Sensitivity Analysis for Linear Structural Equation Models, Longitudinal Mediation With Latent Growth Models and Blended Learning in Biostatistics Education

Sensitivity Analysis for Linear Structural Equation Models, Longitudinal Mediation with Latent Growth Models and Blended Learning in Biostatistics Education

a dissertation presented
by
Adam John Sullivan
to
The Department of Biostatistics

in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy
in the subject of
Biostatistics

Harvard University
Cambridge, Massachusetts
May 2015

Thesis advisor: Tyler J VanderWeele Adam John Sullivan

Sensitivity Analysis for Linear Structural Equation Models, Longitudinal Mediation with Latent Growth Models and Blended Learning in Biostatistics Education

Abstract

In chapter 1, we consider the biases that may arise when an unmeasured confounder is omitted from a structural equation model (SEM) and sensitivity analysis techniques to correct for such biases. We give an analysis of which effects in an SEM are and are not biased by an unmeasured confounder. It is shown that a single unmeasured confounder will bias not just one but numerous effects in an SEM. We present sensitivity analysis techniques to correct for biases in total, direct, and indirect effects when using SEM analyses, and illustrate these techniques with a study of aging and cognitive function.

In chapter 2, we consider longitudinal mediation with latent growth curves. We define the direct and indirect effects using counterfactuals and consider the assumptions needed for identifiability of those effects. We develop models with a binary treatment/exposure followed by a model where treatment/exposure changes with time allowing for treatment/exposure-mediator interaction. We thus formalize mediation analysis with latent growth curve models using counterfactuals, makes clear the assumptions and extends these methods to allow for exposure mediator interactions. We present and illustrate the techniques with a study on Multiple Sclerosis(MS) and depression.

In chapter 3, we report on a pilot study in blended learning that took place during the Fall 2013 and Summer 2014 semesters here at Harvard. We blended the traditional BIO 200: Principles of Biostatistics and created ID 200: Principles of Biostatistics and epidemiology. We used materials from the edX course PH207x: Health in Numbers: Quantitative Methods in Clinical & Public Health Research and used. These materials were used as a video textbook in which students would watch a given number of these videos prior to class. Using surveys as well as exam data we informally assess these blended classes from the student's perspective as well as a comparison of these students with students in another course, BIO 201: Introduction to Statistical Methods in Fall 2013 as well as students from BIO 200 in Fall semesters of 1992 and 1993. We then suggest improvements upon our original course designs and follow up with an informal look at how these implemented changes affected the second offering of the newly blended ID 200 in Summer 2014.

Contents

List of Tables

Listing of figures

To my best friend and wife, Angela Sullivan, who continues to support my dreams. To my father, Tim Sullivan, whose guidance and work ethic has given me the motivation to continue on. To my mother, Bonnie Sullivan, whose unending support and guidance gave me the strength when times got tough. To my grandfather and friend, Lynn Watrous, who has been an amazing friend and teacher my whole life. To my dear friend, mentor and professor, Jake Jacobson, without your advice and support I would never have attempted a graduate degree.

Acknowledgments

Thank you to my wife, Angela Sullivan, whose patience and support has been exactly what I have needed to complete this. Thank you to my parents, Tim and Bonnie Sullivan, your amazing love and support has given me the tools and strength to accomplish anything. Thank you to my friend and grandfather, Lynn Watrous, your support and teaching has continued to propel me in life. I am grateful for every second I get to spend with you or talking with you. Thank you to the rest of my family for being such a great support and keeping me grounded. Thank you to my adviser, Tyler J VanderWeele, your guidance and patience have allowed me to grow tremendously whether it be this dissertation, as a statistician or in my career search. Thank you to my committee, Rebecca Betensky, Sebastien Haneuse and Marcello Pagano. Thank you Rebecca for encouraging me and working with me after my first attempt on my qualifying exams and for your guidance ever since then. Thank you Sebastien for your thoughtful questions and advise that help me see where I need to grow in my understanding. Thank you Marcello for giving me the opportunity to teach with you and

further biostatistics education. Thank you to the rest of the biostatistics department for the amazing opportunity to work with and know all of you. Last but not least thank you Jake Jacobson, your wisdom and support is the only reason I pursued a graduate degree.

Research reported in this publication was supported by the National Institutes of Health under award numbers T32NS048005 and R01ES017876. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

1

Sensitivity analysis for unmeasured confounding in linear structural equation models

Adam J. Sullivan and Tyler J. VanderWeele

1.1 Introduction

Linear structural equation models(LSEMs) are frequently used in many of the social sciences [6, 63]. Many of the models used are complex and contain various interrelationships. These relationships are shown in a model as paths. LSEMs allow researchers to analyze multiple paths at the same time. However it is often easy to think of variables that have been left out of a

model which may have an impact on these relationships. This impact comes in the form of biasing the effects which prompts the need for sensitivity analysis. There is a large literature on sensitivity analysis for unmeasured confounding for a single cause-effect relationship [4, 9–11, 17, 18, 23, 25, 28, 29, 32–34, 37, 39, 40, 46, 57–60, 66, 70, 73, 77]. Here we apply and extend this literature to the setting of LSEMs with many cause-effect relationships and a single unmeasured confounder.

A confounder is an extra variable that is causally related to both the dependent and the independent variable. A confounder may be called an unmeasured confounder if either no data was collected on it or it was left out of the model that was analyzed. LSEMs have very strong assumptions that are made about the functional form of the relationships, the distribution of variables and having included all confounding variables in the model. Many of the assumptions have been ignored in the models when used in practice [72]. With the strong assumptions about confounding it is important to know how sensitive the effects of interest are with respect to unmeasured confounding.

In this paper we will describe what a LSEM is and discuss the basic assumptions. Then we will consider sensitivity analysis where we will show in what circumstances and for which effects an unmeasured confounder would bias the results and what estimates are robust to the bias. We will then discuss what other effects aside from the effects of interest are biased due to unmeasured confounding. We will then give an example and discuss how to

use this sensitivity analysis technique. Then we will consider a missing path analysis in which we will discuss how the bias is affected if certain paths are absent from the model.

1.2 Brief overview of Linear Structural Equation Models

We will begin by considering a brief overview of the basics of an LSEM. We give an overview of the important concepts and language as well as a background to the methods contained in the rest of this paper.

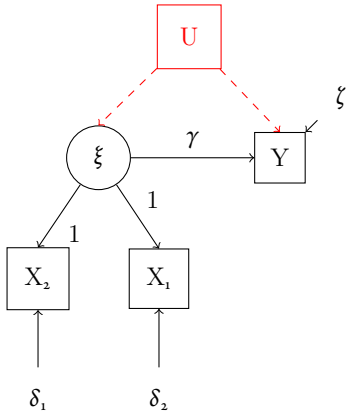


Figure 1.1: Path Diagram Example from Bollen with Newly Added U

Figure 1.1 is a simple path diagram that is shown by Bollen [6]. The model contains both observable and latent variables. Observable variables are variables that have been directly measured and are represented by squares in the figure. Latent variables are variables that are not able to be directly measured but can be inferred from observable variables in the model and are represented by circles in the figure. Relationships between these

variables are represented by the arrows which are called paths. A line with a single arrow represents a causal path, for example in figure 1.1, the arrow $\xi \rightarrow$ Y, represents that ξ has a direct effect on Y.

The final parts of figure 1.1 which have not been mentioned previously are δ_1, δ_2, and ζ. These represent random error effects on X_1, X_2 and Y respectively. Random error is the element of randomness that is not contributed by the other paths in the model. It should also be noted that unless otherwise indicated, these errors are all uncorrelated with each other and with ξ. In a LSEM, correlations are shown by using a double headed arrow path between two variables. In the rest of the path diagrams in this paper we will assume the errors are uncorrelated and leave the error terms out for simplicity. However we note that there is still error associated with each of the relationships as in figure 1.1. From the complete model in Figure 1.1 a LSEM assumes the following mathematical relationships:

$$\begin{aligned} y &= \gamma\xi + \zeta \\ x_1 &= \xi + \delta_1 \\ x_2 &= \xi + \delta_2 \end{aligned}$$

Traditionally LSEM assume that all of the above equations involve variables that are normally distributed as well as that each path follows a linear regression. Various extensions are possible for binary and ordinal variables in which it is assumed that the observed binary or categorical variables is the dichotomized or coarsened version of an underlying latent continuous normally distributed variable [6].

Suppose now we have a variable U in the model that represents a missing variable that was not accounted for in the analysis. This would bias our effect estimates for the effect of ξ on Y. This is also sometimes called an unblocked "backdoor path" [52]. If that variable had not been accounted for then even if there was no effect of ξ on Y (i.e. the arrow between the variables is missing) we would likely find an association between ξ and Y because of U. We would likely believe there is a causal relationship between ξ and Y even though there is no true association present. We need to be very mindful of this when positing models and this paper will detail steps to take in order to assess how sensitive results on an LSEM are to the impact of an unmeasured U. We will consider sensitivity analysis for total effects and for direct and indirect effects which arise when multiple variables and cause-effect relationships are being considered.

We note that in some settings an unmeasured confounder can be explicitly included in a LSEM and in certain settings it is still possible to proceed with estimation of certain effects [6]. Here we deal with the setting when the unmeasured confounder has in fact not been included in the model and the

researcher is interested in how the unmeasured confounder biases effect estimates.

1.3 Confounding and Bias of Effects

1.3.1 Path Analysis with Confounding

Whether or not there is bias for a specific effect in the context of a LSEM, will depend on whether all of the "backdoor paths" between the exposure and outcome are blocked or controlled for. In this section we will determine whether there is bias within the effect estimates that the LSEM has given us due to an incorrect model with a missing confounder.

We will consider what happens when a model is missing a single confounder. We will first consider for which paths a potential confounder might cause bias for the effect estimates. To illustrate this we will use figure 1.2. Suppose, as in figure 1.2 we have A which is an exposure, M which is a mediator, and Y which is an outcome of interest. On a more complex diagram, A, M and Y could be any three variables on the diagram for which we were interested in the various total, direct and indirect effects. Once we have chosen these three variables the analysis of bias can proceed as described below. If we change the three variables chosen as the exposure, mediator and outcome, we could proceed with a similar analysis for these three variables as well. C1 and C2 are measured covariates and U is again an unmeasured confounder. Note that in our models all of the variables are

observable; however any of these variables could be latent as well with multiple observed indicators as is typical in LSEM and this would not change the results nor the sensitivity analysis. We will illustrate the results with latent variables with multiple observed indicators in Section 1.6 below.

The direct effect is the path from the exposure, A, to the outcome, Y, which is not mediated by any other variable. This is represented by the $A \to Y$ path in figure 1.2. An indirect effect is a path from A to Y that goes through one or more other variables. This is referred to as the effect being mediated by other variables. This is represented by the path, $A \to M \to Y$ in figure 1.2. The total effect is the combination of the direct and the indirect effect. The total effect of A on Y would include both the direct path from $A \to Y$ as well as the path $A \to M \to Y$.

