Natural experiments:
innovative methodology in applied health promotion and chronic disease prevention research

Dr. Scott Leatherdale

CIHR-PHAC Chair in Applied Public Health Research

Associate Professor, School of Public Health and Health Systems

University of Waterloo

March 6th 2018

AAHB 18th Annual Scientific Meeting

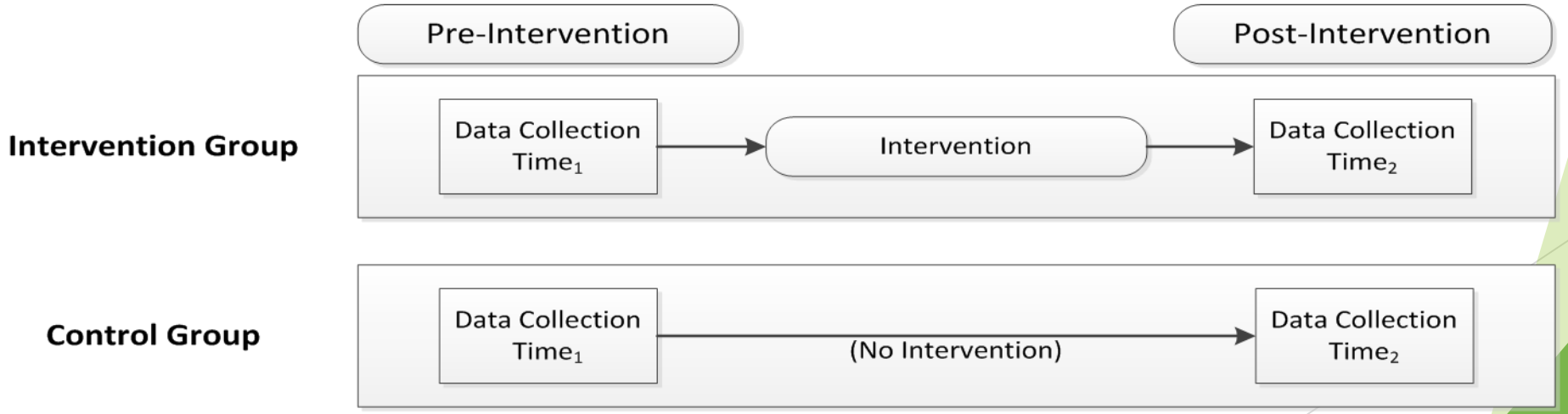
Official objectives of the workshop

- The workshop will enable *all* participants to:

1. Define natural experiments and provide concrete examples of some natural experiments using real world examples,
2. Outline the characteristics and assumptions of natural experiments, including strengths and limitations; and,
3. Explain the design considerations and analytical methods within natural experiment evaluation

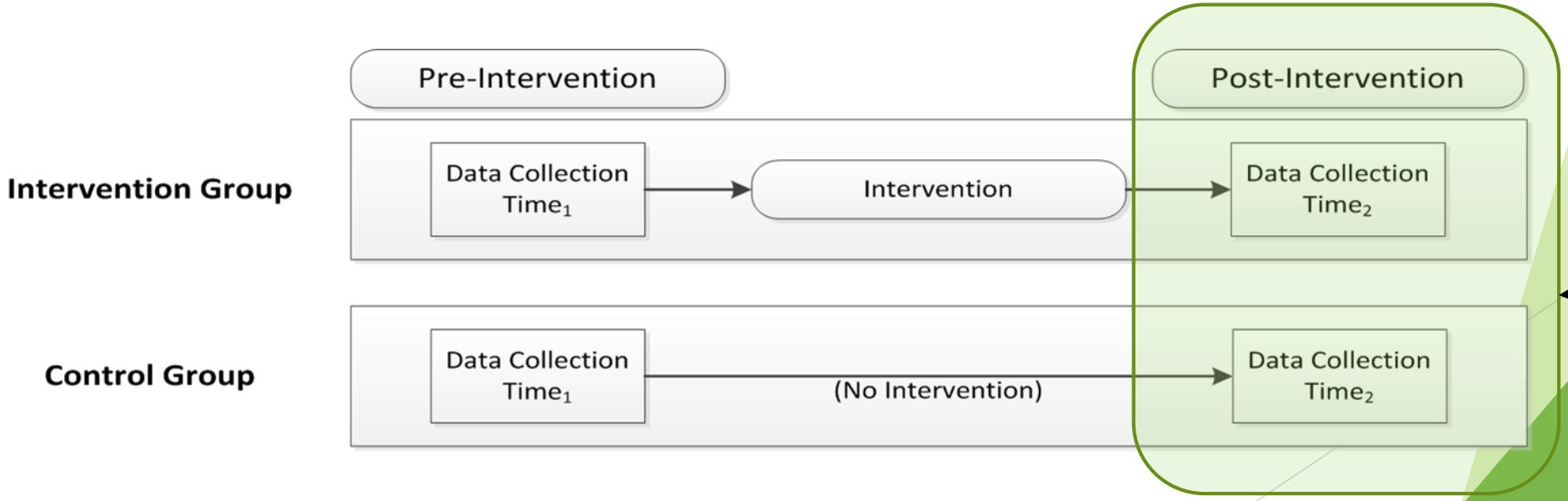
As a brief reminder….

- A primary purpose of an experiment is to manipulate a treatment variable (*intervention*) to determine the effect on a dependent variable (*outcome*).
  - Key features of a traditional experiment include an intervention group, a control group, and pre-intervention (Time1) and post-intervention (Time2) measures from both groups.
  - In an experiment, the analysis compares the difference in the Time1 to Time2 values for the outcome in the intervention group relative to the control group.

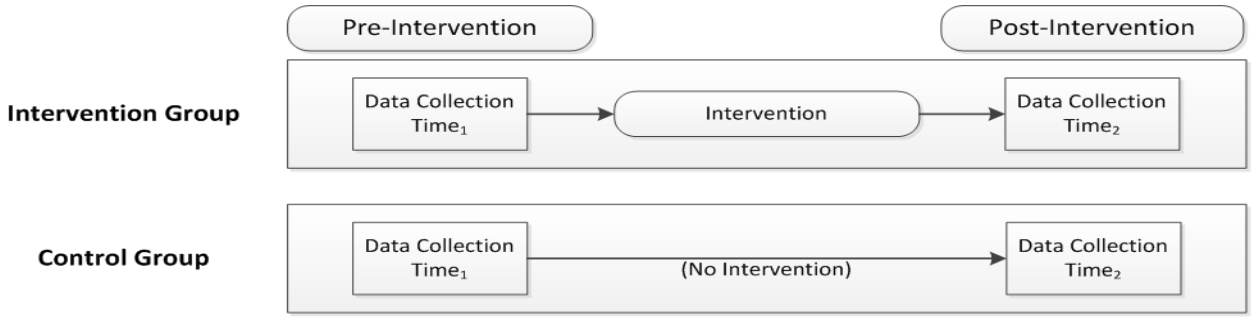


- In general, if the change observed for the outcome is significantly different in the intervention group than the control group, the experiment would suggest that the difference may be due to the intervention.

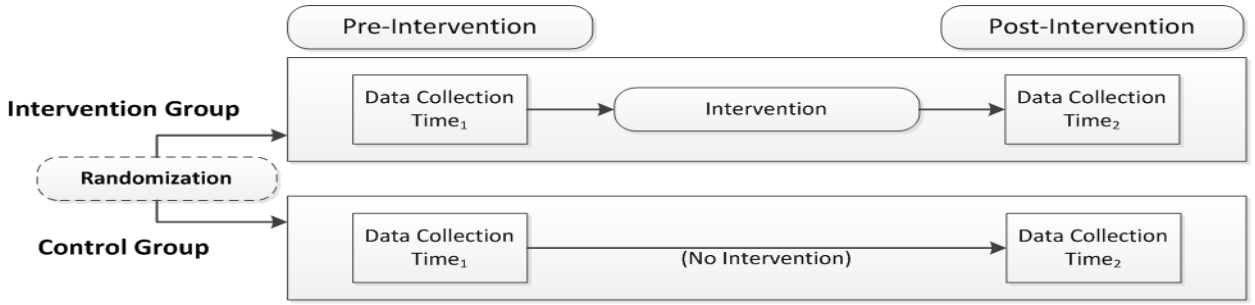
- If the change observed for the outcome is **not** significantly different between the intervention group and the control group, the experiment would suggest that the intervention did not work.



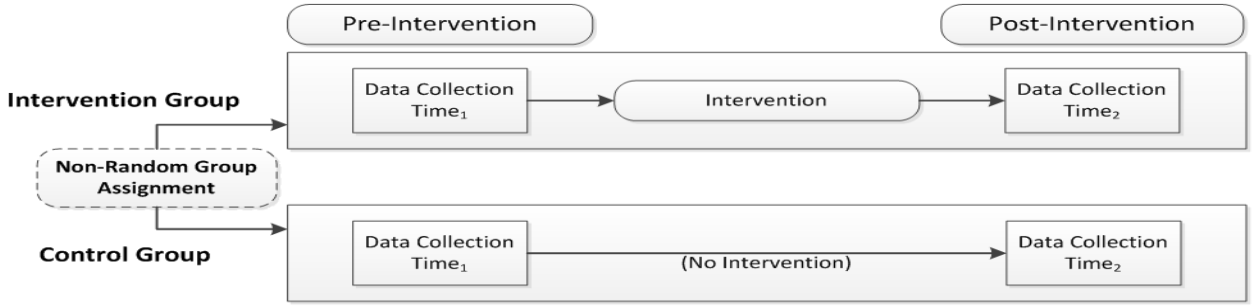
- Basic Experimental Design



- Basic Randomized Experimental Design



- Basic Quasi-Experimental Design



What is a natural experiment?

- A natural experiment is when a particular ***intervention*** is implemented, but the circumstances surrounding the implementation are not under the control of researchers.

What is a natural experiment?

- A natural experiment is when a particular *intervention* has been implemented but the circumstances surrounding the implementation are not under the control of researchers.

- In a natural experiment, the intervention can take multiple forms:
  - Environmental structural change
    - e.g., changes to the built environment (e.g., highways, parks, grocery stores, etc.)
  - Program changes
    - e.g., a new program is implemented or an existing program is altered (e.g., HPV vaccines)
  - Policy changes
    - e.g., a new policy is implemented or an existing policy is altered (e.g., cannabis legalization Federally in Canada)

What is a natural experiment?

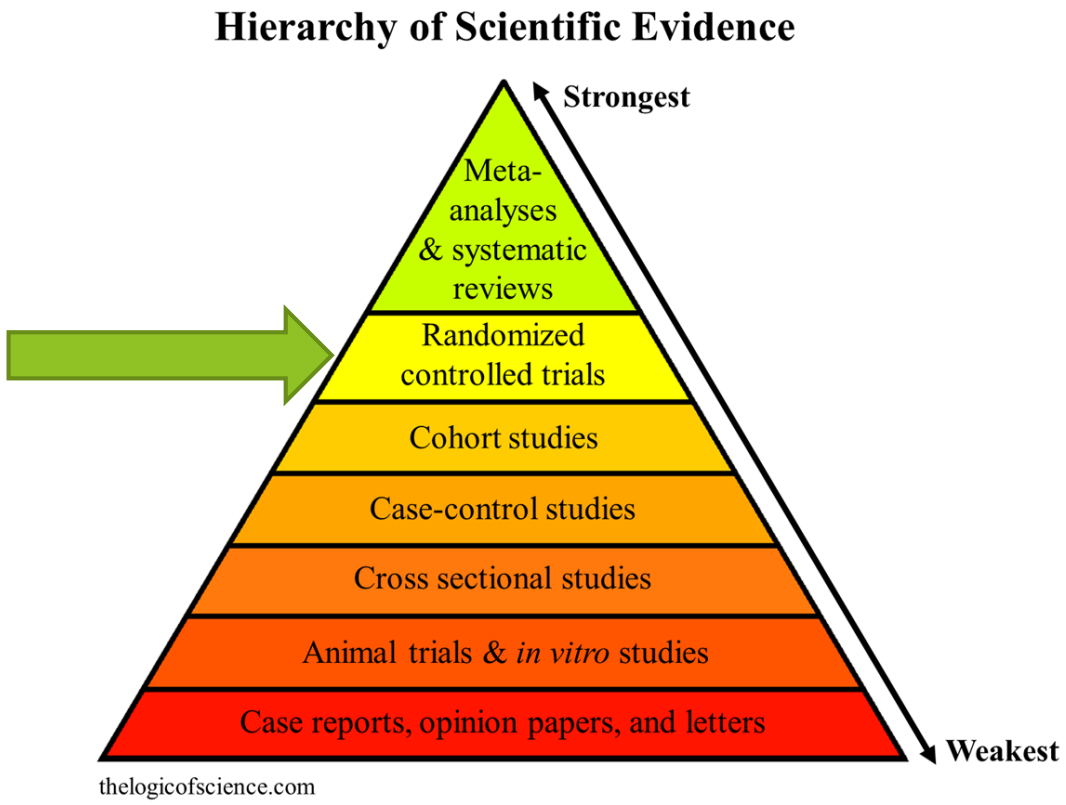
- Two additional features often associated with natural experiments are:

  1. the implementation of the intervention is not dependent on whether or not there is a plan to evaluate the intervention, and

