increase in number
  - Each additional parameter will use a degree of freedom (or more), which maybe a problem if we have limited power
  - These are also nuisance parameters, because we typically don't care about the individual effect estimates of covariates, we simply want to adjust for them, but the model has to estimate them anyway

Conditional effect estimates

- Including several variables in the model however comes at a cost as they increase in number
  - Each additional parameter will use a degree of freedom (or more), which maybe a problem if we have limited power
  - These are also nuisance parameters, because we typically don't care about the individual effect estimates of covariates, we simply want to adjust for them, but the model has to estimate them anyway
- Approaches that minimize the amount of parameters in the model, could therefore be advantageous

Propensity scores

- Propensity scores are an alternative method of covariate adjustment, relying on a participant's (or observation's) propensity to be exposed given covariate values

Propensity scores

- Propensity scores are an alternative method of covariate adjustment, relying on a participant's (or observation's) propensity to be exposed given covariate values
- Based on this score we can adjust for covariates by

Propensity scores

- Propensity scores are an alternative method of covariate adjustment, relying on a participant's (or observation's) propensity to be exposed given covariate values
- Based on this score we can adjust for covariates by
  - Stratifying based on the PS

Propensity scores

- Propensity scores are an alternative method of covariate adjustment, relying on a participant's (or observation's) propensity to be exposed given covariate values
- Based on this score we can adjust for covariates by
  - Stratifying based on the PS
  - Matching based on it

Propensity scores

- Propensity scores are an alternative method of covariate adjustment, relying on a participant's (or observation's) propensity to be exposed given covariate values
- Based on this score we can adjust for covariates by
  - Stratifying based on the PS
  - Matching based on it
  - Including it in a regression model

Propensity scores

- Propensity scores are an alternative method of covariate adjustment, relying on a participant's (or observation's) propensity to be exposed given covariate values
- Based on this score we can adjust for covariates by
  - Stratifying based on the PS
  - Matching based on it
  - Including it in a regression model
  - Weighting based on it (IPW is one option)

Propensity scores

- Propensity scores are an alternative method of covariate adjustment, relying on a participant's (or observation's) propensity to be exposed given covariate values
- Based on this score we can adjust for covariates by
  - Stratifying based on the PS
  - Matching based on it
  - Including it in a regression model
  - Weighting based on it (IPW is one option)
- Since we no longer have to worry about covariates given the PS this is a dimension reduction procedure

Propensity score methods

- Dimension reduction comes at a cost

Propensity score methods

- Dimension reduction comes at a cost
  - We have to rely on a model for the exposure in order to estimate the PS

Propensity score methods

- Dimension reduction comes at a cost
  - We have to rely on a model for the exposure in order to estimate the PS
  - The model has to be correctly specified in addition to the outcome model

Propensity score methods

- Dimension reduction comes at a cost
  - We have to rely on a model for the exposure in order to estimate the PS
  - The model has to be correctly specified in addition to the outcome model
- PS not well defined for polytomous and continuous exposures

Propensity score methods

- Dimension reduction comes at a cost
  - We have to rely on a model for the exposure in order to estimate the PS
  - The model has to be correctly specified in addition to the outcome model
- PS not well defined for polytomous and continuous exposures
- Other issues may also arise

Propensity score methods

- Dimension reduction comes at a cost
  - We have to rely on a model for the exposure in order to estimate the PS
  - The model has to be correctly specified in addition to the outcome model
- PS not well defined for polytomous and continuous exposures
- Other issues may also arise
- Have to weigh limitations against advantages in each case

Exposure-confounder feedback

- Simple confounding structures involving a common cause of exposure and outcome are not a major concern as long as the common cause is measured

Exposure-confounder feedback

- Simple confounding structures involving a common cause of exposure and outcome are not a major concern as long as the common cause is measured

$$\begin{tikzcd} C \arrow[r] \arrow[rr, bend left] & X \arrow[r] & Y \end{tikzcd}$$

Exposure-confounder feedback

- Simple confounding structures involving a common cause of exposure and outcome are not a major concern as long as the common cause is measured

$$\begin{tikzcd} \boxed{C} \arrow[r] \arrow[rr, bend left] & X \arrow[r] & Y \end{tikzcd}$$

Exposure-confounder feedback

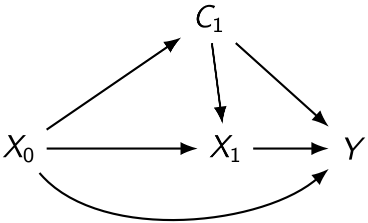
- However, in a time varying setting the relationship can be more complicated

Exposure-confounder feedback

- However, in a time varying setting the relationship can be more complicated
  - Confounder may cause the exposure, which may in turn cause the confounder at a later time-point

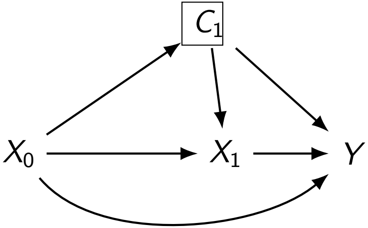
Exposure-confounder feedback

- However, in a time varying setting the relationship can be more complicated
  - Confounder may cause the exposure, which may in turn cause the confounder at a later time-point



Exposure-confounder feedback

- However, in a time varying setting the relationship can be more complicated
  - Confounder may cause the exposure, which may in turn cause the confounder at a later time-point



Exposure-confounder feedback

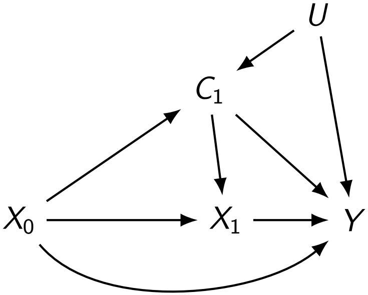
- Traditional regression approaches fail in this case

Exposure-confounder feedback

- Traditional regression approaches fail in this case
- IPW (and other methods) can be effective

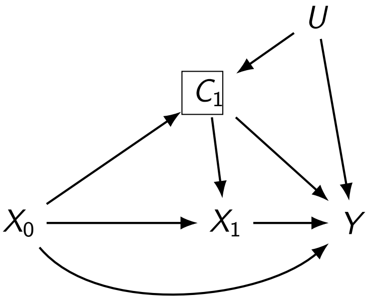
Exposure-confounder feedback

- Traditional regression approaches fail in this case
- IPW (and other methods) can be effective



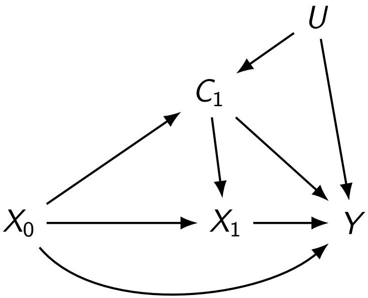
Exposure-confounder feedback

- Traditional regression approaches fail in this case
- IPW (and other methods) can be effective



Exposure-confounder feedback

- Traditional regression approaches fail in this case
- IPW (and other methods) can be effective



Exposure-confounder feedback

- Traditional regression approaches fail in this case
- IPW (and other methods) can be effective