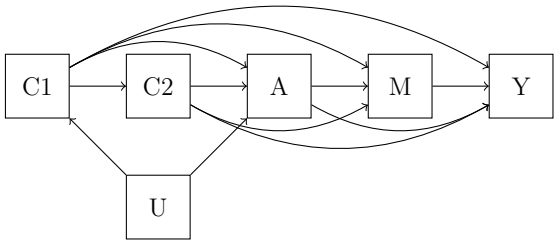


Figure 1.2: Example SEM

1.3.2 Bias of Causal Effects

In this section we will consider diagrams similar to figure 1.2. We will change the location of U in the diagram and examine how this affects the bias of the three types of effects discussed in section 1.3.1 (total, direct and indirect).

We summarize the results in Table 1.1 and formal arguments are given in the appendix. Table 1.1 is laid out so that the rows represent where the left arrow of the U variable points on the pathway from the treatment to the outcome and the columns represent where the right arrow from U points on the pathway.

Total Effects

		Right				
		Pre-Exposure	Exposure	Mediator	Outcome	Post-Outcome
Left	Pre-Exposure	Not Biased	Not Biased	Not Biased	Not Biased	Not Biased
	Exposure			Biased	Biased	Not Biased
	Mediator				Not Biased	Not Biased
	Outcome					Not Biased
	Post-Outcome					Not Biased

Direct / Indirect Effects

		Right				
		Pre-Exposure	Exposure	Mediator	Outcome	Post-Outcome
Left	Pre-Exposure	Not Biased	Not Biased	Not Biased	Not Biased	Not Biased
	Exposure			Indirect Biased	Direct Biased	Not Biased
	Mediator				Direct and Indirect Biased	Not Biased
	Outcome					Not Biased
	Post-Outcome					Not Biased

Table 1.1: Bias of Causal Effects

The cells indicate which of the effects are biased and which are unbiased in each of the various settings. Justification is given in the appendix. Cells in the table with the left arrow indicated to the right of the right arrow were left blank. In most scenarios considered in Table 1.1, the unmeasured variable would not produce bias for the total, indirect and direct effect of A on Y. This is because many of the scenarios in Table 1.1 the unmeasured variable affects two variables which occur either before the exposure or after the outcome and thus does not bias the effects of interest. Table 1.1 gives

only 3 scenarios which results in a bias of the effects of interest. Figure 1.3 shows the three interesting scenarios. The first one shown in figure 1.3a is a case of exposure-mediator confounding. This is confounding where the U has a causal relationship with both the exposure and with the mediator. With this scenario there is bias in the total effect as well as the indirect effect. There is a possibility of an association between A and M purely due to the unmeasured U. This would mean that there may in fact be no indirect effect of $A \rightarrow M \rightarrow Y$ but our estimates would show otherwise. The direct effect however is unbiased in this scenario. The second interesting scenario is shown in figure 1.3b, this is a case of exposure-outcome confounding, meaning that the unmeasured confounder has a causal relationship with the exposure and the outcome. With this scenario there is bias in the total and the direct effect; but the indirect effect is unbiased. Finally the last interesting scenario is shown in figure 1.3c. This is a case of mediator-outcome confounding. In this scenario both the direct and indirect effects are biased. In the next section we will explore the bias that is present in these scenarios and develop sensitivity analysis that can be used to assess the impact of unmeasured variables.

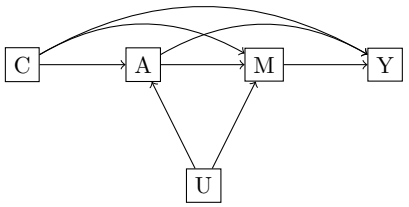
The variables in the models we have considered are very simple. However one can apply the results shown here to a LSEM of any size and complexity. By breaking down the complex model into smaller parts like these scenarios one can assess for bias due to confounding in different parts of the model. If there are intermediate variables between the variables in the diagram chosen

as the exposure A and mediator M and one of the arrows of U is pointed into one of these intermediate variables, the bias analysis would be analogous to the setting in which the arrow of U pointing into M itself. If there are intermediate variables between the variables in the diagram chosen as the mediator M and outcome Y and one of the arrows of U pointed into one of the intermediate variables, the bias analysis would be analogous to the arrow of U pointing into Y.

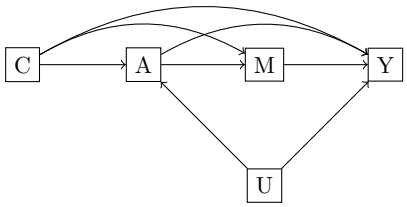
Each of the variables in figure 3 are displayed as a single variable for the purpose of simplicity. However each of these variables can represent a group of variables. For example in each of scenarios the variable C can be used to represent all of the covariates that are being adjusted for. An example of this variable grouping will be shown in section 1.6.

1.4 Scope of Bias throughout a Structural Equation Model

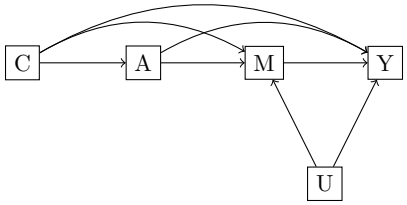
One key feature of an LSEM is the capacity to estimate the effects for any path specified in the model. In section 1.3 we considered the bias created by unmeasured confounding on the main effects of interest. This section will consider what other effects in the LSEM are biased due to the unmeasured confounding. We will show that with unmeasured confounding not just one edge and effect estimate are biased but in fact many effect estimates and even numerous distinct edges will be biased. Specifically, we show in the appendix that for any variable V that has an edge into the variable at the



(a) Exposure-Mediator Confounding



(b) Exposure-Outcome Confounding



(c) Mediator-Outcome Confounding

Figure 1.3: Scenarios with Bias Present

left most edge of the unmeasured confounder, the effect estimate of the edge from V to the variable on the right most edge of the unmeasured confounder will be biased. This means in scenario 1 any effect estimate of edges into M will be biased for any variable that affects A. In scenarios 2 and 3 any effect estimate of edges into the outcome, Y, will be biased for any variable that affects A in scenario 2 or that affects M in scenario 3.

In order to further explore bias we will consider scenario 3, mediator-outcome confounding, in more detail using simulations. The other two scenarios will be considered in the appendix. Recall from section 1.3.2 that the M $\rightarrow$ Y relationship as well as the A $\rightarrow$ Y relationship is known to be biased.

For example, simulations can be used to illustrate that the C $\rightarrow$ Y relationship is also biased. The following steps were done to simulate the data which follows the paths shown in figure 1.3c:

1. C and U are normally distributed with mean 0 and standard deviation 1.
2. 20,000 values are simulated for both C and U.
3. The effects of C$\rightarrow$ A, C$\rightarrow$ M, A$\rightarrow$ M, A$\rightarrow$ Y, U$\rightarrow$ M, U$\rightarrow$ Y and M$\rightarrow$ Y were all set to a moderate effect of 0.6.
4. The effect of C$\rightarrow$ Y was set to 0.
5. $A = 0.6C + \varepsilon_a$, where ε_a is a random error term normally distributed with mean 0 and standard deviation 1.
6. $M = 0.6C + 0.6A + 0.6U + \varepsilon_m$, where ε_m is a random error term normally distributed with mean 0 and standard deviation 1.
7. $Y = 0C + 0.6A + 0.6M + 0.6U + \varepsilon_y$, where ε_y is a random error term normally distributed with mean 0 and standard deviation 1.

This defines the paths exactly as in figure 1.3c and gives us a data set with 5 variables and 20,000 values for each variable. Then using Stata 13 the SEM was fit two ways. The first fit was having the C $\rightarrow$ Y edge in the model, knowing that this has a true effect of 0. The second model fit was leaving the C $\rightarrow$ Y edge out of the model. Results are summarized in Table 1.2.

When the model is fit allowing for the possibility of a $C \rightarrow Y$ edge (first row) the $M \rightarrow Y$ effect is upwards biased, with an estimate of 0.87 with or%

Scenario 3: Mediator-Outcome Confounding						
	C → A	C → M	A → M	C → Y	A → Y	M → Y
True Model	0.6	0.6	0.6	0	0.6	0.6
C → Y in model	0.59 (0.58, 0.61)	0.61 (0.59, 0.63)	0.58 (0.56, 0.60)	-0.17 (-0.19, -0.15)	0.44 (0.42, 0.45)	0.87 (0.86, 0.88)
C → Y not in model	0.60 (0.58, 0.61)	0.61 (0.59, 0.63)	0.58 (0.56, 0.60)	–	0.40 (0.39, 0.42)	0.82 (0.81, 0.83)

Table 1.2: SEM effect estimates from scenario 3 simulation.

confidence interval (0.86, 0.88) instead of 0.6; the $A \to Y$ effect is biased downwards with an estimate of 0.44 (0.42, 0.45) instead of 0.6. We also find that the effect of $C \to Y$ is biased downward with an estimate of -0.17 (-0.19, -0.15) instead of 0. Again this effect was set to zero in the simulations. When leaving the $C \to Y$ path out of the model the effect of $A \to Y$ and $M \to Y$ remain biased downwards and biased upwards respectively. The bias of the $A \to Y$ edge is even larger than before with estimates of 0.40 (0.39, 0.42).

Our simulations in Table 1.2 illustrates that more effects other than the direct and indirect effect are biased. An unmeasured confounder does not just bias a single edge but many edges. A single unmeasured confounder can thus introduce bias for many paths in an LSEM. Correcting biases for a LSEM single unmeasured confounder does not just require correcting one edge but many.

We will give methods in the next section for sensitivity analysis for specific total, direct or indirect effects of interest. But when using such sensitivity analysis it is recommended that in the presence of unmeasured confounding researchers also mention the possibility of the other effects that could be biased significantly as well.

1.5 Sensitivity Analysis for Structural Equation Models Under Unmeasured Confounding

1.5.1 Scenario 1: Exposure - Mediator Confounding

Consider the scenario in Figure 1.3a from Table 1.1. Here we see that the total and indirect effects both are biased. The bias comes from the possibility of an association being present based on the unblocked "backdoor" path that was mentioned in section 1.2. Our goal is to quantify the amount of bias that is present and to assess the robustness of the effects to confounding with sensitivity analysis. Formal mathematical development is given in the appendix (Section 1.9).