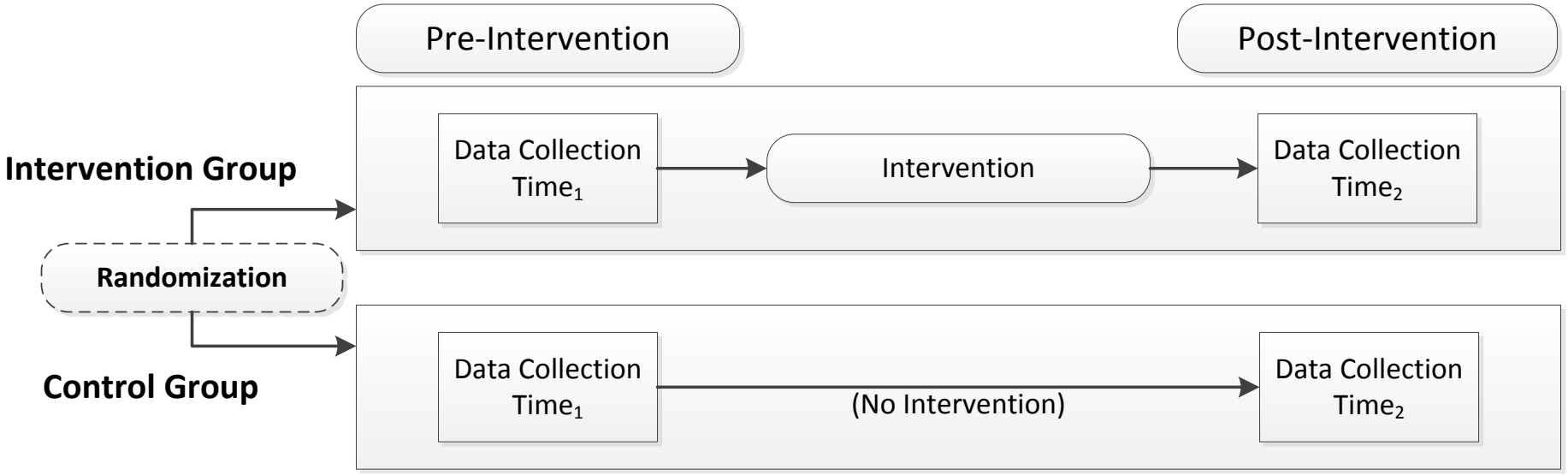
  2. **random allocation** of the intervention is not feasible for ethical, political or other reasons.

Randomized Control Trial

- In many research domains (e.g., health services treatment), the randomized control trial (RCT) is considered by many to be the best way to understand the true impact of an intervention.

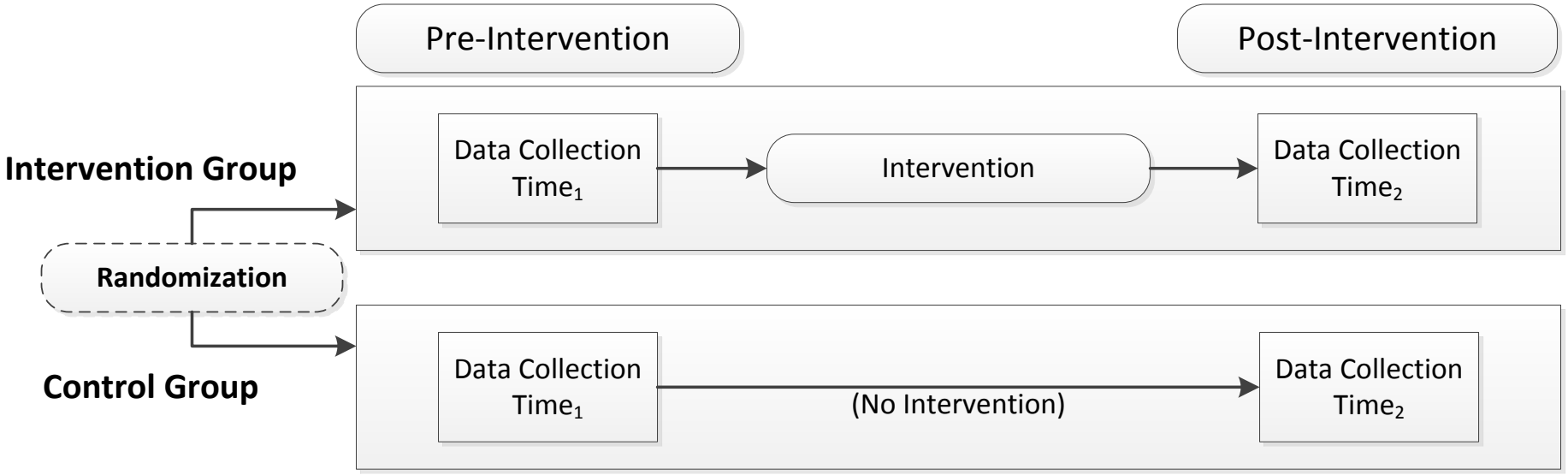


Randomized Control Trial



- After randomization, the two groups of subjects are followed in exactly the same way over time
  - the only differences between them is the *intervention* (treatment) they received.

Randomized Control Trial



- Without random allocation to the treatment or control group, you cant ensure that the groups were the same prior to treatment
  - makes it potentially inaccurate to claim that the observed differences were actually due to the treatment being effective.

The problem with RCTs in Public Health

- In health promotion and chronic disease prevention, an RCT design would often **not** be considered ethical, politically feasible, or appropriate for evaluating the impact of many public health interventions.

The problem with RCTs

- In health promotion and chronic disease prevention, an RCT design would often **not** be considered ethical, politically feasible, or appropriate for evaluating the impact of many public health interventions.

- For example:
  - It is often not ethical to randomize people to different policy conditions
    - e.g., requiring people to move to a particular state (or province 🇨🇦)
  - Program implementation decisions can sometimes be determined due to a variety of considerations, including political or fiscal constraints
    - e.g., certain food programs may be targeted to communities in need (i.e., inner cities) rather than randomizing across all communities at the state-level or nationally

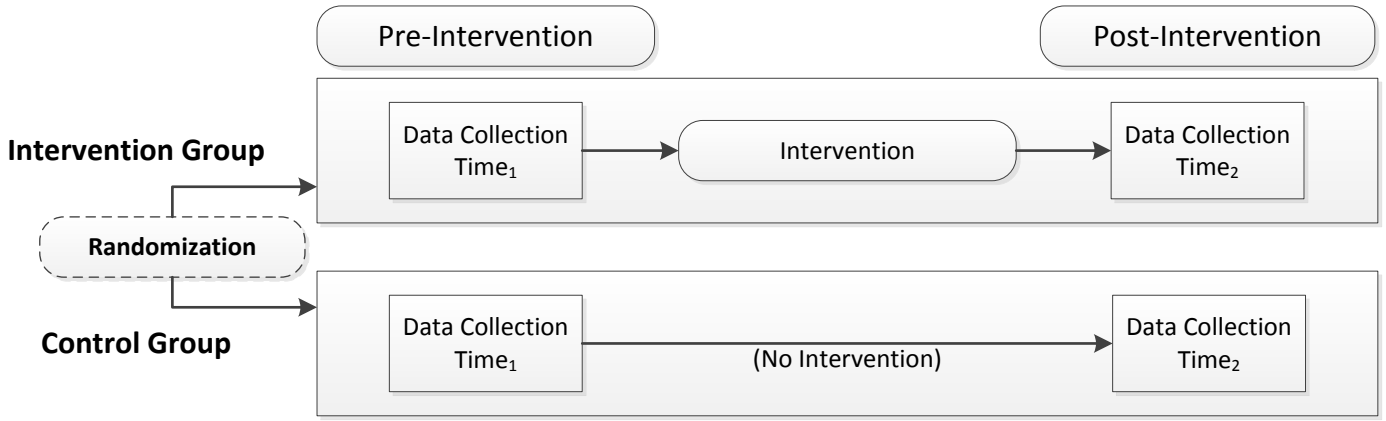
The problem with RCTs

- In health promotion and chronic disease prevention, an RCT design would often **not** be considered ethical, politically feasible, or appropriate for evaluating the impact of many public health interventions.

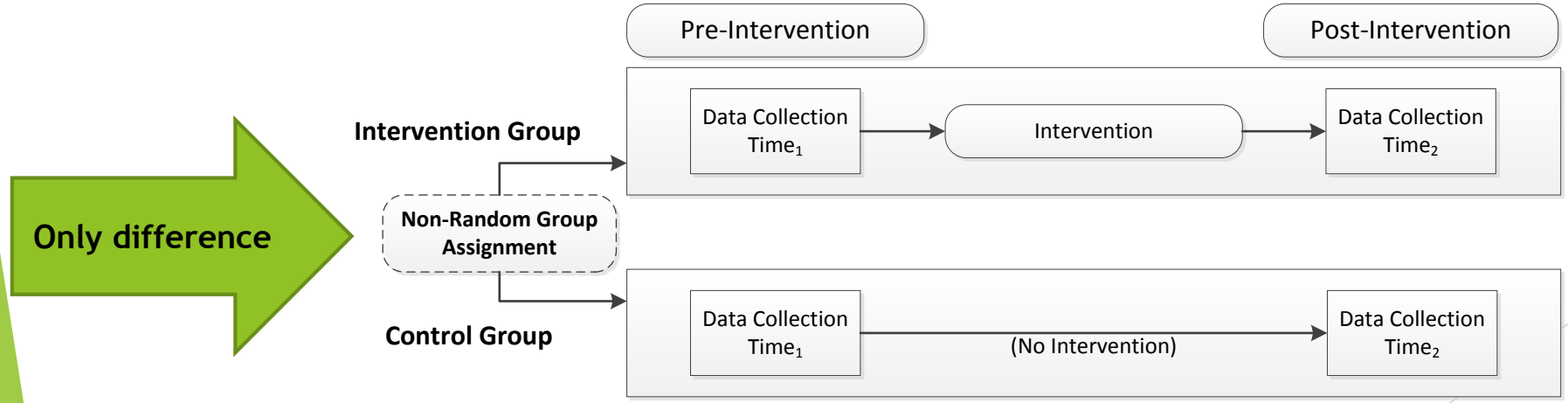
- For example:
  - It is often not ethical to randomize people to different policy conditions
    - e.g., requiring people to move to a particular province
  - Program implementation decisions can sometimes be determined due to a variety of considerations, including political or fiscal constraints
    - e.g., certain food programs are targeted to communities in need (i.e., remote northern communities) rather than randomizing across all Canadian communities

We must look to alternative research methods for determining the impact of certain public health interventions when randomization is not feasible.

Classic RCT design



Classic quasi-experimental design



Why are natural experiments important?

- Public health decision makers are increasingly pushed to provide evidence-based prevention interventions, however, a large gap exists between
  a) the type of research evidence that is **available**
  b) the type of research evidence that is **needed**

  to inform the public health prevention agenda.

Why are natural experiments important?

- Public health decision makers are increasingly pushed to provide evidence-based prevention interventions, however, a large gap exists between
  a) the type of research evidence that is **available**
  b) the type of research evidence that is **needed**

  to inform the public health prevention agenda.

- However, in many cases:
  - Timely evidence is often not available
  - Contextually relevant evidence is often not available

The issues faced by decision makers

1. Much of the available evidence intended for use within public health is from overly controlled research designs (RCTs) that:
   a) do not mimic real-world conditions,
   b) that usually fail to consider different contextual confounders, and
   c) often ignore the various competing interests that exist in public health practice.

The issues faced by decision makers

1. Much of the available evidence intended for use within public health is from overly controlled research designs (RCTs) that:
   a) do not mimic real-world conditions,
   b) that usually fail to consider different contextual confounders, and
   c) often ignore the various competing interests that exist in public health practice.

and

2. The need to be cautious about making important public health decisions based solely on extremely biased evidence derived from non-experimental observational studies that lack methodological rigor.

Missed Opportunities

- In the US, there are tens of thousands of natural experiment interventions (i.e., ongoing structural, program, and policy changes) implemented by decision makers and stakeholders annually that are not evaluated.
  - I am sure everyone in the room can think of program or policy changes that were not adequately evaluated

Missed Opportunities

- In the US, there are tens of thousands of natural experiment interventions (i.e., ongoing structural, program, and policy changes) implemented by decision makers and stakeholders annually that are not evaluated.
  - I am sure everyone in the room can think of program or policy changes that were not adequately evaluated

- These unevaluated natural experiments currently represent numerous lost opportunities for generating timely practice-based evidence by determining:
  - what works,
  - for whom, and
  - in what context.

Why does this matter?

- Although evidence from the evaluation of natural experiments may be imperfect, we know that:

  1. It can often be more relevant to public health stakeholders because it produces evidence of intervention ***impact*** that reflects the actual realities of intervention implementation; and,

  2. It can identify promising interventions in circumstances when decision makers implement innovative new interventions that have not been tried or evaluated elsewhere.

Why does this matter?

- Although evidence from the evaluation of NEs may be imperfect:

  1. It can often be more relevant to public health stakeholders because it produces evidence of intervention ***impact*** that reflects the actual realities of intervention implementation; and,

  2. It can identify promising interventions in circumstances when decision makers implement innovative new interventions that have not been tried or evaluated elsewhere.

- However, it is equally important to recognize that the strength of evidence derived from natural experiments can be highly variable depending on the strength of the research design used in the evaluation.

What is a natural experimental study?

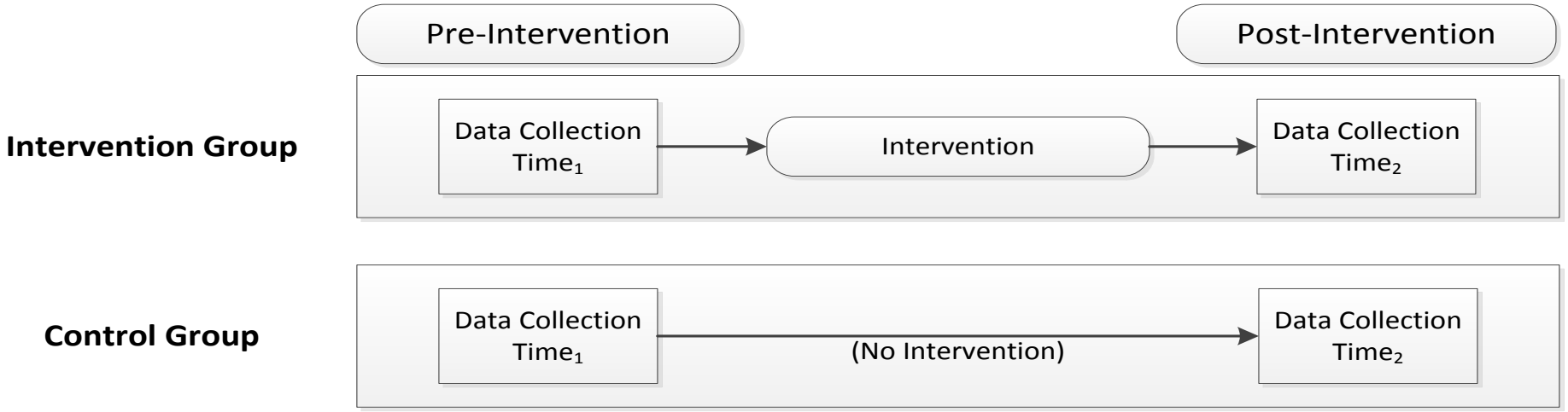
- When a natural experiment occurs, the methodological tool used to evaluate the impact of the intervention is called a *natural experimental study*.

What is a natural experimental study?

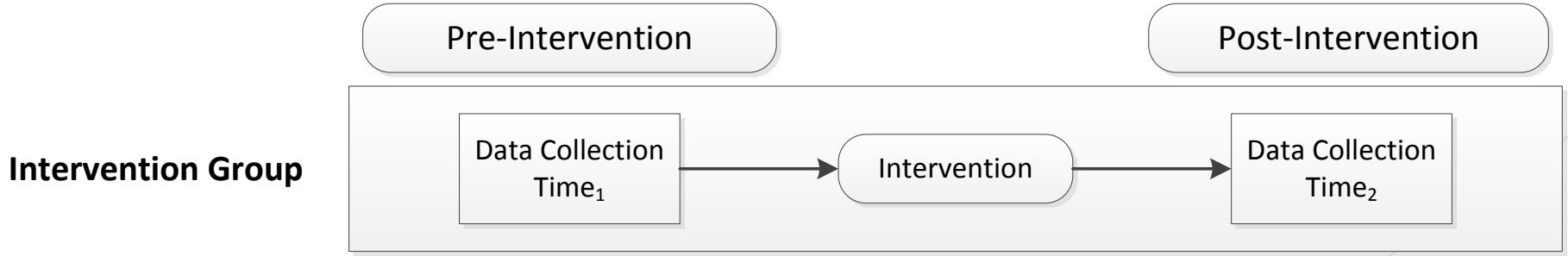
- When a natural experiment occurs, the methodological tool used to evaluate the impact of the intervention is called a *natural experimental study*.

- Natural experimental studies can be either:
  - an experimental design (*intervention and control groups*), or
  - a non-experimental design (*no control group*).

Traditional experimental design



Non-experimental design (no control group)



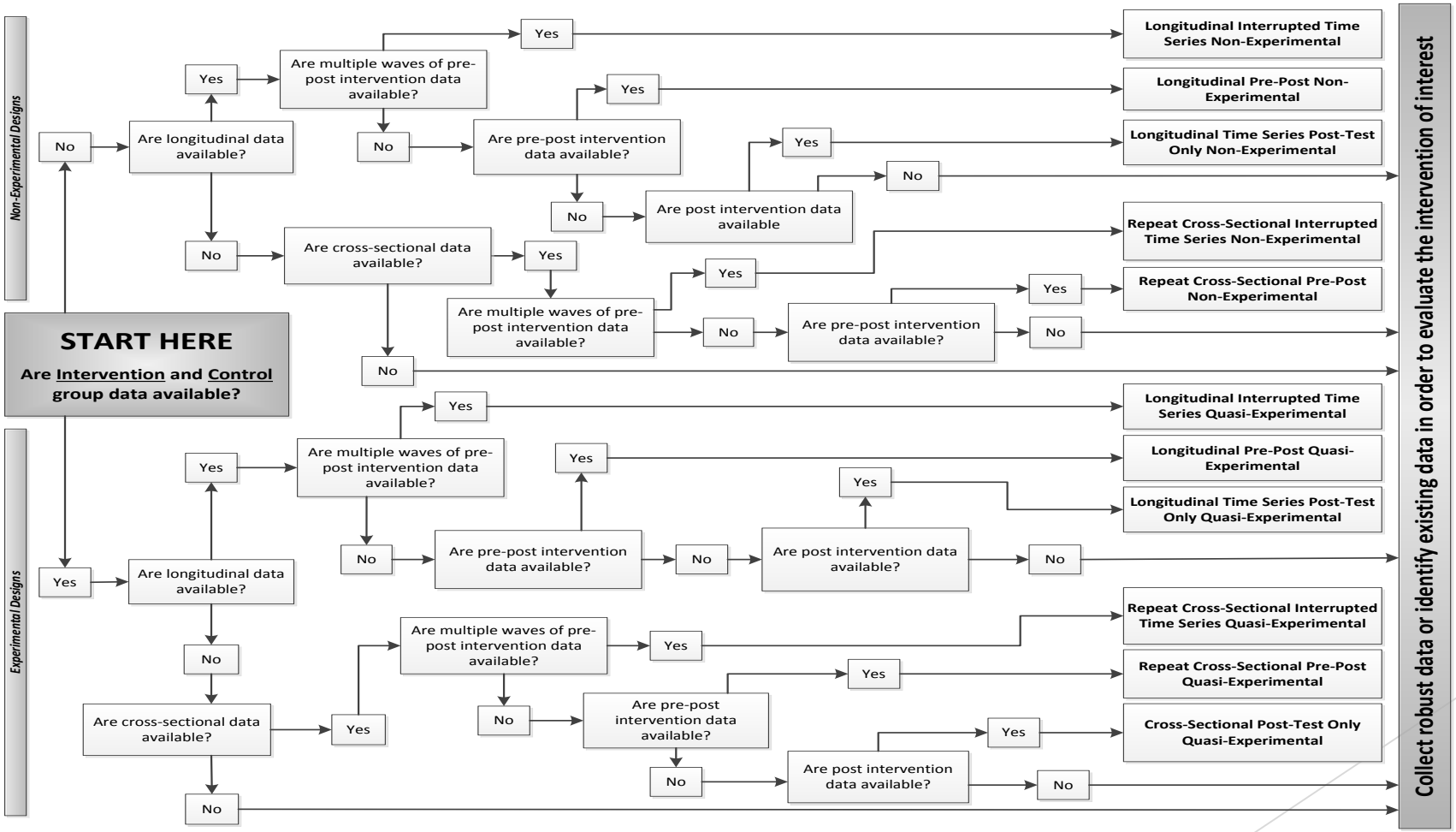
What is a natural experimental study?

- When a natural experiment occurs, the methodological tool used to evaluate the impact of the intervention is called a *natural experimental study*.

- Natural experimental studies can be either:
  - an experimental design (*intervention and control groups*), or
  - a non-experimental design (*no control group*).

The design largely depends on the type of data that are available when the natural experiment occurred.

So what design options are available for evaluating natural experiments?

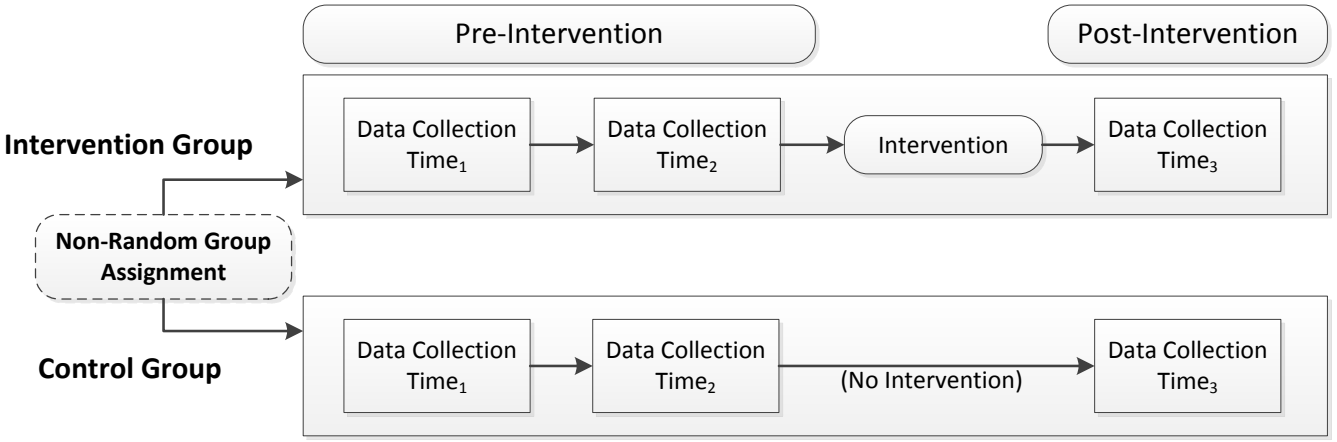


Experimental designs appropriate for evaluating natural experiments

- There are at least six different experimental designs appropriate for evaluating natural experiments
  - Longitudinal interrupted time series quasi-experimental
  - Longitudinal pre-post quasi-experimental
  - Longitudinal time series post-test only quasi-experimental
  - Repeat cross-sectional interrupted time series quasi-experimental
  - Repeat cross-sectional pre-post quasi-experimental
  - Cross-sectional post-test only quasi-experimental

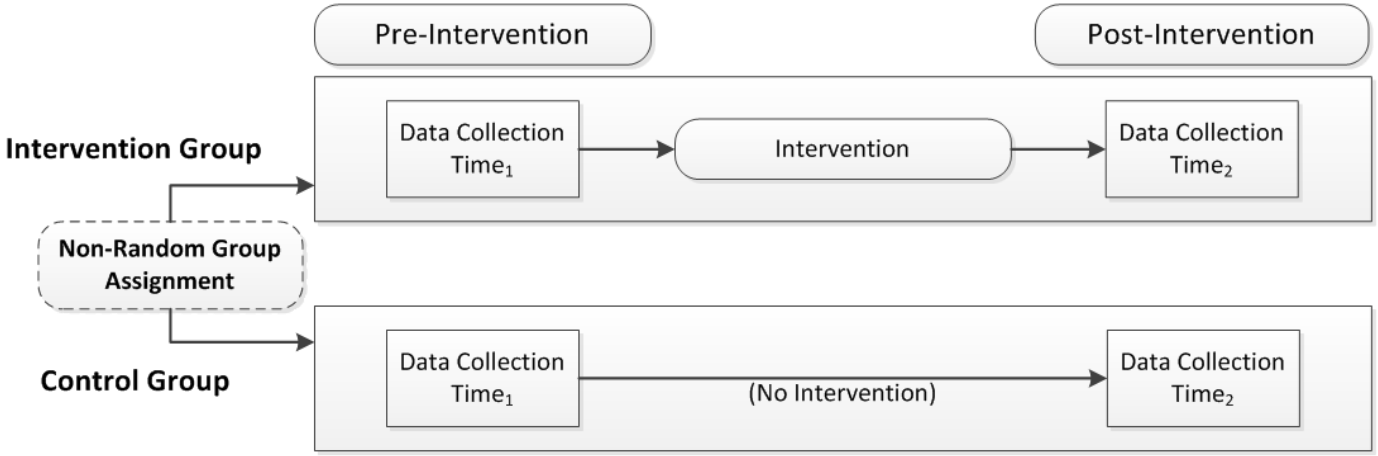
- Each design has particular strengths and weaknesses.

Longitudinal Interrupted Time Series Quasi-Experimental



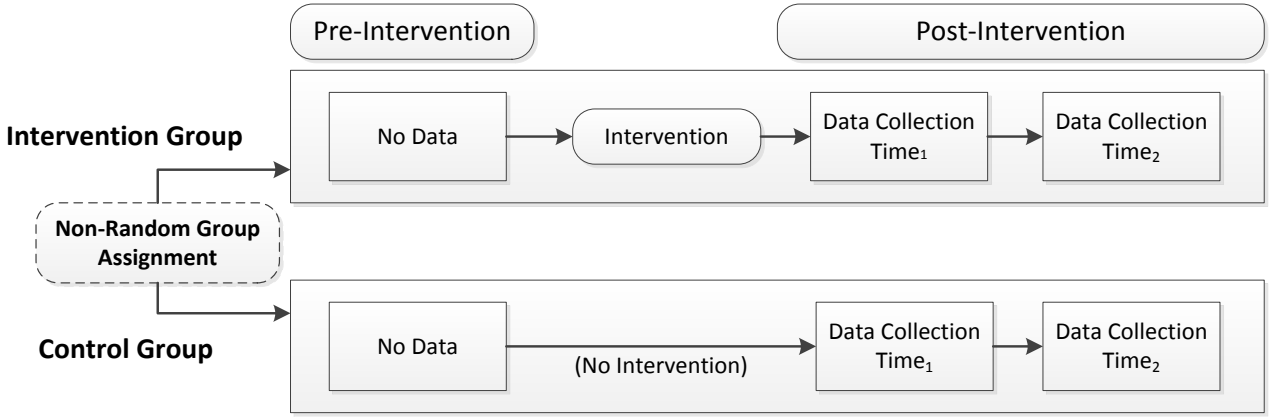
- Requires longitudinal data and 1 (or more) control groups
- The additional pre-intervention measure (Time1) is used to determine if the intervention and control groups were experiencing differences in their pre-intervention trajectories.
- This design is strong with respect to internal validity, and the additional pre-intervention and/or post-intervention measures add further strength.