1.5.1.1 Total Effect Bias and Correction.

Consider the total effect under exposure-mediator confounding. This confounding will bias the $A \rightarrow M$ relationship and this in turn biases the total effect for the $A \rightarrow Y$ relationship. We will consider sensitivity analysis in this setting. Let B_{add} denote the difference between the quantity estimated by the LSEM (ignoring U) and the true causal effect of A on Y (i.e. what we would have obtained had we included U as well). Suppose that the effect of U on Y is constant across strata of A (i.e. U and A do not interact in their effects on Y, this is typically assumed in a LSEM) and the mean of U is additive in both

C and A, VanderWeele [71] and Lin [39] have shown that the bias is given by

$$B_{add} = \gamma d$$

where γ is the mean effect of U on Y, and $d = m_1 - m_0$, where m_0 and m_1 are the means of U for two different levels of A being compared. If U were binary m_0 and m_1 would be the prevalences of U in the two different levels of A being compared. The parameter γ is the estimated effect that we would see if we regressed Y on U and M. Once both γ and d are specified we can then subtract the bias γd from the estimate of the effect of A on Y from the LSEM to get a corrected estimate of the effect of A on M. We can also subtract the bias factor γd from both limits of the 95% confidence intervals in order to also get a corrected 95% confidence interval.

In general it is helpful to vary the values specified for both γ and d, as this will allow one to assess the sensitivity of the estimate of the effect of A on Y. In most cases the true values of γ and d are unknown and need to be specified. We vary our specifications for both γ and d. This leads us to having a range of values for the bias, B_{add}. Subtracting this range of values from the estimate of the total effect we then have a range of values for the correct total effect. We can then take the range of values for the bias, B_{add}, and subtract them from the upper and lower bounds of the confidence interval around the total effect in order to correct the confidence intervals as well.[1]

[1]Alternatively, it would have been possible to use sensitivity analysis to correct the effect of A on M first and then use this to get a corrected total effect as well, but obtaining corrected confidence intervals by correcting the $A \rightarrow M$ relationship first is considerably more

1.5.1.2 Indirect Effect Bias and Correction.

In section 3, we noted that the direct effect is unbiased by exposure-mediator confounding so we do not need to do sensitivity analysis for this effect. However, the indirect effect is biased by exposure-mediator confounding. In section 1.2 it was noted that the total effect is the combination of the direct and indirect effects. Since the direct effect is unbiased and since

$$\text{Total Effect} = \text{Direct Effect} + \text{Indirect Effect}$$

the bias for the indirect effect will be the same as the bias for the total effect. Once the sensitivity analysis parameters γ and d are specified, we can thus take the same bias factor γd and subtract this from the indirect effect and both limits of its confidence interval to get a corrected estimate and confidence interval for the indirect effect. We can use this approach to assess the sensitivity of the indirect effect. As we vary the parameters, we can take the range of values for the bias factor of the total effect used previously and subtract them from the indirect effect and its confidence interval. This again gives us a corrected indirect effect as well as corrected confidence interval in each case.

---

complicated, whereas the correction approach presented here for estimates and confidence intervals is relatively straightforward and as will be seen below we will likewise be able to obtain similar corrected estimates and confidence intervals for the direct and indirect effects as well.

1.5.2 Scenario 2: Exposure - Outcome Confounding

Consider the scenario in Figure 1.3b from Table 1.1 of unmeasured exposure-outcome confounding. Here we see that both the direct and total effects are biased. As noted in Table 1.1 the indirect effect is unbiased.

1.5.2.1 Direct effect Bias and Correction.

Examining the direct effect we see that the unmeasured confounding biases the direct effect $A \rightarrow Y$ relationship. For sensitivity analysis for the direct effect we can proceed in a similar manner as in Section 5.2. If U is additive in both C and A and the effect of U on Y is constant across strata of A then if we define B_{add} as the difference between our estimator using the LSEM ignoring the unmeasured confounder and the true direct effect, then we have that:

$$B_{add} = \gamma d,$$

where γ is the mean effect of U on Y and $d = (m_1 - m_0)$, where m_1 and m_0 are the mean of U for the two different levels of the exposure A being compared. We again specify values for γ and d and calculate $B_{add} = \gamma d$. This represents the bias for the direct effect of A on Y. We can again subtract this bias factor from our estimate of the direct effect and both limits of its confidence interval to obtain a corrected estimate and confidence interval. We can then vary the values of both γ and d and obtain a range of values and confidence intervals for the direct effect to assess its sensitivity to unmeasured confounding.

1.5.2.2 Total Effect Bias and Correction.

As discussed in section 3, the indirect effect is not biased by unmeasured exposure-outcome confounding, but the total effect is biased by such confounding. As in section 1.5.1.1 we know that

$$\text{Total Effect} = \text{Direct Effect} + \text{Indirect Effect}.$$

Since the indirect effect is unbiased, for any given level of the sensitivity parameters γ and d, we can take the bias factor for the direct effect found previously and subtract this from the total effect and both limits of its confidence interval to obtain a corrected total effect and confidence interval. We can do this for a range of values of the sensitivity analysis parameters for the total effect as well.

1.5.3 Scenario 3: Mediator - Outcome Confounding.

Consider the scenario in Figure 1.3c from Table 1.1 with unmeasured mediator-outcome confounding. Here we see that the direct and indirect effects are both biased by such unmeasured mediator-outcome confounding. The total effect in this scenario is however unbiased.

1.5.3.1 Direct Effect Bias and Correction.

Consider the direct effect of A on Y. If we are able to assume the direct effect of U on the outcome Y is the same for all levels of a (i.e. no interaction

between U and A) and the expected value of U is additive in A and (M, C) (i.e. $E[U|a, m, c] = g(a) + h(m, c)$ for some functions g and h) then if we define the bias factor B_{add} to be the difference between our estimator of the direct effect and the true direct effect we have [73]:

$$B_{add} = \delta\gamma$$

where δ is the difference in the mean value of U, conditional on M, between the two levels of A we are comparing and γ is the direct effect of U on the outcome Y, not through M

As before we will need to specify both δ and γ. Once we have specified a range of values for both δ and γ we have a range of values for B_{add}. To assess the sensitivity of the direct effect we subtract B_{add} from the estimate and both limits of the confidence interval. We then have obtained a corrected estimate and confidence interval for the direct effect.

1.5.3.2 Indirect Effect Bias and Correction.

We noted in section 3 that the total effect is unbiased by unmeasured mediator-outcome confounding. Since

$$\text{Total Effect} = \text{Direct Effect} + \text{Indirect Effect}.$$

the bias for the indirect effect will be of equal magnitude but opposite sign as that of the direct effect. We can thus add the bias factor for the direct effect

to the estimate and both limits of the confidence interval of the indirect effect to get a corrected indirect effect. We can do this for a range of values for the sensitivity analysis parameters so that we can assess the sensitivity of the indirect effect to the unmeasured confounding.

1.5.4 Discussion of Sensitivity Findings.

We now will consider the range of values for our effects once we have specified a range of values for the sensitivity parameters. Three things may happen with these ranges:

1. The range of the effects will contain zero.
2. The range of the effects will be in the same direction as the results of the data analysis.
3. the range of the effects is in the opposite direction as the results of the data analysis.

If the range of the effect contains zero this means that it is possible that the effect of interest is sensitive to unmeasured confounding. We would have evidence that the effect which was seen before sensitivity analysis may be due to the unmeasured confounding.

If the range of values is in the same direction of the effect originally found and does not include zero, this would indicate that the effect is relatively robust to unmeasured confounding at least over the range of sensitivity parameters considered.

If part of the range of the values is in the opposite direction of the effect originally found then it is possible that the confounding is strong enough to have changed the direction of the effect.

In some cases, for an estimate effect to be reduced to zero, very large values of the sensitivity analysis parameters may be required. This then would provide evidence that in fact the effect under study is actually present, and not entirely due to unmeasured confounding alone.

1.6 An Example

In this section we give an example of the use and interpretation of a sensitivity analysis in the context of a published study on cognitive function. This example comes for the work of Charlton et al [12]. This paper considered the relationship between age and working memory. Charlton et al [12] also evaluated whether DTI(diffusion tensor imaging) measured white matter mediated the relationship between age and information processing speed, working memory, flexibility and fluid intelligence. Charlton et al [12] contained 118 subjects ages 50-90 with mean age 70 and standard deviation 10.5. Processing speed, working memory, flexibility and fluid intelligence were assessed by standardized neuropsychological tests (see table 1 in Charlton et al for more information on these tests). Based on their study Penke & Deary [54] proposed the model which is shown in figure 1.4. This model contains one outcome called the general factor of cognitivity or g factor which is a latent variable and is inferred by processing speed, working memory,

flexibility and fluid intelligence. Their model includes a direct effect of age on the g-factor outcome as well as an effect mediated by DTI mean diffusivity.

Figure 1.4 here represents figure 1 in Penke & Deary with an added unmeasured confounder of the relationship DTI mean diffusivity and the g factor. There is the possibility, for example, of a genetic or biological factor that would lead to an increase in DTI mean diffusivity(decrease in white matter integrity [12]) and also a decrease in g factor. For the model fit by Penke and Deary, as indicated in Figure 1.4, the direct effect of age on g factor was -0.65 (-0.67, -0.62); the indirect effect of age on g factor through DTI mean diffusivity is 0.0077 (0.0077, 0.0078). The indirect effect is found by multiplying the effect of age on DTI mean diffusivity by the effect of DTI mean diffusivity on g factor ($0.77 * 0.01 = 0.0077$). All of the coefficients have been standardized, meaning that the original variables in the model were transformed into variables with mean 0 and standard deviation 1. We will use the sensitivity analysis outlined in section 1.5.3 to consider how an unmeasured mediator-outcome confounder might change these estimates

We begin by assessing the sensitivity of the direct effect. This has a bias of the form $B_{add}(m) = \delta\gamma$, where for a fixed level of DTI mean diffusivity= m, δ is the difference in prevalence of U between two ages one standard deviation apart and γ is the effect of U on the g factor. We begin with considering what values for δ and γ would suffice to eliminate the direct effect. This could be done if the difference in prevalence of U at fixed m in ages one standard deviation apart was δ=0.13, and the corresponding γ was -5. This

would lead to $B_{add} = (0.13)(-5) = -0.65$. If we subtracted this bias from the model's estimate of the direct effect, we would obtain $-0.65 - -0.65 = 0$. With a corrected 95% confidence interval of (-0.02,0.03), and this would suffice to completely eliminate the effect. Alternatively, a difference in prevalences of 0.26 with an effect of U on Y of 2.5 would likewise give a bias factor of (0.26)(-2.5)=-0.65 and suffice to eliminate the direct effect. While a difference in prevalences of 0.13 (or possibly even 0.26) might be considered plausible, an effect size of U on Y of 5 or even 2.5 standard deviations is probably unlikely. That such extreme values for γ would be required to eliminate the effect suggests that the direct effect is reasonably robust to unmeasured confounding.