Longitudinal Pre-Post Quasi-Experimental



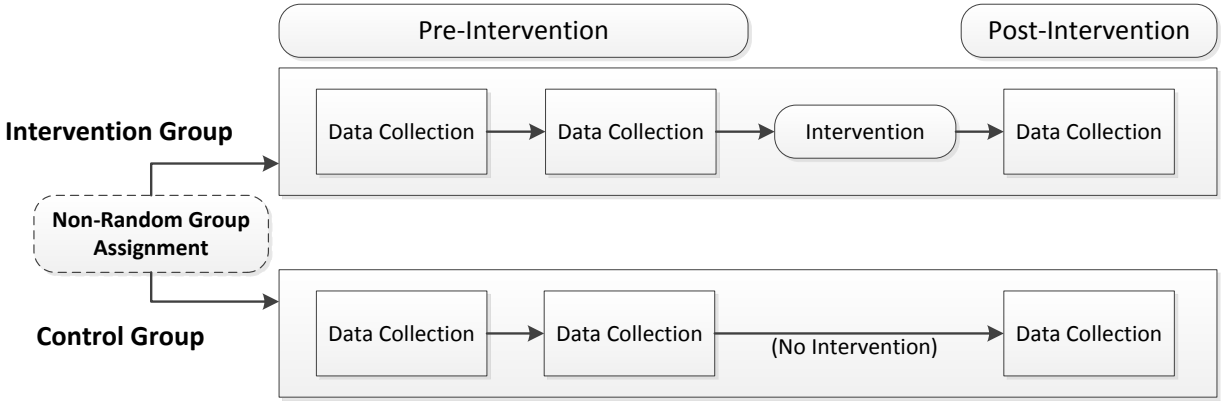
- This quasi-experimental design is also strong with respect to internal validity.

Longitudinal Time Series Post-Test Only Quasi-Experimental



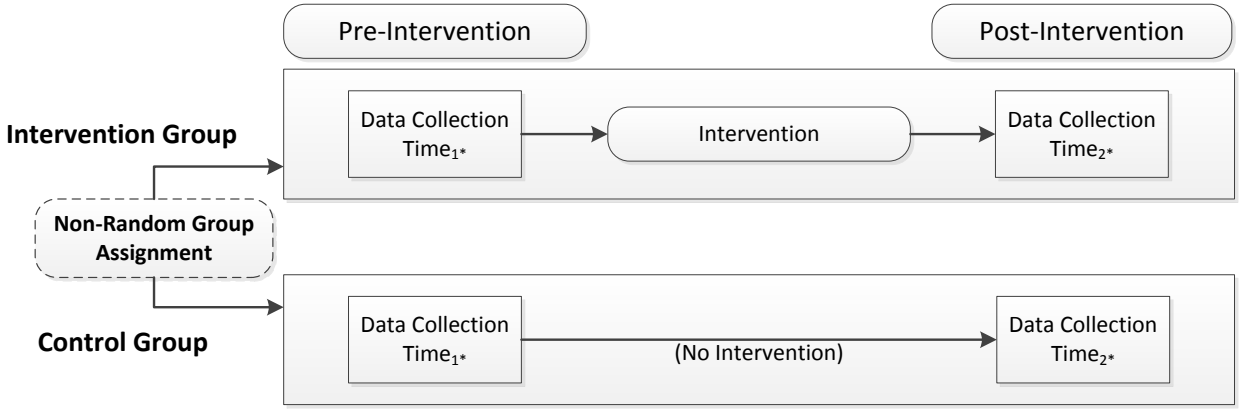
- Requires longitudinal data but with no baseline data
- The additional post-intervention measure (Time2) is used to determine if the intervention and control groups experienced differences in their post-intervention trajectories.

Repeat Cross-Sectional Interrupted Time Series Quasi-Experimental



- Requires repeat cross-sectional data
- Similar to classic quasi-experimental design, except that the respondents at Time 1 are not the same respondents at Time 2 or Time 3
- As such, design is **weak** with respect to internal validity since there may be substantial selection bias introduced by unmeasured changes in the composition of the samples over time.

Repeat Cross-sectional Pre-Post Quasi-Experimental

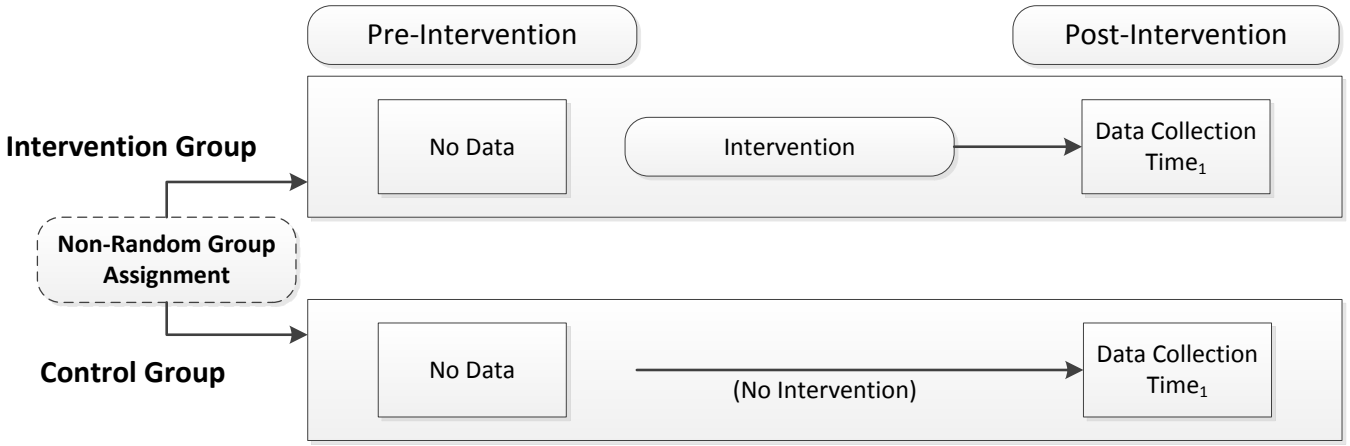


* Note: the respondents who participated at $Time_1$ are not the same respondents who participated at $Time_2$

- Similar to classic quasi-experimental design, except that the respondents at Time 1 are not the same respondents at Time 2

- As such, design is **weak** with respect to internal validity since there may be substantial selection bias introduced by unmeasured changes in the composition of the samples over time.

Cross-sectional Post-Test Only Quasi-Experimental



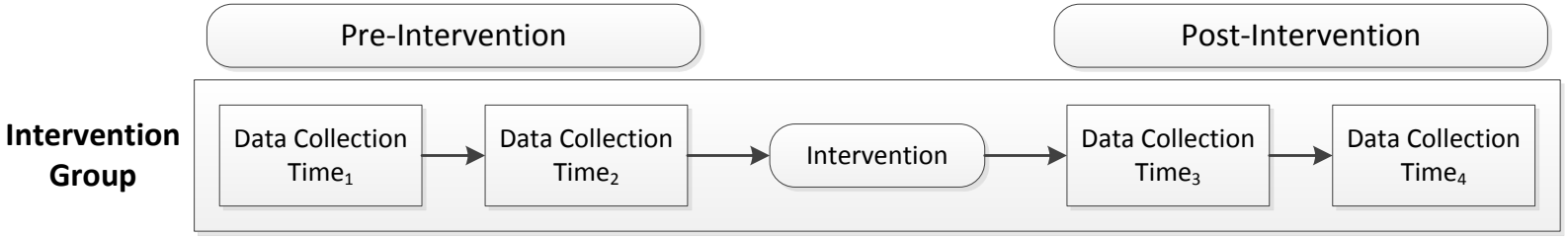
- Requires cross-sectional data
- Seems to be a common approach used in the natural experiment literature because it is easy and data are often available.
- This design is very weak with respect to internal validity since there are no baseline data.

Non-Experimental designs appropriate for evaluating natural experiments

- There are at least five different non-experimental designs appropriate for evaluating natural experiments
  - Longitudinal interrupted time series non-experimental
  - Longitudinal pre-post non-experimental
  - Longitudinal time series post-test only non-experimental
  - Repeat cross-sectional interrupted time series non-experimental
  - Repeat cross-sectional pre-post non-experimental

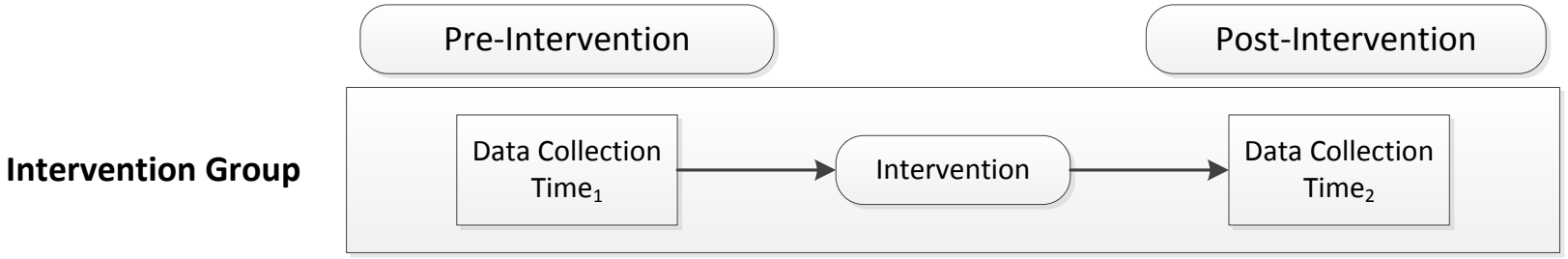
- While these designs are not as robust as experimental designs, the evidence derived from non-experimental evaluations of natural experiments may be considered better than having no evidence at all.

Longitudinal Interrupted Time Series Non-Experimental



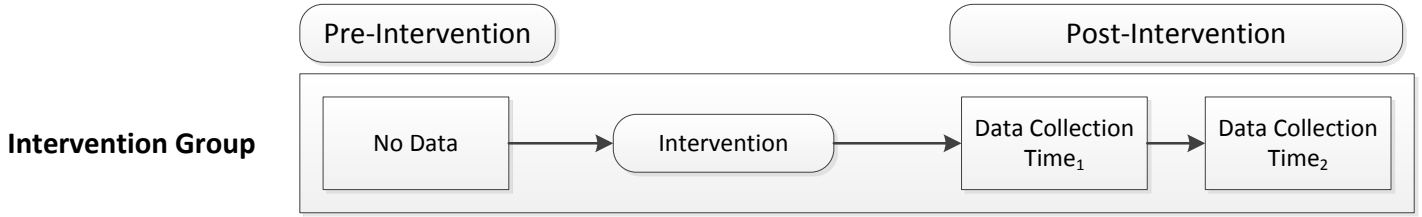
- Requires longitudinal data
- The additional pre-intervention and post-intervention measures help control for temporal or secular changes observed in the outcome observed in the intervention group.
- This design is very weak with respect to internal validity since there are no control group data to compare to.
- Strongest of the non-experimental designs.
  - Note: the evidence a non-experimental design (i.e., no control group) is generally considered weaker than the evidence derived from an experimental design

Longitudinal Pre-Post Non-Experimental



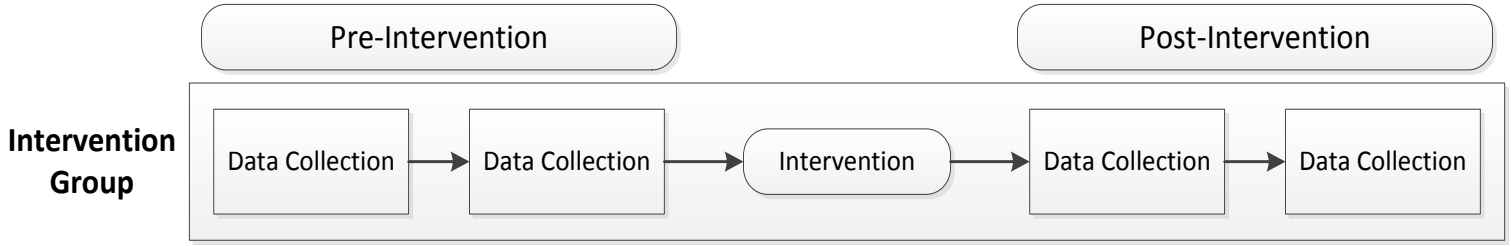
- Requires longitudinal data
- This design is **very weak** with respect to internal validity since there are no control group data to compare to.