In contrast, with the indirect effect, a prevalence difference of $\delta = 0.05$ and an effect of U on Y of $\gamma = -0.154$ standard deviations would give a bias factor of $B_{add} = (0.05) - (0.154) = -0.0077$ which would suffice to explain away the indirect effect. This is a much more modest scenario than we considered for the direct effect. Note that, as in section 5.3, we add this bias factor to the indirect effect to get the corrected indirect effect, while we subtracted it from the direct effect to get the corrected direct effect. If instead we specified the prevalence difference to be δ=0.05 and the effect of U on Y were $\gamma = 0.3$ standard deviations, this would would give a bias factor of $B_{add} = (0.05)(0.3) = 0.015$ and a corrected estimate and confidence interval of -.0073 (-.0073, -.0072) which was the opposite direction of the initial effect. We see that much less unnmeasured confounding is needed to eliminate the

indirect effect than was the case with the direct effect. With the indirect effect even a fairly modest amount of confounding could reverse the direction of the effect. In this example we can be fairly confident that the direct effect is robust to unmeasured confounding, but we see also that, due to the possibility of unmeasured confounding we cannot really draw conclusions about the indirect effect.

To illustrate sensitivity analysis further, we could also try specifying γ to be similar in magnitude to the effect of other variables on the g factor. For example, if we specified the effect of U on Y to be of the same magnitude as the effect of age on the g-factor, -0.65, and specified the prevalence difference to be δ=0.13 once again, we would have $B_{add} = (0.13)(-0.65) = -0.0845$. When we subtract the bias factor from our estimated direct effect and both limits of the confidence interval, we get a corrected direct effect estimate of $-0.5655(-0.5855, -0.5355)$, which is not very different from the initial estimate. When we add this bias factor to the indirect effect and both limits of its confidence interval we obtain $-0.0768(-0.0768, -0.0767)$, which again is the reverse direction of the initial indirect effect estimate.

We can also consider how unmeasured confounding would affect our direct and indirect effect estimates if the effect of U on the g factor were exactly one standard deviation. If we keep $d = 0.13$ and we let $\gamma = -1$ we have a bias factor of $B_{add} = (0.13)(-1) = -0.13$, and the corrected direct effect would be $-0.52(-0.54, -0.49)$ and the corrected indirect effect would be $-0.1223(-0.1223, -0.1222)$. If we keep $d = 0.13$ and we let $\gamma = 1$ we have a bias

factor of $B_{add} = (0.13)(1) = 0.13$, the corrected direct effect would be $-0.78(-0.8, -0.75)$ and the corrected indirect effect would be $0.1377(0.1377, 0.1378)$. With $\gamma = 1$ or $\gamma = -1$ these would be a fairly large genetic effects; however the direct effect still would not be eliminated, but the direction of the indirect effect is again reversed. Again the indirect effect is sensitive to unmeasured confounding. It is thus possible that there is no effect of DTI mean diffusivity on g factor and that any effect seen in the model is due to the unmeasured confounding.

In summary, our sensitivity analysis suggests that it is unlikely any unmeasured confounder of DTI mean diffusivity and g factor would explain away the direct effect of age on g factor. It also suggests, however, that the indirect effect of age on g factor through DTI mean diffusivity is highly sensitive to unmeasured confounding.

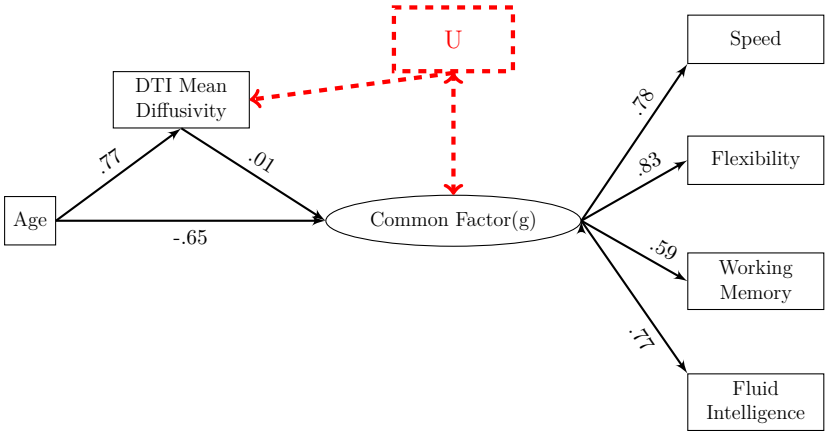


Figure 1.4: Penke and Deary: Figure 1 with Added Confounding

1.7 Missing Path Analysis

In this section we will consider what happens to the biases considered above when specific paths are absent. We will consider the three scenarios discussed in section 1.5 and remove one path at a time checking whether effects remain biased. Table 1.3 displays the results if specific edges are absent and removed from the model. The table is split into 3 sections for the total, direct and indirect effect. If we consider the total effect section of the table and then if the $A \to Y$ path is absent, we see that in the presence of exposure-mediator confounding the total effect is still biased. However had we been able to remove the $M \to Y$ path, the total effect would be unbiased. This means that if it is known that the $M \to Y$ path does not exist in the model (i.e. it is reasonable to assume that it has no effect) then the result would be that the total and direct effects are both unbiased. When removing either of these effects the indirect effect would be equal to zero.

This table will help the researcher to determine whether or not a sensitivity analysis is appropriate for a particular model. We note that removing a path should only be done if it is truly known on substantive grounds that it is absent. A key assumption that we have explored in this paper is that the LSEM is complete. This means we cannot remove a path that should be there without introducing further bias.

Table 1.3 corresponds to the biases if one of the arrows ($A \to M$, $A \to M$ or $M \to Y$)is missing but we have an unmeasured confounder not included in the

Total effect

	Missing Arrows		
	$A \to Y$	$A \to M$	$M \to Y$
Exposure-Mediator Confounding	Biased	Biased	Unbiased
Exposure-Outcome Confounding	Biased	Biased	Biased
Mediator-Outcome Confounding	Unbiased	Unbiased	Biased

Direct effect

	Missing Arrows		
	$A \to Y$	$A \to M$	$M \to Y$
Exposure-Mediator Confounding	Unbiased	Unbiased	Unbiased
Exposure-Outcome Confounding	Biased	Biased	Biased
Mediator-Outcome Confounding	Biased	Biased	Unbiased

Indirect effect

	Missing Arrows		
	$A \to Y$	$A \to M$	$M \to Y$
Exposure-Mediator Confounding	Biased	Biased	Unbiased
Exposure-Outcome Confounding	Unbiased	Unbiased	Unbiased
Mediator-Outcome Confounding	Biased	Unbiased	Biased

Table 1.3: Missing Path Analysis

model and we simply fit the SEM. However if we know for certain that one of the edges ($A \to M$, $A \to M$ or $M \to Y$) were missing then without fitting the SEM we could still estimate certain effects. For example, if there were an unmeasured exposure-mediator confounder and if we knew that the $A \to M$ edge were missing we would know the the indirect effect is 0 even though the SEM would not estimate as 0 if we ignored the unmeasured confounding.

1.8 Discussion

This paper considers sensitivity analysis for LSEMs. Three scenarios, exposure-outcome confounding, exposure-mediator confounding and mediator-outcome confounding were found to have bias associated with either the total, direct or indirect effects, as well as potentially numerous others. Section 1.5 showed a straight forward sensitivity analysis for each of the three scenarios. We also showed the result that when you have an unmeasured confounder other effects in the model will be biased as well. Specifically, for any variable V that has an edge into the variable at the left most edge of the unmeasured confounder, the effect estimate of the edge from V to the variable on the right most edge of the unmeasured confounder will be biased. This is the case for all potential variables V and thus many edges may be biased by a single unmeasured confounder. For example in exposure-outcome confounding, any other edge into the outcome will have a biased effect estimate if that variable also affects the exposure. This paper then showed situations in which the bias of the primary effects of interest would be absent if it is known that certain edges in the model were absent. However this should only be done if there is a valid reason for assuming that a specific edge does not exist. Otherwise removing an edge will leave the model incorrectly specified and the results would be biased. Theoretical explanations for the work shown in this paper are discussed in the appendix.

We want to conclude by considering the importance and impact of what

this kind of analysis can do. When unmeasured confounding is present we cannot be sure if our results are reliable. This can lead to incorrect practices and policies that are driven from the models we create and analyze. This sensitivity analysis to unmeasured confounding allows a researcher to assess the strength of the effects in the presence of confounding. If the sensitivity analysis suggests that the results are not robust to unmeasured confounding then it is possible to suggest that further research needs to be done with data collected on the confounder as well. If results do hold up under sensitivity analysis then there can be more confidence in the effects that have been estimated.

Furthermore we note that while we chosen to use simplified models with a single confounder in this paper these techniques can be applied to a model of any size and any path where the confounder lies. By breaking the model down into the smaller scenarios that we have discussed, we can use the sensitivity analysis techniques to asses the individual paths and correct for some of the bias that may be present.

1.9 Appendix

1.9.1 Arguments Concerning Table 1.1.

We use theory from causal diagrams here to demonstrate the points in the text. We refer the reader to Pearl's (2000) textbook for theory and terminology on causal diagrams. We will consider the scenarios in Table 1.1

and explain the biases that are present. With all of the models, U has been left out of the actual analysis even though it is in the figure.

For the exposure-mediator confounder, shown in figure 1.3a, we see that the total and indirect effects are biased. We condition on C in the analysis but this does not block all backdoor paths from A to Y. The reason the total effect is biased is because there exists an unblocked backdoor path from A to Y. That path would be $A \leftarrow U \rightarrow M \rightarrow Y$. If there was no effect of A on Y and of A on M, then the total effect of A on Y would be 0 , but with the path above left uncontrolled for our analysis this would show an effect of A on Y. This is why the total effect would be biased. For the indirect effect in this scenario we see that even if there were not a path from A to M, there would appear to be an indirect effect because of the path: $A \leftarrow U \rightarrow M \rightarrow Y$. However, the direct effect is unbiased in the presence of an exposure-mediator confounder (note that if there is no direct $A \leftarrow Y$ edge then A will be independent of Y conditional on M even if the unmeasured variable U is present).

For the exposure-outcome confounder, shown in figure 1.3b, we see that the total and direct effects are biased. For the total effect if there is no path from A to Y and from A to M, then there would be no total effect from A on Y. However because of U there would be a path $A \leftarrow U \rightarrow Y$ giving rise to association. This would suggest there is an effect even when there is not one. For the direct effect we see that if there is no direct path from A to Y, there is an open path $A \leftarrow U \rightarrow Y$ when conditioning on M. Again we would see an

effect even when there is not one. For the indirect effect we can see that the only path from A which goes through M is $A \rightarrow M \rightarrow Y$ and the exposure-outcome confounder would not generate bias.