Longitudinal Time Series Post-Test Only Non-Experimental



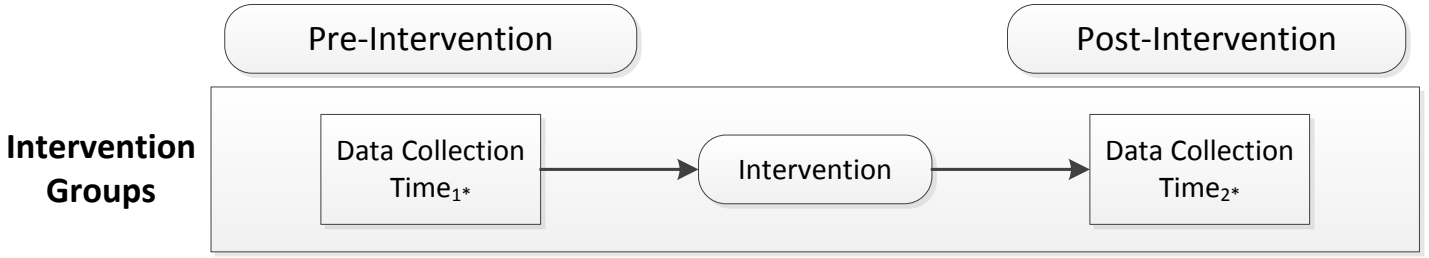
- Requires longitudinal data but with no baseline data
- The additional post-intervention measure (Time2) is used to determine if the intervention group experienced changes in the post-intervention trajectory.
- This design is very weak with respect to internal validity since there are no control group data to compare to and no baseline data.

Repeat Cross-Sectional Interrupted Time Series Non-Experimental



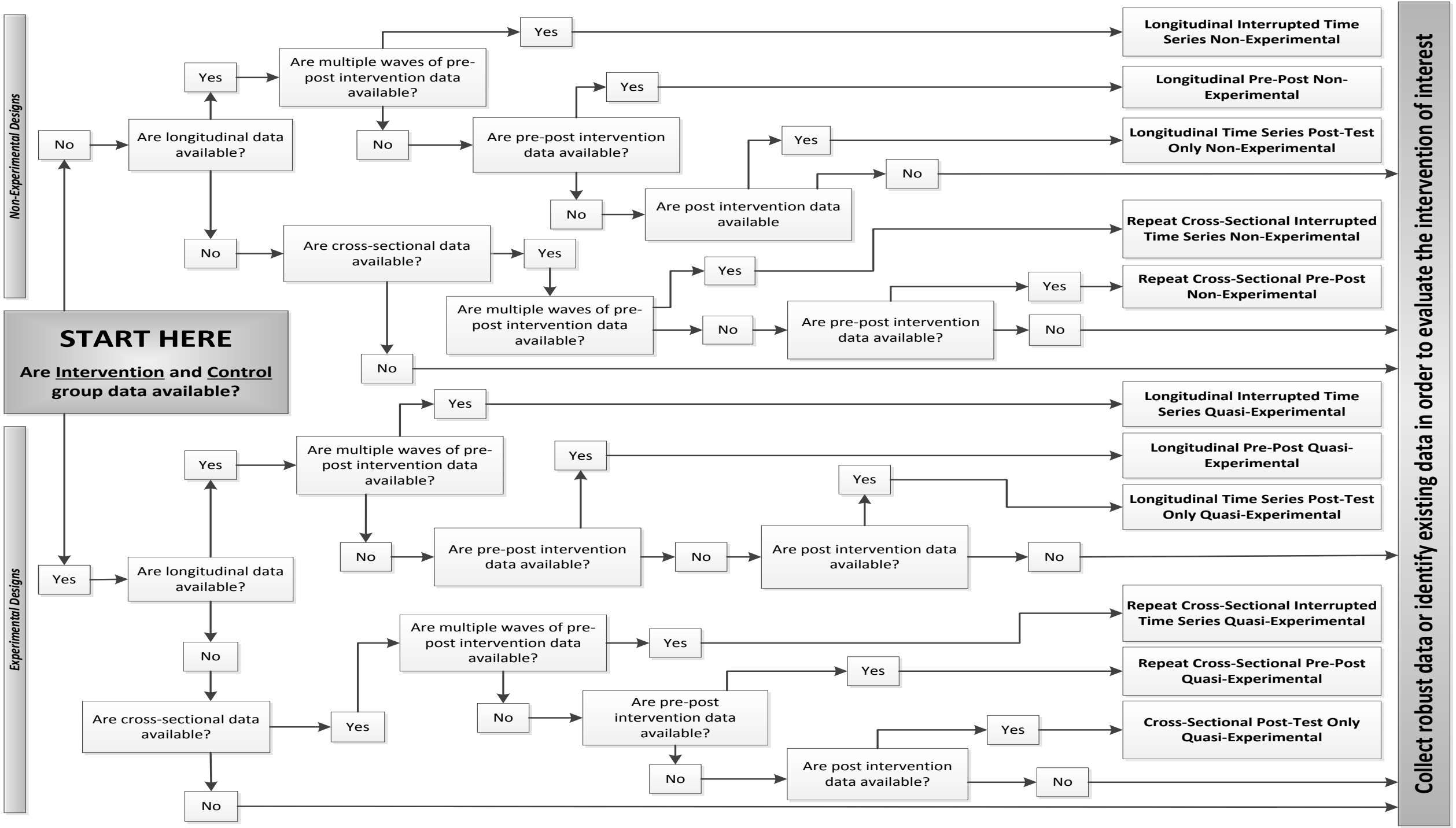
- Requires repeat cross-sectional data
- The additional pre- and post-intervention measures (Time2) are used to determine if the intervention group experienced changes in the post-intervention trajectory.
- Respondents at Time 1 are not the same respondents at Time 2 or Time 3 or Time 4
- This design is very weak with respect to internal validity since there are no control group data to compare to and there may be substantial selection bias introduced by unmeasured changes in the composition of the samples over time.

Repeat Cross-Sectional Pre-Post Non-Experimental

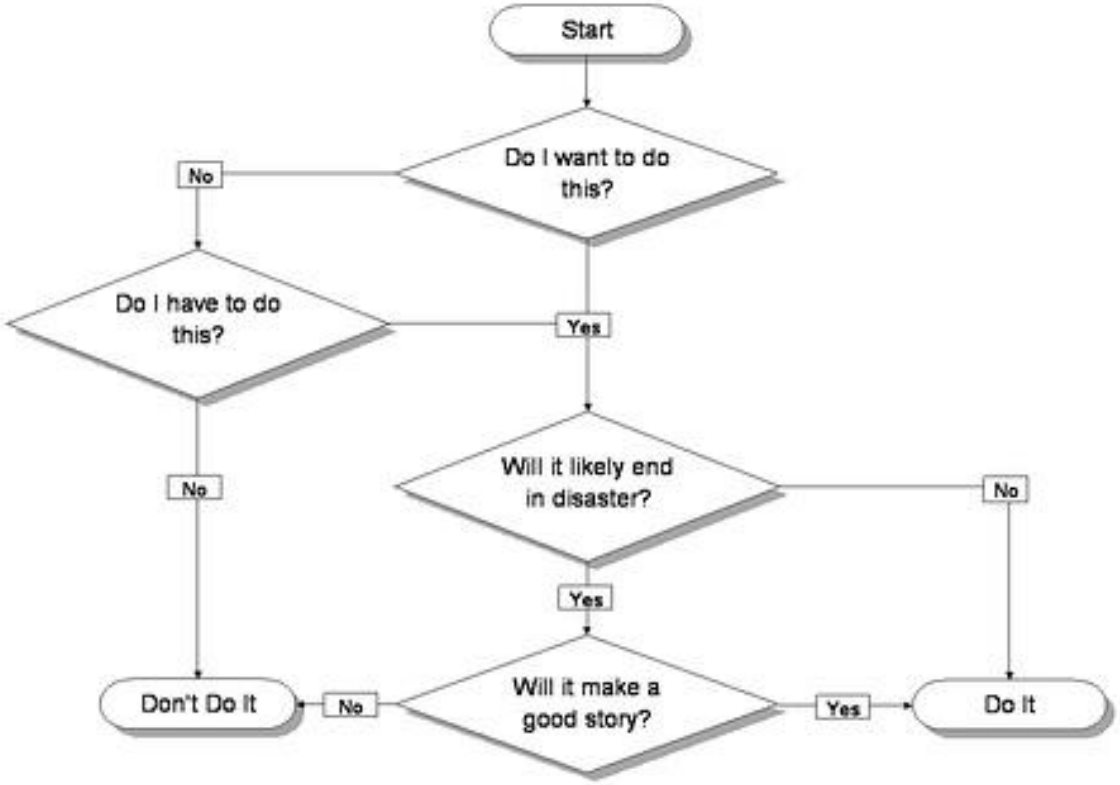


* Note: the respondents who participated at $Time_1$ are not the same respondents who participated at $Time_2$

- Requires repeat cross-sectional data
- Similar to non-experimental interrupted time series design, except that the respondents at Time 1 are not the same respondents at Time 2
- As such, this design is considerably weaker than the alternatives given the potential for substantial bias to be introduced by unmeasured changes in:
  - the composition of the samples over time, or
  - due to other unmeasured contextual confounders influencing the outcomes over time


Non-Experimental Designs
Experimental Designs
START HERE
Are Intervention and Control group data available?
No
Yes
Are longitudinal data available?
Are multiple waves of pre-post intervention data available?
Are pre-post intervention data available?
Are post intervention data available
Are cross-sectional data available?
Longitudinal Interrupted Time Series Non-Experimental
Longitudinal Pre-Post Non-Experimental
Longitudinal Time Series Post-Test Only Non-Experimental
Repeat Cross-Sectional Interrupted Time Series Non-Experimental
Repeat Cross-Sectional Pre-Post Non-Experimental
Longitudinal Interrupted Time Series Quasi-Experimental
Longitudinal Pre-Post Quasi-Experimental
Longitudinal Time Series Post-Test Only Quasi-Experimental
Repeat Cross-Sectional Interrupted Time Series Quasi-Experimental
Repeat Cross-Sectional Pre-Post Quasi-Experimental
Cross-Sectional Post-Test Only Quasi-Experimental
Collect robust data or identify existing data in order to evaluate the intervention of interest

When is a natural experiment worth evaluating?



When to evaluate a natural experiment?

- The value of evaluating a natural experiment may depend on a range of factors, including:
  - the size of the population affected;
  - the likely impact of the intervention;
  - political interest and will to act on the evidence;
  - alignment with informing current and emerging public health priorities; and,
  - the practicalities of collecting new data or using existing data to evaluate the intervention impact.

When to evaluate a natural experiment?

- The value of evaluating a natural experiment may depend on a range of factors, including:
  - the size of the population affected;
  - the likely impact of the intervention;
  - political interest and will to act on the evidence;
  - alignment with informing current and emerging public health priorities; and,
  - the practicalities of collecting new data or using existing data to evaluate the intervention impact.

Researchers should work with decision makers to ***identify*** and ***prioritize*** which natural experiments warrant evaluation based on data availability and the evidence needed to shape decision-making.

Remember, there are 4 core strengths of natural experimental studies

1. Evaluating Government Policy

- The impact of policy is best understood by using natural experimental study designs.
  - Not only is it important to demonstrate to the public the outcomes of policies implemented by governments (especially when policy implementation activities are largely supported by public funds), the robust evaluation of government policies by researchers can improve government accountability.

Remember, there are 4 core strengths of natural experimental studies

2. Effectiveness vs. Efficacy

- Evidence of policy effectiveness in real-world conditions is more relevant than evidence of policy efficacy in unrealistic over-controlled research settings.
  - When it comes to developing policy, stakeholders require the *best available* evidence as opposed to the most rigorous evidence possible.

- Remember, although conclusions regarding policy impact derived from natural experimental evaluations must be drawn cautiously, when appropriately designed and implemented, **robust** natural experiment evaluations using experimental designs can provide strong evidence of policy impact with robust internal and external validity.

Remember, there are 4 core strengths of natural experimental studies

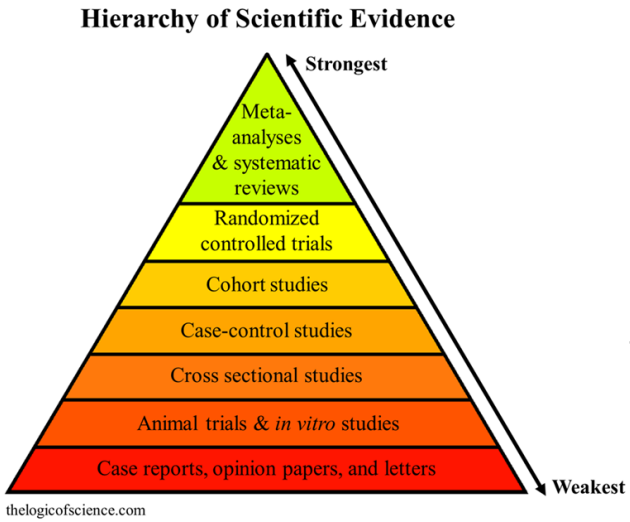
3. Generalizability to other contexts

- Policy evaluation evidence derived from a robust natural experiment evaluation using an experimental design can potentially be generalized to and across different types of populations and settings.
  - Having robustly matched control group(s) similar to the intervention group can make generalization of policy impact more feasible.

Remember, there are 4 core strengths of natural experimental studies

4. Timely Evidence

- While many consider systematic reviews to provide the strongest evidence, substantial time is required for the publication of sufficient number of similar policy evaluation studies to enable a systematic review.
  - Not only are studies evaluating the impact of polices rare in the literature, policy evaluation studies typically take longer than traditional studies of intervention impact (e.g., RCT of clinical treatment).



So many choices… how to decide

So many choices… how to decide

- Dr. Geoff Fong
  - PI of the International Tobacco Control Project (ITC)
    - ITC is the worlds largest international tobacco control policy evaluation project
  - 2015 American Cancer Society, Luther L. Terry Award for “Outstanding Research Contribution”
  - Global ‘thought-leader’ on natural experiments
    - Ranked as #1 threat globally to the tobacco industry

So many choices… how to decide

- According to Geoff, there are four strategies **all** researchers should ideally strive for when designing a study to evaluate a natural experiment in the form of government policy:

Fong et al. 2006

1. using a non-randomized **experimental design** (i.e., quasi-experimental) with an appropriately matched control group rather than a non-randomized non-experimental design with no control group;

So many choices… how to decide

- According to Geoff, there are four strategies **all** researchers should ideally strive for when designing a study to evaluate a natural experiment in the form of government policy:

Fong et al. 2006

1. using a non-randomized experimental design (i.e., quasi-experimental) with an appropriately matched control group rather than a non-randomized non-experimental design with no control group;

2. use a **longitudinal cohort design** where participants are measured on the same outcome variable(s) and correlates at one or more points in time before and after the policy is implemented;

So many choices… how to decide

3. measure and control for individual-level variables that may confound the effectiveness of the policy; and,

So many choices… how to decide

3. measure and control for individual-level variables that may confound the effectiveness of the policy; and,

4. measure and control for contextual-level variables that may also confound the effectiveness of the policy.

So many choices… how to decide

3. measure and control for individual-level variables that may confound the effectiveness of the policy; and,

4. measure and control for contextual-level variables that may also confound the effectiveness of the policy.

- The combination of these four approaches would create a powerful research design for evaluating natural experiments when governments implement new public health policies.

So many choices… how to decide

3. measure and control for individual-level variables that may confound the effectiveness of the policy; and,

4. measure and control for contextual-level variables that may also confound the effectiveness of the policy.

- The combination of these four approaches would create a powerful research design for evaluating natural experiments when governments implement new public health policies.

This evidence would closely match the robust internal validity generally associated with an RCT, when making causal inferences about the impact of a particular policy on an outcome.

Methods to reduce bias in natural experiments

- According to Cook and Campbell, bias in a research design can impact:

Cook & Campbell, 1979

- Internal Validity
  - The validity of inferences about whether observed covariation between the actual intervention (A) and the presumed outcome (B) reflects a causal relationship from A to B, as those variables were manipulated or measured; and,

- External Validity
  - The validity of inferences about whether the cause–effect relationship between the actual intervention (A) and the presumed outcome (B) holds over variation in persons, settings, treatment variables, and measurement variables.

Bias due to confounding

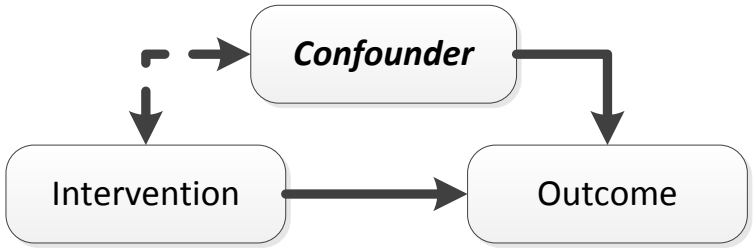
- Confounding is a situation in which the effect of an intervention on an outcome is distorted by the presence of another variable.
  - Positive confounding
    - when the observed intervention effect is biased **away** from the null hypothesis
      - (i.e., falsely assuming the intervention is effective)
  - Negative confounding
    - when the observed intervention effect is biased **toward** the null hypothesis
      - (i.e., falsely assuming the intervention is not effective)

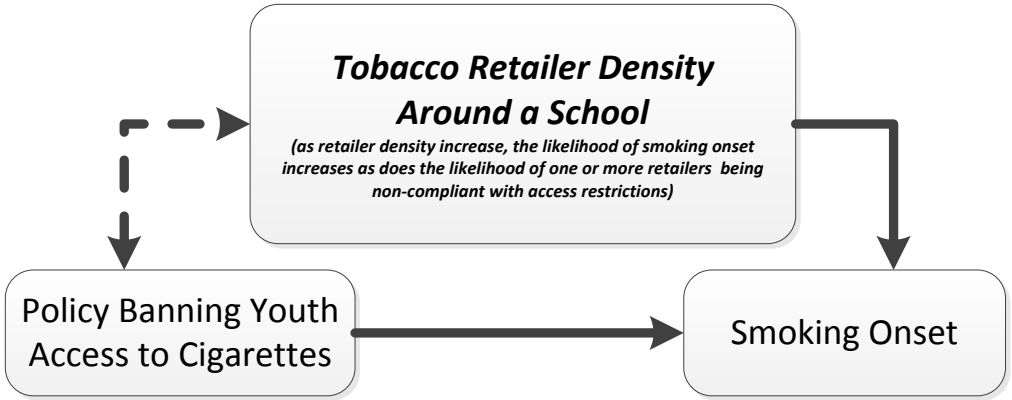
Bias due to confounding

- Confounding is a situation in which the effect of an intervention on an outcome is distorted by the presence of another variable.
  - Positive confounding
    - when the observed intervention effect is biased **away** from the null hypothesis
      - (i.e., falsely assuming the intervention is effective)
  - Negative confounding
    - when the observed intervention effect is biased **toward** the null hypothesis
      - (i.e., falsely assuming the intervention is not effective)

- A confounder is a third variable (other than the intervention and outcome) that is associated with both the intervention and the outcome of interest.
  - The confounder(s) leads to bias by distorting the magnitude and strength of the actual relationship between the intervention and the outcome.

Bias due to confounding





Bias due to confounding

- A RCT provides a robust mechanism for controlling for confounders at the individual-level (i.e., via randomization), hence the robust internal validity of this design.

- However, a RCT does not typically provide a robust means for adjusting for contextual confounders resulting in **lower external validity**.
  - RCT data are generally collected in homogenous contexts rather than heterogeneous contexts

Bias due to confounding

- The lack of randomization of intervention exposure in a natural experiment means natural experimental studies are prone to individual-level bias via confounding.

- However, researchers evaluating natural experiments can limit the amount of bias associated with confounding by:
  - Using quasi-experimental research designs with pre- and post-test measures, and
  - Measuring and adjusting for important contextual characteristics

Matching in experimental research designs for natural experiments

- It is important to note that including a control group in a non-randomized experiment will **not** automatically eliminate potential bias.
  - If the control group differs from the intervention group at baseline for one or more key variables that may impact the effect of the intervention (i.e., a confounder), than these baseline differences can be an important source of bias if not adjusted for properly.

Matching in experimental research designs for natural experiments

- It is important to note that including a control group in a non-randomized experiment will **not** automatically eliminate potential bias.
  - If the control group differs from the intervention group at baseline for one or more key variables that may impact the effect of the intervention (i.e., a confounder), than these baseline differences can be an important source of bias if not adjusted for properly.

- In order to help reduce bias, the control group(s) should ideally be as similar to the intervention group during the pre-intervention observations as possible.

Analytical matching techniques

- When it is not feasible to have robust matching of the intervention and control group at baseline, researchers can used various analytical techniques to adjust for potential differences in the baseline characteristics of the intervention and control groups.

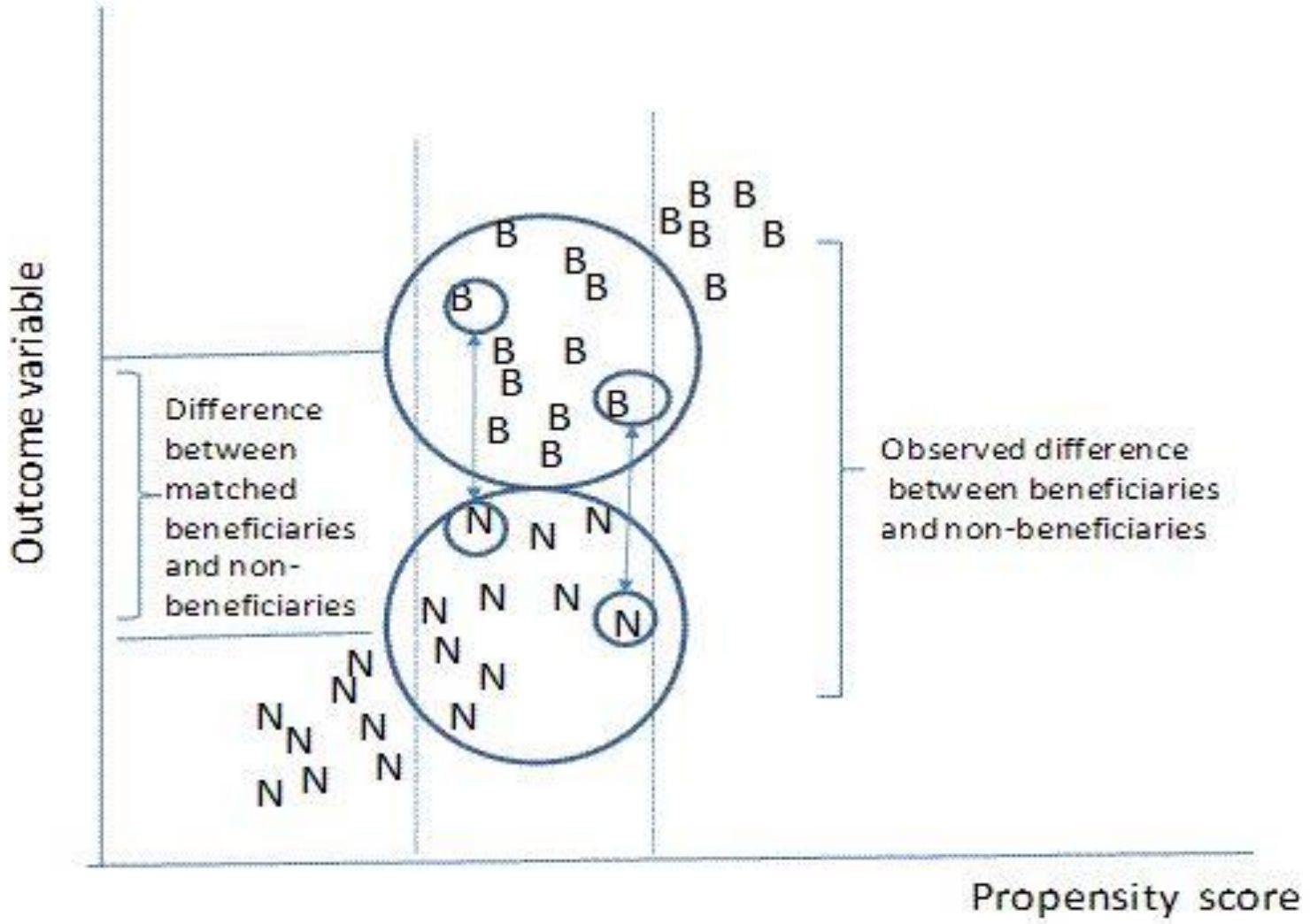
Analytical matching techniques

- When it is not feasible to have robust matching of the intervention and control group at baseline, researchers can used various analytical techniques to adjust for potential differences in the baseline characteristics of the intervention and control groups.

1. Propensity Scores

- Propensity score matching (PSM) uses baseline data to match an individual in the treatment group to a similar individual in the control group
  - This is not based on every single observable characteristic, but on their propensity score (i.e., the likelihood of being in the intervention group given their observable characteristics).
  - PSM helps to ensure that the average characteristics of the treatment and control groups are similar in an effort to obtain an unbiased estimate of the intervention effect at follow-up

Figure 1. A graphical representation of matching on the propensity score



https://s-media-cache-ak0.pinimg.com/564x/37/e9/d2/37e9d20c275fdb292fd9704814513bd9.jpg

Analytical matching techniques

2. Regression Discontinuity

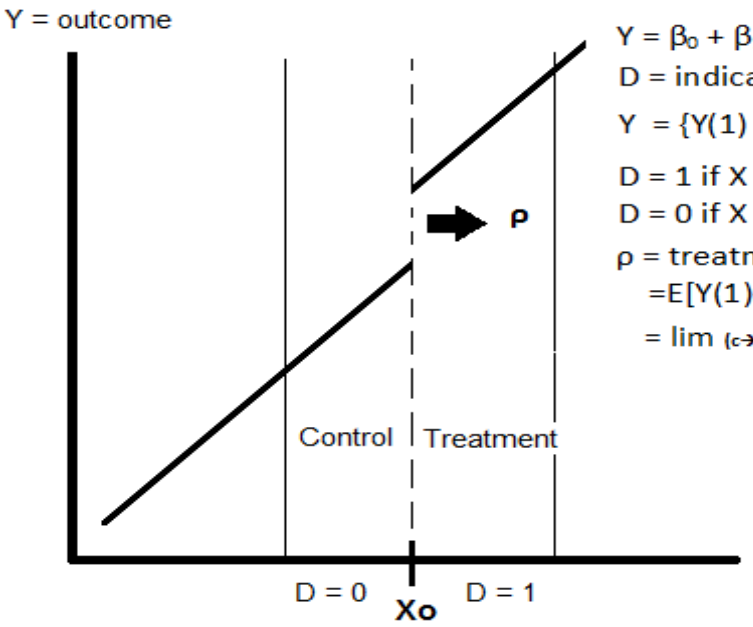
- Can be used in a quasi-experimental study when there is some kind of ***criterion*** or ***threshold*** that must be met before an individual can participate in the intervention being evaluated.
  - For instance, research where an intervention is targeted to particular at-risk groups who are most in need of or would benefit the most from the intervention.
    - e.g., providing an income subsidy to families identified as low income

Analytical matching techniques

2. Regression Discontinuity

- Can be used in a quasi-experimental study when there is some kind of criterion or threshold that must be met before an individual can participate in the intervention being evaluated.
  - For instance, research where an intervention is targeted to particular at-risk groups who are most in need of or would benefit the most from the intervention.