For the mediator-outcome confounder, shown in figure 1.3c, we see that the direct and indirect effects are biased. For the direct effect if we condition on M we should block all paths from A to Y through M, in this case we condition on a collider and open up the path $A \rightarrow M \leftarrow U \rightarrow Y$ and would have association between A and Y conditional on M even if there were no true effect. For the indirect effect we see that even if there were not an effect of M on Y we would see association between M and Y because of the path $M \leftarrow U \rightarrow Y$,and thus if we ignore the mediator-outcome confounder, we may find an indirect effect even when it is absent. For the total effect we see that all backdoor paths from A to Y are blocked by C even if U is present.

1.9.2 Theoretical Explanation for Bias of Additional Paths

In section 1.4 it was noted that in the presence of an unmeasured confounder, U, omitted from the LSEM that for any variable V that has an edge into the variable at the left most edge of the unmeasured confounder, the effect estimate of the edge from V to the variable on the right most edge of the unmeasured confounder will be biased. In the case of exposure-outcome confounding and mediator-outcome confounding all estimated edges into the outcome will be biased for all variables that have edges into the exposure or the mediator respectively. In the case of

exposure-mediator confounding all estimated edges into the mediator will be biased for all variables with edges into the exposure as well.

We will consider exposure-mediator confounding first as in Figure 1.3a. When LSEM estimates the effect of $A \rightarrow M$ it conditions on A. This blocks the path $C \rightarrow A \rightarrow M$. However, because of U, A is a collider and when we condition on it we open up the path $C \rightarrow A \leftarrow U \rightarrow M$. Thus even if there was no direct effect of C on M we would have a non-zero estimate for the $C \rightarrow M$ edge in a LSEM which ignored U. We next consider exposure-outcome confounding which is shown in figure 1.3b. When the LSEM estimates the effect of $A \rightarrow Y$ it conditions on A. This blocks the paths $C \rightarrow A \rightarrow Y$ and $C \rightarrow A \rightarrow M \rightarrow Y$. However because A is again a collider it opens up the path $C \rightarrow A \leftarrow U \rightarrow Y$. This introduces a new unblocked path from C to Y and would bias our estimate of the effect for the edge, C on Y.

Finally, we consider mediator-outcome confounding last which is shown in figure 1.3c. When the LSEM estimates the effect of $M \rightarrow Y$ it conditions on M. This blocks the path $C \rightarrow A \rightarrow M \rightarrow Y$. However, M is now a collider on the path $C \rightarrow A \rightarrow M \leftarrow U \rightarrow Y$. This introduces a new path from C to Y and would bias our estimate of the effect for the edge of C on Y.

2

Longitudinal Mediation with Latent Growth Curve

Adam J. Sullivan, Douglas D. Gunzler, Nathan Morris, Tyler J. VanderWeele

2.1 Introduction

There is a large body of published literature on mediation analysis [31, 36, 53, 56, 67–69, 74, 75]. Almost all of this literature has considered mediation analysis for a single exposure, a single mediator and a single outcome all at one point in time. However in many studies longitudinal data is available and often not used. Instead empirical analysis often rely on the cross sectional models which do not allow for exploiting the temporal sequence of these variables. In addition, it has been shown that

cross-sectional mediation analysis typically generates substantially biased estimates of longitudinal parameters even under the ideal conditions when mediation is complete [43]. The use of longitudinal models would allow for less bias and stronger claims of causality.

In the literature there are three main types of longitudinal models currently in use. The models are the autoregressive model [16, 27], latent growth curve models [14, 21, 41, 48, 50, 76] and latent difference score models [22, 44, 45]. In this paper the focus is on advancing the methodology of mediation with latent growth curve models. We make three major contributions to the literature. We put the models into a formal causal framework so that they may be accurately used to make causal inferences. We then clarify the assumptions needed in order to make causal inferences. Finally we extend existing methodology to allow for interaction to be assessed with these models.

We first consider a latent growth curve model with binary treatment/exposure. With this model we consider the assumptions needed for identifiabilty of the direct and indirect effects. We define the direct and indirect effects, using counterfactuals, in the presence of interaction. We then consider the scenario where there is a longitudinal treatment/exposure. We consider the assumptions needed for identification and define the direct and indirect effects using counterfactuals which allows for the presence of interaction. We finish this paper with an data analysis example.

2.2 Definition of Model

When there is repeated measures data for the mediator and outcome, mediation models can be fit using latent growth curve (LGC) modeling [14, 21, 41, 48, 50, 76]. We use the parallel process model as shown by MacKinnon [41] in which separate growth curves are specified for the mediator and outcome. The treatment/exposure can also have a specified growth curve, or as with a randomized trial, it can be binary. With these growth models there are latent factors included. The first of these factors is the intercept or average baseline of the subjects at the first measurement occasion. The second factor is the slope or the trajectory of the growth after the first measurement occasion. When using these models in the mediation setting we examine the mediating relationships of these latent factors among the growth models.

We begin with Model 1 shown in Figure 2.1. We have a binary treatment, X_i; longitudinal mediator, M_1, M_2 and M_3; and a longitudinal outcome, Y_1, Y_2 and Y_3. With this model X_i affects both the intercept and slope of the mediator and outcome growth models. The intercept and the slope of the mediator growth model also both effect the intercept and the slope of the outcome growth model.

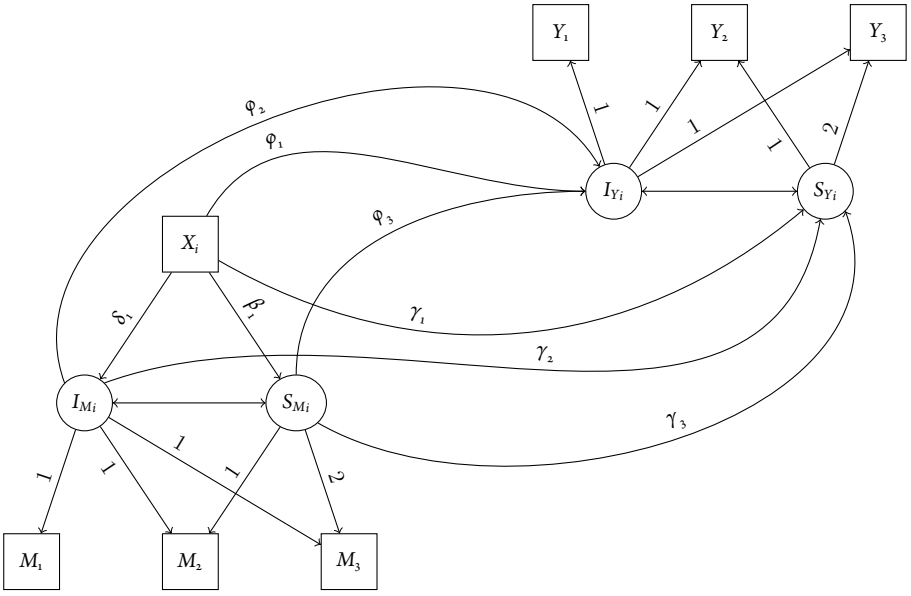


Figure 2.1: Model 1: Without Interaction, covariates C left out for simplicity

More formally, equations (2.1) - (2.6) below specify the relationships shown in Figure 2.1 with $1, \ldots, t$ measurement occasions. We have the following growth curve for the mediator:

$$M_{it} = I_{Mi} + S_{Mi}t + \varepsilon_{Mit} \tag{2.1}$$

$$I_{Mi} = \delta_0 + \delta_1 X_i + \delta_2' C + \nu_{I_{Mi}} \tag{2.2}$$

$$S_{Mi} = \beta_0 + \beta_1 X_i + \beta_2' C + \nu_{S_{Mi}} \tag{2.3}$$

and the following growth curve for the outcome:

$$Y_{it} = I_{Yi} + S_{YI}t + \varepsilon_{Yit} \tag{2.4}$$

$$I_{Yi} = \varphi_0 + \varphi_1 X_i + \varphi_2 I_{Mi} + \varphi_3 S_{Mi} + \varphi_4 X_i I_{Mi} + \varphi_5 X_i S_{Mi} + \varphi_6' C + \nu_{I_{Yi}} \tag{2.5}$$

$$S_{Yi} = \gamma_0 + \gamma_1 X_i + \gamma_2 I_{Mi} + \gamma_3 S_{Mi} + \gamma_4 X_i I_{Mi} + \gamma_5 X_i S_{Mi} + \gamma_6' C + \nu_{S_{Yi}} \tag{2.6}$$

Where $\mathrm{E}[\varepsilon_{Mit}] = \mathrm{E}[\varepsilon_{Yit}] = \mathrm{E}[\nu_{I_{Mi}}] = \mathrm{E}[\nu_{S_{Mi}}] = \mathrm{E}[\nu_{I_{Yi}}] = \mathrm{E}[\nu_{S_{Yi}}] = 0$ and where $\varepsilon_{Mit}, \varepsilon_{Yit}, (\nu_{I_{Mi}}, \nu_{S_{Mi}}) and (\nu_{I_{Yi}}, \nu_{S_{Yi}})$ are mutually independent and where C denotes baseline covariates which, as discussed below we select to represent the set of exposure-mediator, exposure-outcome and mediator-outcome confounders.

Equations 2.1 and 2.4 specify the growth models for individual i's mediator and outcome data respectively at time t. Both include an intercept, slope and error component. Equations 2.2-2.3 and 2.5-2.6 specify the intercept and slope functions for the mediator and outcome models respectively. Note that equations 2.5 and 2.6 allow for treatment/exposure-mediator interaction. In the absence interaction we specify $\varphi_4 = \varphi_5 = \gamma_4 = \gamma_5 = 0$.

We use counterfactual notation $I_{Yxm_1m_2}$, $S_{Yxm_1m_2}$, I_{Mx} and S_{Mx}, where $I_{Yxm_1m_2}$ denotes the value of the intercept model for Y if we were to set $X = x$, $I_M = m_1$

and $S_M = m_2$; $S_{Yxm_1m_2}$ denotes the value of the slope model for Y if we were to set $X = x$, $I_M = m_1$ and $S_M = m_2$; I_{Mx} denotes the value of the intercept model for M if we were to set $X = x$ and S_{Mx} denotes the value of the slope model for M if we were to set $X = x$. We use $Y_{xm_1m_2}$ to denote the counterfactual outcome Y if we were to set $X = x$, $I_M = m_1$ and $S_M = m_2$. The natural direct effect for two values of the exposure, x and x^*, is defined as $\mathrm{E}\left[Y_{xI_{Mx^*}S_{Mx^*}} - Y_{x^*I_{Mx^*}S_{Mx^*}}\right]$ and expresses how much the intercept and slope of the outcome process would change on average if the treatment/exposure were changed from level $x^* = 0$ to $x = 1$ but for each individual the intercept and slope of the mediator process is kept at the level it would have taken under the absence of the treatment/exposure. The natural indirect effect for two values of the exposure, x and x^*, is defined as $\mathrm{E}\left[Y_{xI_{Mx}S_{Mx}} - Y_{xI_{Mx^*}S_{Mx^*}}\right]$ and expresses how much the intercept and slope of the outcome process would change on average if the treatment/exposure was controlled at level $x = 1$ but the intercept and slope of the mediator process were changed from the level they would take if the treatment/exposure was changed from $x^* = 0$ to $x = 1$.