- In this design, participants are assigned to the intervention group or control group based on their score for a pre-determined pre-intervention measure of interest.
  - Assumes that participants with values above and below the threshold are different on an important construct or trait of interest.



$Y = \beta_0 + \beta_1 x + \rho D + e$

D = indicator for treatment assignment

$Y = \{Y(1) \text{ if } D = 1; Y(0) \text{ if } D = 0\}$

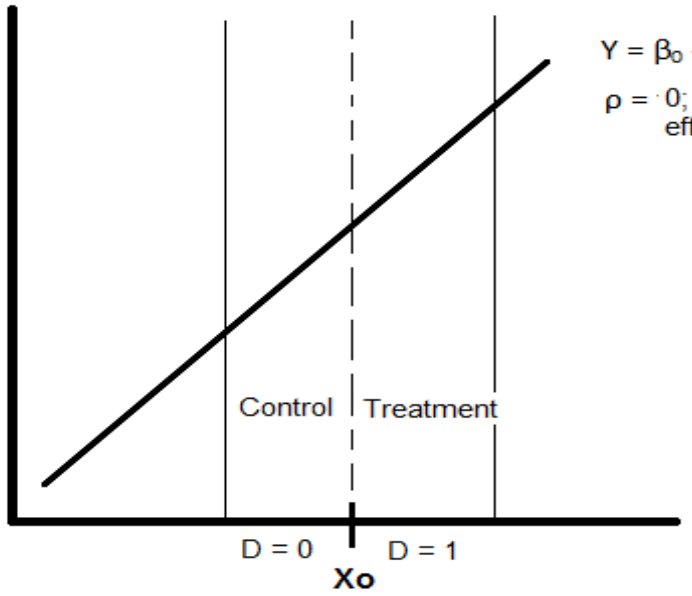
$D = 1 \text{ if } X > X_0$

$D = 0 \text{ if } X < X_0$

ρ = treatment effect

$= E[Y(1)-Y(0) \mid X=X_0]$

$= \lim_{(c \to 0)} E[Y \mid X=X_0-c] - E[Y \mid X=X_0+c]$



$Y = \beta_0 + \beta_1 x + e$

$\rho = 0$; no treatment effect

By properly controlling for the value of the '*criterion variable*' in the regression equation at both pre-test and post-test, one can account for any unobserved differences between the treatment and comparison group

- http://econometricsense.blogspot.ca/2013/05/regression-discontinuity-designs.html

Estimating the effect of alcohol consumption on mortality using the minimum drinking age in the US (Carpenter & Dobkin, 2009)

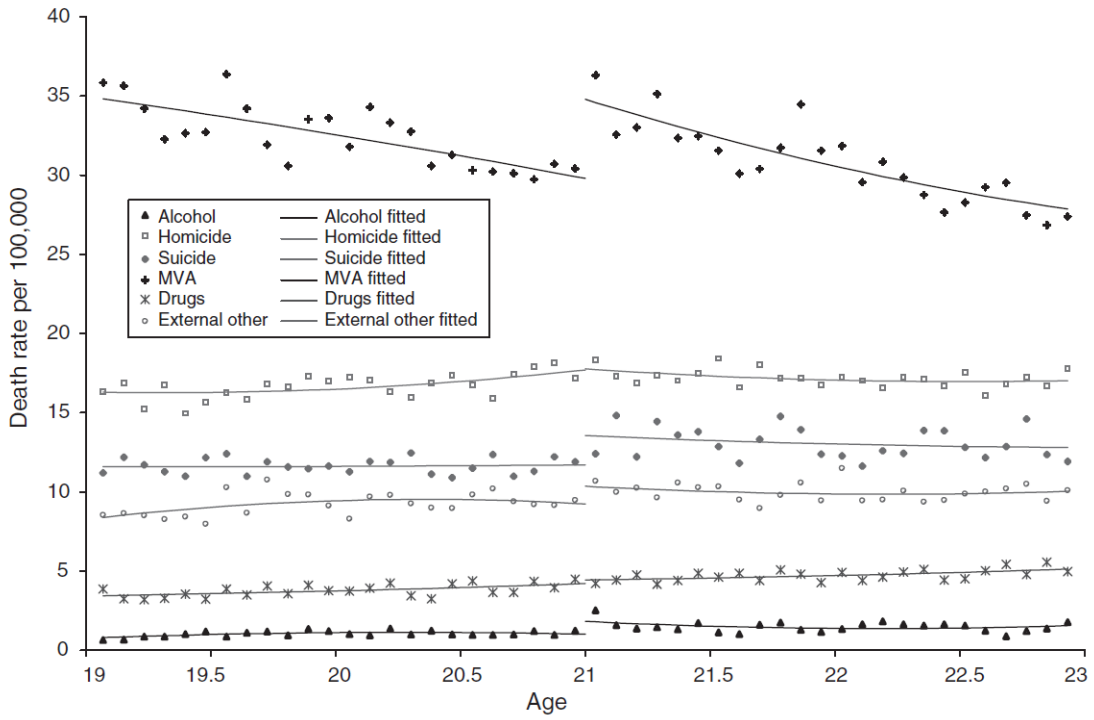


FIGURE 4. AGE PROFILES FOR DEATH RATES BY EXTERNAL CAUSE

- Figure 4 plots the age profile of external deaths separately by cause.

Estimating the effect of alcohol consumption on mortality using the minimum drinking age in the US (Carpenter & Dobkin, 2009)

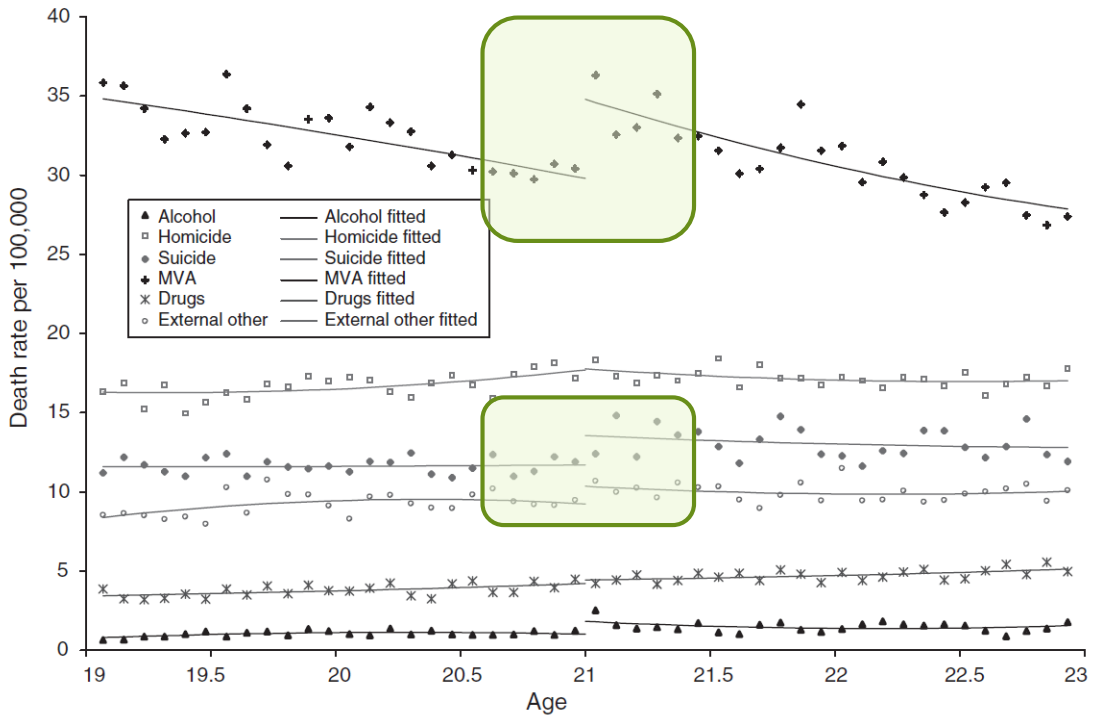


FIGURE 4. AGE PROFILES FOR DEATH RATES BY EXTERNAL CAUSE

- Figure 4 shows a large and noticeable increase in motor vehicle accidents at age 21
  - Note, there were smaller but noticeable discontinuities in alcohol-related deaths (e.g., alcohol overdoses) and suicides after age 21.

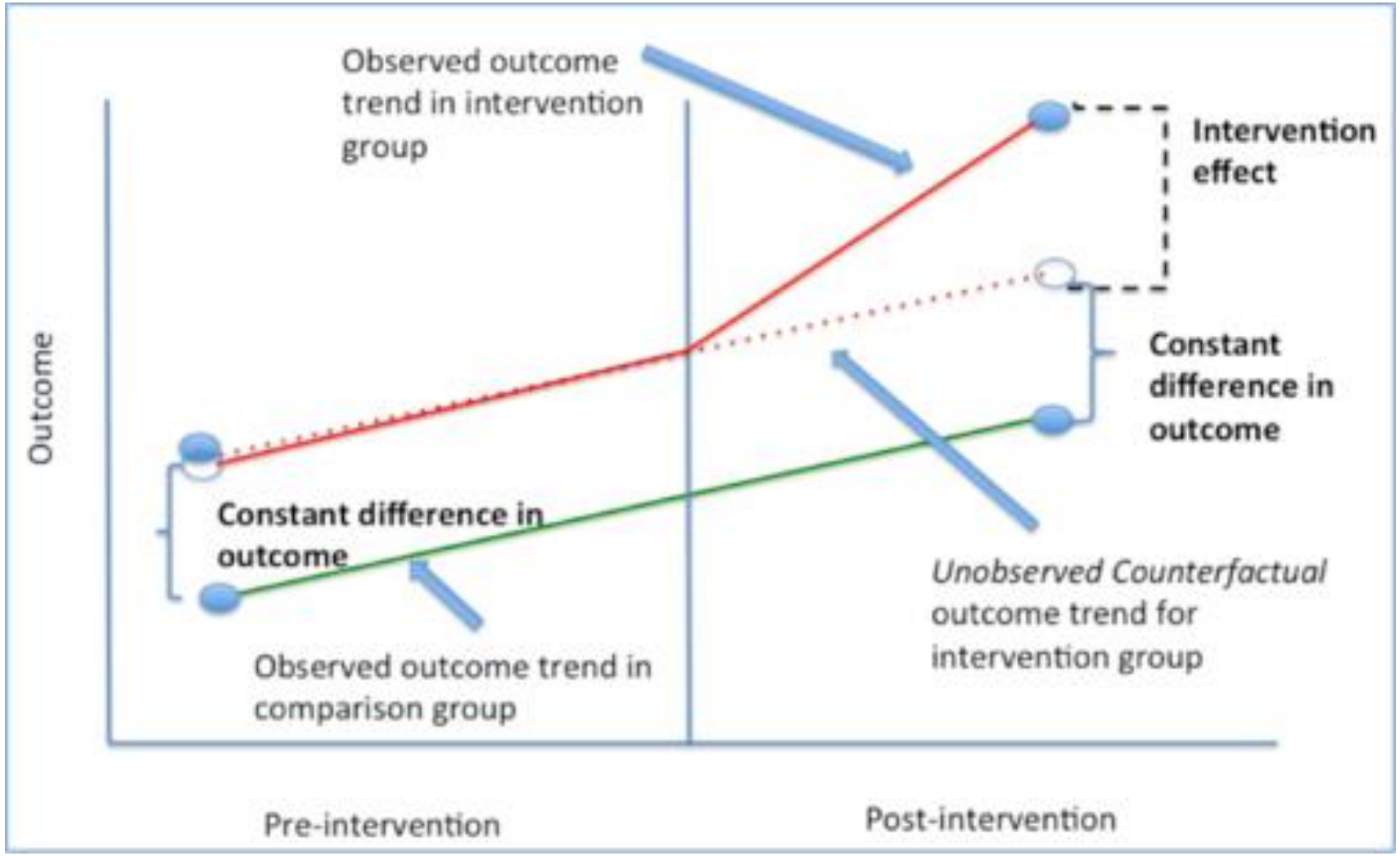
http://pubs.aeaweb.org/doi/pdfplus/10.1257/app.1.1.164

Analytical matching techniques

3. Difference-in-Difference Models

- DID evaluates the differential effect of an intervention on a 'treatment group' relative to a 'control group' in a natural experiment.
  - Compares the mean change between Time 1 and Time 2 in the outcome variable for the treatment group, compared to the mean change between Time 1 and Time 2 in the outcome variable for the control group.

- By measuring the average change for the treatment and control group, a DID attempts to mitigate the effects of confounding and selection bias.



https://www.mailman.columbia.edu/research/population-health-methods/difference-difference-estimation

Examining the impact of changes in school tobacco control policies and programs on current smoking (Leatherdale & Cole, 2015)

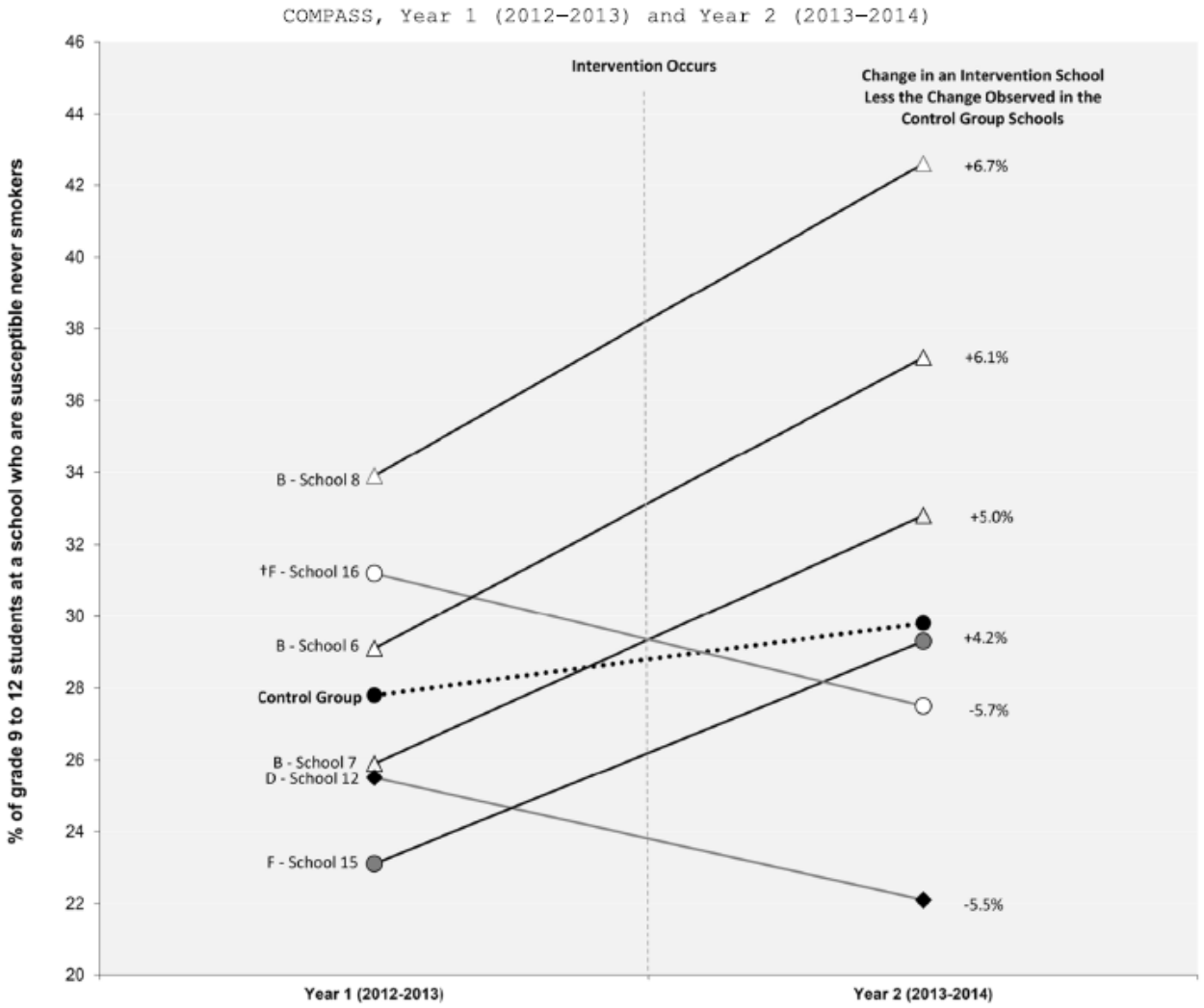
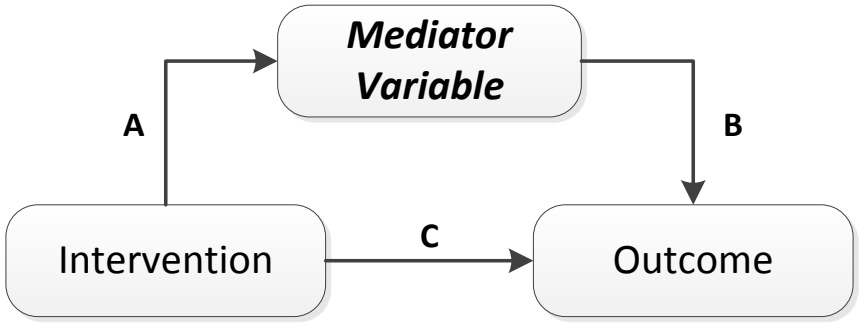


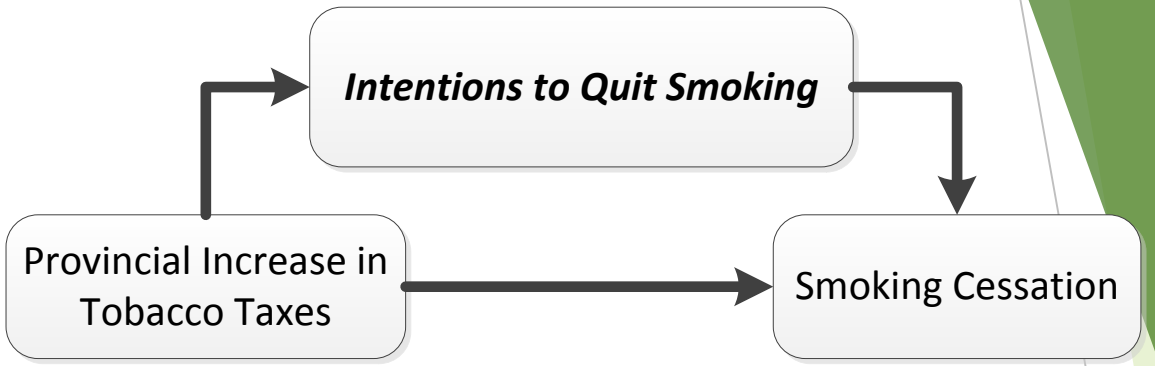
Figure 2 Difference-in-difference results of the school-level tobacco control programs or policies with significant pre-post differences in the school-level prevalence of susceptible never smokers. † Indicates a program or policy was stopped or removed. B - Effective and Enforced Tobacco Control Policies; D - Cessation Interventions; F - Aligned and Coordinated Interventions (Staff Training). Control Schools reported no changes to their tobacco control programs and policies between Y_1 and Y_2.

Mediators in a natural experiment

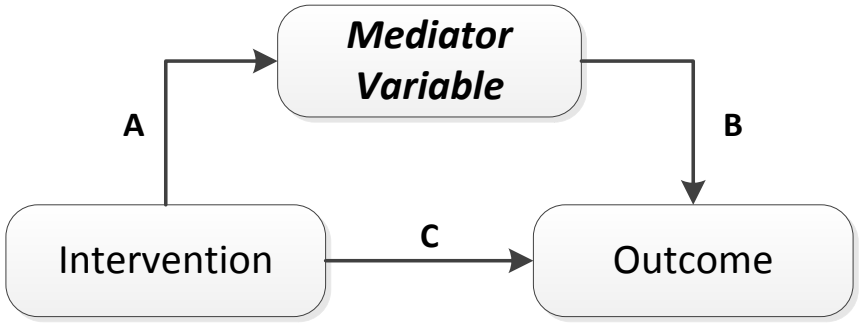
- Rather than a direct causal relationship between the intervention and the outcome, a mediation model suggests that the intervention influences the mediator variable, which in turn influences the outcome.
  - Understanding the presence and importance of a mediator variable can help researchers and stakeholders to better understand the mechanisms driving the relationship between the intervention and outcome.

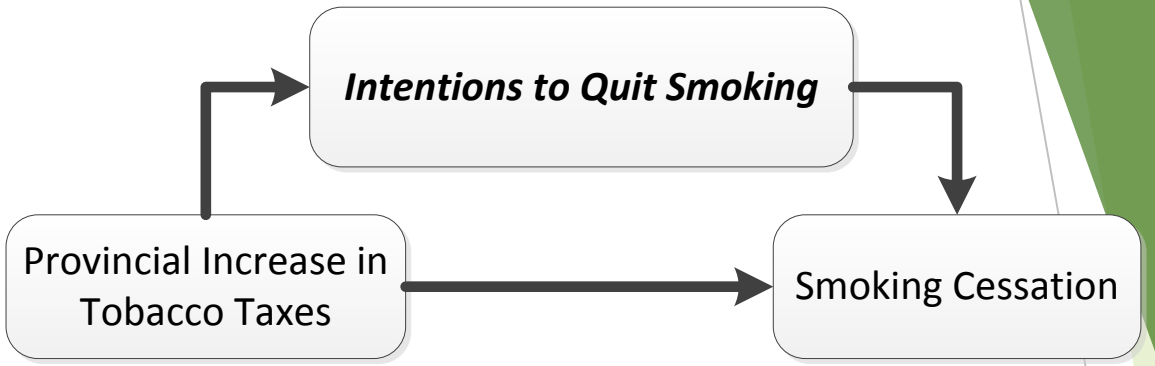
- Mediators may help to provide a more complete story of the mechanisms behind how the intervention ‘caused’ change in the outcome of interest.





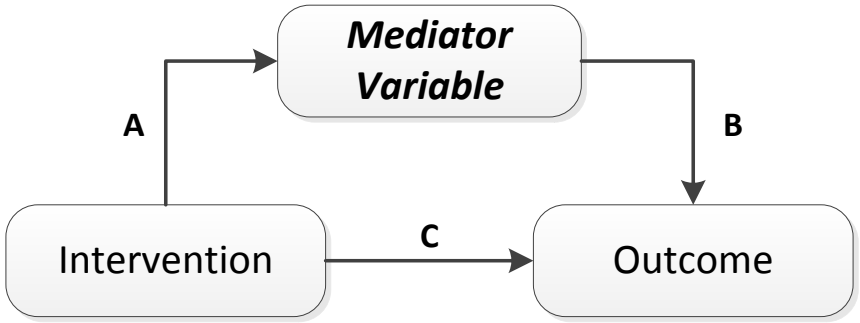
- The direct effect (C) measures the change in the outcome from the intervention value when the mediator variable remains constant.

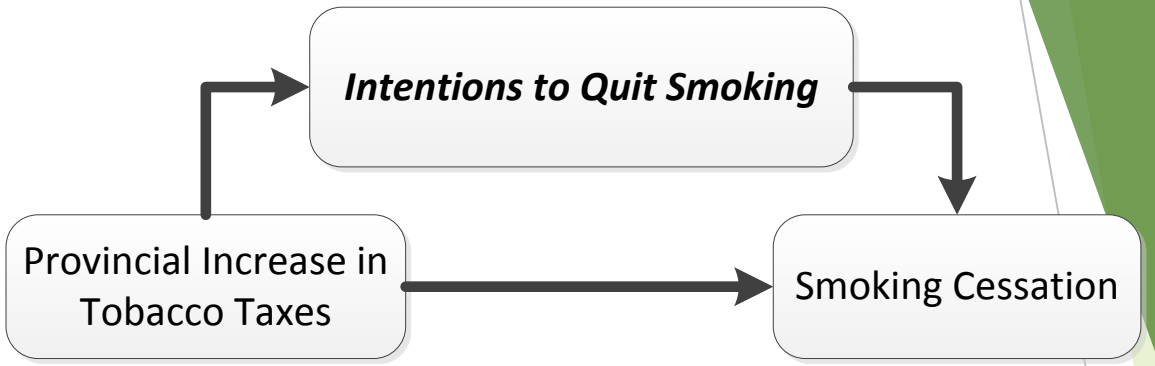




- The direct effect (C) measures the change in the outcome from the intervention value when the mediator variable remains constant.

- The indirect effect (A & B) measures the change in the outcome when the intervention value remains constant and the mediator variable changes by the amount it would have changed had the intervention value actually changed.





- There are a number of analytical approaches to test mediation, such as:
  - Linear regression

    (Baron & Kenny, 1996)
  - Logistic regression

    (MacKinnon & Dwyer, 1993)
  - Structural equation modelling

    (Kline, 2015)

Do data systems to evaluate natural experiments currently exist?

- Considering the breadth of research designs available for evaluating natural experiments, there are ample opportunities to build the evidence base moving forward using existing data systems.

- This will help to:
  - Save time
  - Save valuable research resources ($$$)
  - Fill important gaps in our understanding of how to best intervene

Do ‘IDEAL’ data systems for evaluating natural experiments currently exist?

- In Canada, we are privileged by having at least two existing data systems that are ideal for evaluating natural experiments:
  - International Tobacco Control Policy Evaluation Project [ITC]
  - The COMPASS system [COMPASS]

- These strategically placed systems allow for ***multiple***, ***ongoing***, ***timely***, ***cost-effective***, and ***robust*** evaluations in different priority domains and populations.

The International Tobacco Control Policy Evaluation Project (ITC)

- ITC is the first international research program for the systematic evaluation of key policies of the WHO Framework Convention on Tobacco Control (FCTC) at the population-level.

- The ITC Project conducts longitudinal cohort surveys in more than 28 countries to enable the robust evaluation of changes to **national tobacco control policies** resulting from the implementation of the FCTC using quasi-experimental research methods.

- For more information on the ITC methods and results, please refer to http://www.itcproject.org/

The COMPASS System (COMPASS)

- COMPASS is the world's largest and most comprehensive longitudinal natural experiment evaluation system focused on youth health.

- COMPASS collects longitudinal data from 70,000+ students and 100+ schools annually in Ontario, Alberta, Quebec, British Columbia & Nunavut.

- COMPASS data enables the robust evaluation of changes to **provincial- or local-level programs, policies and built environment resources** on youth in multiple domains:
  - Substance use (tobacco use [including e-cigarettes], alcohol use, marijuana use)
  - Obesity and correlates of obesity (poor diet, physical activity, sedentary behaviour)
  - Mental health, bullying, resilience, and academic achievement.

- For more information on COMPASS, please refer to https://uwaterloo.ca/compass-system/ .

Conclusions

- Demand for high-quality evidence of intervention impact in the domain of applied health promotion and chronic disease prevention research will continue to grow.
- The evaluation of natural experiment represents a huge untapped resource for health promotion and chronic disease prevention stakeholders.
- Evidence from natural experiments can be timely, affordable, and robust.

My motivation.

Questions?

- Thank you.