We let $A \amalg B|C$ denote that A is independent of B conditional on C. We show below that the natural direct and indirect effects are identified if:

(C1) $I_{Y_{I_M S_M}}, S_{Y_{I_M S_M}} \amalg X|C$ (no unmeasured confounding for the exposure-outcome relationship)

(C2) $I_{Y_{I_M S_M}}, S_{Y_{I_M S_M}} \amalg I_M, S_M|X, C$ (no unmeasured confounding for the mediator-outcome relationship)

(C3) $I_{M_x}, S_{M_x} \amalg X|C$ (no unmeasured confounding for the exposure-mediator relationship)

(C4) $I_{Y_{m_1,m_2}}, S_{Y_{m_1,m_2}} \amalg I_{M_{x^*}}, S_{M_{x^*}}|C$ (no mediator-outcome confounders which are affected by the exposure)

Proposition: For any function u if (C1) - (C4) hold then
$\mathrm{E}\left[u\left(I_{Y_{xI_{M_{x^*}}S_{M_{x^*}}}}, S_{Y_{xI_{M_{x^*}}S_{M_{x^*}}}}\right)\right] = \sum_{c,m_1,m_2} \mathrm{E}\left[u\left(I_Y, S_Y\right)|x, m_1, m_2, c\right] \Pr\left(m_1, m_2|x^*, c\right) \Pr(c)$

Proof:

$$\mathrm{E}\left[u\left(I_{Yx I_{M_{x^*}} S_{M_{x^*}}}, S_{Yx I_{M_{x^*}} S_{M_{x^*}}}\right)\right] = \sum_c \mathrm{E}\left[u\left(I_{Yx I_{M_{x^*}} S_{M_{x^*}}}, S_{Yx I_{M_{x^*}} S_{M_{x^*}}}\right)|C=c\right]\Pr(C=c)$$

(Iterated Expectations)

$$= \sum_{c,m_1,m_2} \mathrm{E}\left[u\left(I_{Yxm_1m_2}, S_{Yxm_1m_2}\right)|C=c, I_{Mx^*}=m_1, S_{Mx^*}=m_2\right]\Pr\left(I_{Mx^*}=m_1, S_{Mx^*}=m_2|C=c\right)\Pr(C=c)$$

(Iterated Expectations)

$$= \sum_{c,m_1,m_2} \mathrm{E}\left[u\left(I_{Yxm_1m_2}, S_{Yxm_1m_2}\right)|C=c\right]\Pr\left(I_{Mx^*}=m_1, S_{Mx^*}=m_2|X=x^*, C=c\right)\Pr(C=c)$$

(C4 & C3)

$$= \sum_{c,m_1,m_2} \mathrm{E}\left[u\left(I_{Yxm_1m_2}, S_{Yxm_1m_2}\right)|X=x, C=c\right]\Pr\left(I_M=m_1, S_M=m_2|X=x^*, C=c\right)\Pr(C=c)$$

(C1 & consistency)

$$= \sum_{c,m_1,m_2} \mathrm{E}\left[u\left(I_{Yxm_1m_2}, S_{Yxm_1m_2}\right)|X=x, I_M=m_1, S_M=m_2, C=c\right]\Pr\left(I_M=m_1, S_M=m_2|X=x^*, C=c\right)\Pr(C=c)$$

(C2)

$$= \sum_{c,m_1,m_2} \mathrm{E}\left[u\left(I_Y, S_Y\right)|X=x, I_M=m_1, S_M=m_2, C=c\right]\Pr\left(I_M=m_1, S_M=m_2|X=x^*, C=c\right)\Pr(C=c)$$

(consistency)

$$= \sum_{c,m_1,m_2} \mathrm{E}\left[u\left(I_Y, S_Y\right)|x, m_1, m_2, c\right]\Pr\left(m_1, m_2|x^*, c\right)\Pr(c)$$

This completes the proof.

Then if we replace x with x^* we get:

$$\mathrm{E}\left[u\left(I_{Yx^*I_{Mx^*}S_{Mx^*}}, S_{Yx^*I_{Mx^*}S_{Mx^*}}\right)\right] = \textstyle\sum_{c,m_1,m_2} \mathrm{E}\left[S_Y|x^*, m_1, m_2, c\right]\Pr\left(m_1, m_2|x^*, c\right)\Pr(c)$$

from this it follows with $u(I_Y, S_Y) = I_Y + S_Y t + \varepsilon_Y$ that the average natural direct effect is given by:

$$\mathrm{E}\left[u\left(I_{Yx I_{Mx^*}S_{Mx^*}}, S_{Yx I_{Mx^*}S_{Mx^*}}\right) - u\left(I_{Yx^*I_{Mx^*}S_{Mx^*}}, S_{Yx^*I_{Mx^*}S_{Mx^*}}\right)\right]$$
$$= \sum_{c,m_1,m_2} \left\{\mathrm{E}\left[u\left(I_Y, S_Y\right)|x, m_1, m_2, c\right] - \mathrm{E}\left[u\left(I_Y, S_Y\right)|x^*, m_1, m_2, c\right]\right\}\Pr\left(m_1, m_2|x^*, c\right)\Pr(c)$$

If we replace x^* with x we would get:

$$\mathrm{E}\left[u\left(I_{Yx I_{M_x} S_{M_x}}, S_{Y x I_{M_x} S_{M_x}}\right)\right] = \sum_{c,m_1,m_2} \mathrm{E}\left[u) I_Y, S_Y\right) | x, m_1, m_2, c] \Pr\left(m_1, m_2 | x, c\right) \Pr(c)$$

from this

it follows with $u(I_Y, S_Y) = I_Y + S_Y t + \varepsilon_Y$ that the natural indirect effect is given by:

$$\begin{aligned}
&\mathrm{E}\left[u\left(I_{Y x I_{M_x} S_{M_x}}, S_{Y x I_{M_x} S_{M_x}}\right) - u\left(I_{Y x I_{M_{x^*}} S_{M_{x^*}}}, S_{Y x I_{M_{x^*}} S_{M_{x^*}}}\right)\right] \\
&\quad = \sum_{c,m_1,m_2} \mathrm{E}\left[u\left(I_Y, S_Y\right) | x, m_1, m_2, c\right] \left\{\Pr\left(m_1, m_2 | x, c\right) - \Pr\left(m_1, m_2 | x^*, c\right)\right\} \Pr(c) \\
&\quad = \sum_{c,m_1,m_2} \mathrm{E}\left[u\left(I_Y, S_Y\right) | x, m_1, m_2, c\right] \Pr\left(m_1, m_2 | x, c\right) \Pr(c) - \mathrm{E}\left[u\left(I_Y, S_Y\right) | x, m_1, m_2, c\right] \Pr\left(m_1, m_2 | x^*, c\right) \Pr(c)
\end{aligned}$$

With the model shown in Figure 2.1 we have that

$Y = u\left(I_Y, S_Y\right) = I_Y + S_Y t + \varepsilon_Y$. Thus given (2.5) and (2.6) we have

$$\begin{aligned}
\mathrm{E}\left[u\left(I_Y, S_Y\right) | x, m_1, m_2, c\right] &= \varphi_0 + \varphi_1 x + \varphi_2 m_1 + \varphi_3 m_2 + \varphi_4 x m_1 + \varphi_5 x m_2 + \varphi_6' c \\
&\quad + (\gamma_0 + \gamma_1 x + \gamma_2 m_1 + \gamma_3 m_2 + \gamma_4 x m_1 + \gamma_5 x m_2 + \gamma_6' c) t \quad (2.7)
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{E}\left[u\left(I_Y, S_Y\right) | x^*, m_1, m_2, c\right] &= \varphi_0 + \varphi_1 x^* + \varphi_2 m_1 + \varphi_3 m_2 + \varphi_4 x^* m_1 + \varphi_5 x^* m_2 + \varphi_6' c \\
&\quad + (\gamma_0 + \gamma_1 x^* + \gamma_2 m_1 + \gamma_3 m_2 + \gamma_4 x^* m_1 + \gamma_5 x^* m_2 + \gamma_6' c) t \quad (2.8)
\end{aligned}$$

Therefore the average natural direct effect is

$$\sum_{c,m_1,m_2} \left\{ (\varphi_1 + \varphi_4 m_1 + \varphi_5 m_2 + \gamma_1 t + \gamma_4 m_1 t + \gamma_5 m_2 t)(x - x^*) \right\} \Pr(m_1, m_2 | x^*, c) \Pr(c)$$

$$= (\varphi_1 + \varphi_4 \mathrm{E}[M_1 | x^*, c] + \varphi_5 \mathrm{E}[M_2 | x^*, c] + \gamma_1 t + \gamma_4 \mathrm{E}[M_1 | x^*, c] t + \gamma_5 \mathrm{E}[M_2 | x^*, c] t)(x - x^*)$$

$$= (\varphi_1 + \varphi_4(\delta_0 + \delta_1 x^* + \delta_3' c) + \varphi_5(\beta_0 + \beta_1 x^* + \beta_3' c) + \gamma_1 t + \gamma_4(\delta_0 + \delta_1 x^* + \delta_2' c)t + \gamma_5(\beta_0 + \beta_1 x^* + \beta_2' c)t)(x - x^*)$$

Given (2.2), (2.3), (2.5) and (2.6) we have

$$\begin{aligned} \mathrm{E}\left[u\left(I_{Yx I_{M_x} S_{M_x}}, S_{Yx I_{M_x} S_{M_x}}\right)\right] &= (\varphi_2 + \gamma_2 t)\delta_0 + (\varphi_3 + \gamma_3 t)\beta_0 + \varphi_0 + \gamma_0 t \\ &+ (\varphi_1 + \gamma_1 t + (\varphi_2 + \gamma_2 t)\delta_1 + (\varphi_3 + \gamma_3 t)\beta_1 + (\varphi_4 + \gamma_4 t)(\delta_0 + \delta_2' c) \\ &+ (\varphi_5 + \gamma_5 t)(\beta_0 + \beta_2' c))x \\ &+ ((\varphi_4 + \gamma_4 t)\delta_1 + (\varphi_5 + \gamma_5 t)\beta_1)x^2 + (\varphi_6' + \gamma_6' t + (\varphi_2 + \gamma_2 t)\delta_2' \\ &+ (\varphi_3 + \gamma_3 t)\beta_2')c \end{aligned}$$

$$
\begin{aligned}
\mathrm{E}\left[u\left(I_{Y_{x}I_{M_{x^*}}S_{M_{x^*}}}, S_{Y_{x}I_{M_{x^*}}S_{M_{x^*}}}\right)\right] &= (\varphi_2 + \gamma_2 t)\delta_0 + (\varphi_3 + \gamma_3 t)\beta_0 + \varphi_0 + \gamma_0 t \\
&+ (\varphi_1 + \gamma_1 t + (\varphi_4 + \gamma_4 t)(\delta_0 + \delta_2' c) + (\varphi_5 + \gamma_5 t)(\beta_0 + \beta_2' c))x \\
&+ ((\varphi_2 + \gamma_2 t)\delta_1 + (\varphi_3 + \gamma_3 t)\beta_1)x^* \\
&+ ((\varphi_4 + \gamma_4 t)\delta_1 + (\varphi_5 + \gamma_5 t)\beta_1)xx^* + (\varphi_6' + \gamma_6' t + (\varphi_2 + \gamma_2 t)\delta_2' + (\varphi_3 + \gamma_3 t)\beta_2')c
\end{aligned}
$$

Therefore the average natural indirect effect is

$$
\begin{aligned}
&\sum_{c,m_1,m_2} \mathrm{E}\left[u\left(I_Y, S_Y\right)|x, m_1, m_2, c\right] \Pr\left(m_1, m_2|x, c\right) \Pr(c) \\
&\qquad - \mathrm{E}\left[u\left(I_Y, S_Y\right)|x, m_1, m_2, c\right] \Pr\left(m_1, m_2|x^*, c\right) \Pr(c) \\
&= (\varphi_1 + \gamma_1 t + (\varphi_2 + \gamma_2 t)\delta_1 + (\varphi_3 + \gamma_3 t)\beta_1 + (\varphi_4 + \gamma_4 t)(\delta_0 + \delta_2' c) \\
&\qquad + (\varphi_5 + \gamma_5 t)(\beta_0 + \beta_2' c))x + ((\varphi_4 + \gamma_4 t)\delta_1 + (\varphi_5 + \gamma_5 t)\beta_1)x^2 \\
&\qquad - ((\varphi_4 + \gamma_4 t)\delta_1 + (\varphi_5 + \gamma_5 t)\beta_1)xx^* \\
&\qquad - (\varphi_1 + \gamma_1 t + (\varphi_4 + \gamma_4 t)(\delta_0 + \delta_2' c) + (\varphi_5 + \gamma_5 t)(\beta_0 + \beta_2' c))x \\
&\qquad - ((\varphi_2 + \gamma_2 t)\delta_1 + (\varphi_3 + \gamma_3 t)\beta_1)x^* \\
&= ((\varphi_2 + \gamma_2 t)\delta_1 + (\varphi_3 + \gamma_3 t)\beta_1)(x - x^*) \\
&\qquad + ((\varphi_4 + \gamma_4 t)\delta_1 + (\varphi_5 + \gamma_5 t)\beta_1)(x^2 - xx^*)
\end{aligned}
$$

As discussed previously in the absence of interaction we specify $\varphi_4 = \varphi_5 = \gamma_4 = \gamma_5 = 0$. This leads to the following direct effect:

$$\sum_{c,m_1,m_2} \left\{ (\varphi_1 + \varphi_4 m_1 + \varphi_5 m_2 + \gamma_1 t + \gamma_4 m_1 t + \gamma_5 m_2 t)(x - x^*) \right\} \Pr(m_1, m_2 | x^*, c) \Pr(c)$$
$$= (\varphi_1 + \gamma_1 t)(x - x^*)$$

and the following indirect effect:

$$\sum_{c,m_1,m_2} \mathrm{E}\left[u\left(I_Y, S_Y\right) | x, m_1, m_2, c\right] \Pr(m_1, m_2 | x, c) \Pr(c)$$
$$- \mathrm{E}\left[u\left(I_Y, S_Y\right) | x, m_1, m_2, c\right] \Pr(m_1, m_2 | x^*, c) \Pr(c)$$
$$= ((\varphi_2 + \gamma_2 t)\delta_1 + (\varphi_3 + \gamma_3 t)\beta_1)(x - x^*)$$

The equations modeled here differ from that of the ones presented by MacKinnon [41] in that the intercept of the outcome is not a cause of the slope of the mediator. This follows because if I_Y affected S_M then confounding assumption (C4) would be violated because I_Y would be a mediator-outcome confounder (i.e. a common cause of S_M and Y) that was itself affected by the exposure. These equations here unlike those of MacKinnon also allow for exposure/treatment-mediator interaction.

2.3 Model with Growth Curve for Treatment/Exposure

In Section 2.2 the models were developed under the assumption of a binary treatment/exposure. This is often the case in randomized trials. In this section we consider the model displayed in Figure 2.2. This model allows for the treatment/exposure to change with time and fits a growth curve for this as well.

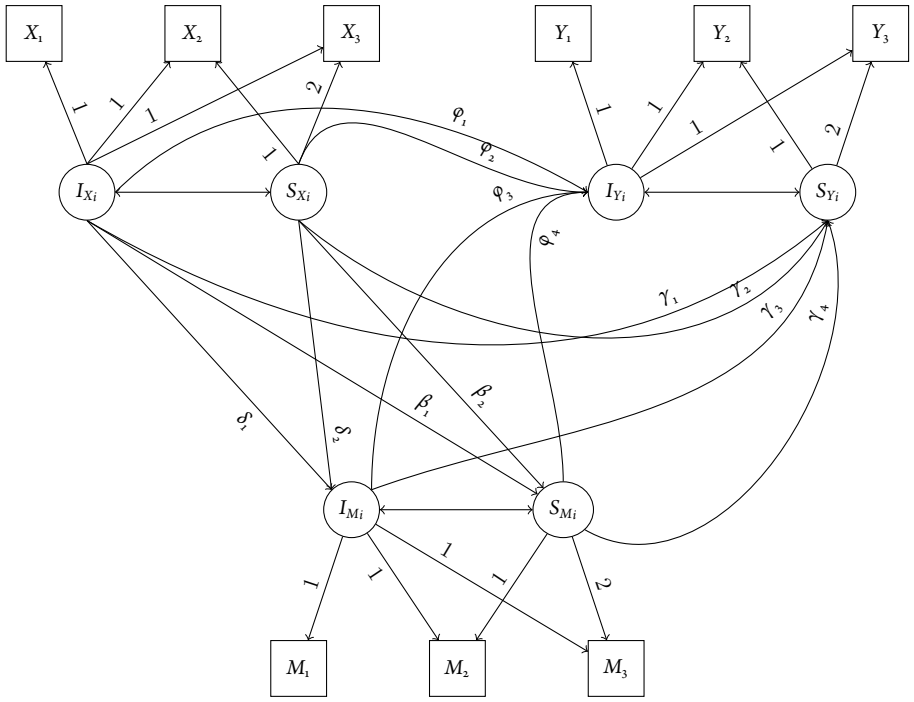


Figure 2.2: Model 2: Without Interaction, covariates C left out for simplicity

More formally, equations (2.9) - (2.17) specify the relationships shown in Figure 2.2 with $1, \ldots, t$ measurement occasions. We have the following growth curve for the treatment/exposure:

$$X_{it} = I_{Xi} + S_{Xi}t + \varepsilon_{Xit} \tag{2.9}$$

$$I_{Xi} = \rho_0 + \nu_{I_{Xi}} \tag{2.10}$$

$$S_{Xi} = \lambda_0 + \nu_{S_{Xi}} \tag{2.11}$$

the following growth curve for the mediator:

$$M_{it} = I_{Mi} + S_{Mi}t + \varepsilon_{Mit} \tag{2.12}$$

$$I_{Mi} = \delta_{o} + \delta_{1}I_{Xi} + \delta_{2}S_{Xi} + \delta_{3}'C + \nu_{I_{Mi}} \tag{2.13}$$

$$S_{Mi} = \beta_{o} + \beta_{1}I_{Xi} + \beta_{2}S_{Xi} + \beta_{3}'C + \nu_{I_{Mi}} \tag{2.14}$$

and the following growth curve for the outcome:

$$Y_{it} = I_{Yi} + S_{YI}t + \varepsilon_{Yit} \tag{2.15}$$

$$I_{Yi} = \varphi_0 + \varphi_1 I_{Xi} + \varphi_2 S_{Xi} + \varphi_3 I_{Mi} + \varphi_4 S_{Mi} + \varphi_5 I_{Xi} I_{Mi} + \varphi_6 I_{Xi} S_{Mi} + \varphi_7 S_{Xi} I_{Mi} + \varphi_8 S_{Xi} S_{Mi} + \varphi_9' C + \nu_{I_{Yi}} \tag{2.16}$$

$$S_{Yi} = \gamma_0 + \gamma_1 I_{Xi} + \gamma_2 S_{Xi} + \gamma_3 I_{Mi} + \gamma_4 S_{Mi} + \gamma_5 I_{Xi} I_{Mi} + \gamma_6 I_{Xi} S_{Mi} + \gamma_7 S_{Xi} I_{Mi} + \gamma_8 S_{Xi} S_{Mi} + \gamma_9' C + \nu_{S_{Yi}} \tag{2.17}$$

Where $\mathrm{E}[\varepsilon_{Xit}] = \mathrm{E}[\varepsilon_{Mit}] = \mathrm{E}[\varepsilon_{Yit}] = \mathrm{E}[\nu_{I_{Xi}}] = \mathrm{E}[\nu_{S_{Xi}}] = \mathrm{E}[\nu_{I_{Mi}}] = \mathrm{E}[\nu_{S_{Mi}}] = \mathrm{E}[\nu_{I_{Yi}}] = \mathrm{E}[\nu_{S_{Yi}}] = 0$ and where $\varepsilon_{X_{it}}, \varepsilon_{M_{it}}, \varepsilon_{Y_{it}}, (\nu_{I_{Xi}}, \nu_{S_{Xi}}), (\nu_{I_{Mi}}, \nu_{S_{Mi}}) and (\nu_{I_{Yi}}, \nu_{S_{Yi}})$ are mutually independent and where C denotes baseline covariates which as discussed below we select to represent exposure-mediator, exposure-outcome and mediator-outcome confounders.

Equations 2.9, 2.12 and 2.15 specify the growth model for individual i's treatment/exposure, mediator and outcome respectively. Equations 2.10-2.11, 2.13-2.14 and 2.16-2.17 specify the intercept and slope for the exposure, mediator and outcome respectively. Note that equations 2.16 and 2.17 allow for exposure-mediator interaction. In the absence of interaction we specify $\varphi_5 = \varphi_6 = \varphi_7 = \varphi_8 = \gamma_5 = \gamma_6 = \gamma_7 = \gamma_8 = 0$.

We use counterfactual notation $I_{Yx_1x_2m_1m_2}$, $S_{Yx_1x_2m_1m_2}$, $I_{Mx_1x_2}$ and $S_{Mx_1x_2}$, where $I_{Yx_1x_2m_1m_2}$ denotes the value of the intercept model for Y if we were to set $I_X = x_1$, $S_X = x_2$, $I_M = m_1$ and $S_M = m_2$; $S_{Yx_1x_2m_1m_2}$ denotes the value of the slope model for Y if

we were to set $I_X = x_1$, $S_X = x_2$, $I_M = m_1$ and $S_M = m_2$; $I_{M_{x_1x_2}}$ denotes the value of the intercept model for M if we were to set $I_X = x_1$ and $S_X = x_2$ and $S_{M_{x_1x_2}}$ denotes the value of the slope model for M if we were to set $X = x$. We use $Y_{x_1x_2m_1m_2}$ to denote the counterfactual outcome Y if we were to set $I_X = x_1$, $S_X = x_2$, $I_M = m_1$ and $S_M = m_2$. The natural direct effect for two values of the intercept function of the treatment/exposure x_1 and x_1^* and for two values of the slope function of the treatment/exposure x_2 and x_2^*, is defined as $\mathrm{E}\left[Y_{x_1x_2I_{M_{x_1^*x_2^*}}S_{M_{x_1^*x_2^*}}} - Y_{x_1^*x_2^*I_{M_{x_1^*x_2^*}}S_{M_{x_1^*x_2^*}}}\right]$ and expresses how much the intercept and slope of the outcome process would change on average if the intercept and slope functions of the treatment/exposure were changed from levels $x_1^* = x_2^* = 0$ to $x_1 = a_1$ and $x_2 = a_2$ but for each individual the intercept and slope of the mediator process is kept at the level it would have taken under the absence of the treatment/exposure. The natural indirect effect for two values of the intercept function of the treatment/exposure x_1 and x_1^* and for two values of the slope function of the treatment/exposure x_2 and x_2^*, is defined as $\mathrm{E}\left[Y_{x_1x_2I_{M_{x_1x_2}}S_{M_{x_1x_2}}} - Y_{x_1x_2I_{M_{x_1^*x_2^*}}S_{M_{x_1^*x_2^*}}}\right]$ and expresses how much the intercept and slope of the outcome process would change on average if the if the intercept and slope functions of the treatment/exposure were controlled at levels $x_1 = a_1$ and $x_2 = a_2$ but the intercept and slope of the mediator process were changed from the level they would take if the if the intercept and slope functions of the treatment/exposure functions were changed from $x_1^* = x_2^* = 0$ to $x_1 = a_1$ and $x_2 = a_2$.

We show below that the natural direct and indirect effects are identified if:

(C5) $I_{Y_{I_M S_M}}, S_{Y_{I_M S_M}} \amalg I_X, S_X | C$ (no unmeasured confounding for the exposure-outcome relationship)

(C6) $I_{Y_{I_M S_M}}, S_{Y_{I_M S_M}} \amalg I_M, S_M | I_X, S_X, C$ (no unmeasured confounding for the mediator-outcome relationship)

(C7) $I_{M_{x_1 x_2}}, S_{M_{x_1 x_2}} \amalg I_X, S_X | C$ (no unmeasured confounding for the exposure-mediator relationship)

(C8) $I_{Y_{m_1,m_2}}, S_{Y_{m_1,m_2}} \amalg I_{M_{x_1^* x_2^*}}, S_{M_{x_1^* x_2^*}} | C$ (no mediator-outcome confounders which are affected by the exposure)

Proposition: For any function u if (C5) - (C-8) hold then
$\mathrm{E}\left[u\left(I_{Y_{x_1 x_2} I_{M_{x_1^* x_2^*}} S_{M_{x_1^* x_2^*}}}, S_{Y_{x_1 x_2} I_{M_{x_1^* x_2^*}} S_{M_{x_1^* x_2^*}}}\right)\right] = \sum_{c,m_1,m_2} \mathrm{E}\left[u\left(I_Y, S_Y\right) | x_1, x_2, m_1, m_2, c\right] \Pr\left(m_1, m_2 | x_1^*, x_2^*, c\right) \Pr(c)$

Proof:

$$
\begin{aligned}
&\mathrm{E}\left[u\left(I_{Y_{x_1x_2I_{M_{x_1^*x_2^*}}S_{M_{x_1^*x_2^*}}}}, S_{Y_{x_1x_2I_{M_{x_1^*x_2^*}}S_{M_{x_1^*x_2^*}}}}\right)\right] \\
&\qquad = \sum_{c} \mathrm{E}\left[u\left(I_{Y_{x_1x_2I_{M_{x_1^*x_2^*}}S_{M_{x_1^*x_2^*}}}}, S_{Y_{x_1x_2I_{M_{x_1^*x_2^*}}S_{M_{x_1^*x_2^*}}}}\right)|C=c\right]\Pr(C=c) \\
&&\text{(Iterated Expectations)} \\
&= \sum_{c,m_1,m_2} \mathrm{E}\left[u\left(I_{Y_{x_1x_2m_1m_2}}, S_{Y_{x_1x_2m_1m_2}}\right)|C=c, I_{M_{x_1^*x_2^*}}=m_1, S_{M_{x_1^*x_2^*}}=m_2\right] \\
&\qquad\qquad \times \Pr\left(I_{M_{x_1^*x_2^*}}=m_1, S_{M_{x_1^*x_2^*}}=m_2|C=c\right)\Pr(C=c) \\
&&\text{(Iterated Expectations)} \\
&= \sum_{c,m_1,m_2} \mathrm{E}\left[u\left(I_{Y_{x_1x_2m_1m_2}}, S_{Y_{x_1x_2m_1m_2}}\right)|C=c\right]\Pr\left(I_{M_{x_1^*x_2^*}}=m_1, S_{M_{x_1^*x_2^*}}=m_2|I_X=x_1^*, S_X=x_2^*, C=c\right) \\
&\qquad \times \Pr(C=c) && \text{((C8) \& (C7))} \\
&= \sum_{c,m_1,m_2} \mathrm{E}\left[u\left(I_{Y_{x_1x_2m_1m_2}}, S_{Y_{x_1x_2m_1m_2}}\right)|I_X=x_1, S_X=x_2, C=c\right] \\
&\qquad\qquad \times \Pr\left(I_M=m_1, S_M=m_2|I_X=x_1^*, S_X=x_2^*, C=c\right)\Pr(C=c) \\
&&\text{((C5) \& consistency)} \\
&= \sum_{c,m_1,m_2} \mathrm{E}\left[u\left(I_{Y_{x_1x_2m_1m_2}}, S_{Y_{x_1x_2m_1m_2}}\right)|I_X=x_1, S_X=x_2, I_M=m_1, S_M=m_2, C=c\right] \\
&\qquad\qquad \times \Pr\left(I_M=m_1, S_M=m_2|I_X=x_1^*, S_X=x_2^*, C=c\right)\Pr(C=c) && \text{((C6))} \\
&= \sum_{c,m_1,m_2} \mathrm{E}\left[u\left(I_Y, S_Y\right)|I_X=x_1, S_X=x_2, I_M=m_1, S_M=m_2, C=c\right] \\
&\qquad\qquad \times \Pr\left(I_M=m_1, S_M=m_2|I_X=x_1^*, S_X=x_2^*, C=c\right)\Pr(C=c) \\
&&\text{(consistency)} \\
&= \sum_{c,m_1,m_2} \mathrm{E}\left[u\left(I_Y, S_Y\right)|x_1, x_2, m_1, m_2, c\right]\Pr\left(m_1, m_2|x_1^*, x_2^*, c\right)\Pr(c)
\end{aligned}
$$

This completes the proof.

Then if we replace x with x^* we get:

$$\mathrm{E}\left[u\left(I_{Yx_1^*x_2^*I_{Mx_1^*x_2^*}S_{Mx_1^*x_2^*}}, S_{Yx_1^*x_2^*I_{Mx_1^*x_2^*}S_{Mx_1^*x_2^*}}\right)\right] = \sum_{c,m_1,m_2} \mathrm{E}\left[u\left(I_Y, S_Y\right)|x_1^*, x_2^*, m_1, m_2, c\right] \Pr\left(m_1, m_2|x_1^*, x_2^*, c\right) \Pr(c)$$

from this it follows with $u(I_Y, S_Y) = I_Y + S_Y t + \varepsilon_Y$ that the average natural direct effect is given by:

$$\mathrm{E}\left[u\left(I_{Yx_1x_2I_{Mx_1^*x_2^*}S_{Mx_1^*x_2^*}}, S_{Yx_1x_2I_{Mx_1^*x_2^*}S_{Mx_1^*x_2^*}}\right) - u\left(I_{Yx_1^*x_2^*I_{Mx_1^*x_2^*}S_{Mx_1^*x_2^*}}, S_{Yx_1^*x_2^*I_{Mx_1^*x_2^*}S_{Mx_1^*x_2^*}}\right)\right] =$$

$$\sum_{c,m_1,m_2} \left\{\mathrm{E}\left[u\left(I_Y, S_Y\right)|x_1, x_2, m_1, m_2, c\right] - \mathrm{E}\left[u\left(I_Y, S_Y\right)|x_1^*, x_2^*, m_1, m_2, c\right]\right\} \Pr\left(m_1, m_2|x_1^*, x_2^*, c\right) \Pr(c)$$

If we replaced x^* with x we would get:

$$\mathrm{E}\left[u\left(I_{Yx_1x_2I_{Mx_1x_2}S_{Mx_1x_2}}, S_{Yx_1x_2I_{Mx_1x_2}S_{Mx_1x_2}}\right)\right] = \sum_{c,m_1,m_2} \mathrm{E}\left[u\left(I_Y, S_Y\right)|x_1, x_2, m_1, m_2, c\right] \Pr\left(m_1, m_2|x_1, x_2, c\right) \Pr(c)$$

from

this it follows with $u(I_Y, S_Y) = I_Y + S_Y t + \varepsilon_Y$ that the natural indirect effect is given by:

$$\mathrm{E}\left[u\left(I_{Yx_1x_2I_{Mx_1x_2}S_{Mx_1x_2}}, S_{Yx_1x_2I_{Mx_1x_2}S_{Mx_1x_2}}\right) - u\left(I_{Yx_1x_2I_{Mx_1^*x_2^*}S_{Mx_1^*x_2^*}}, S_{Yx_1x_2I_{Mx_1^*x_2^*}S_{Mx^*}}\right)\right] =$$

$$\sum_{c,m_1,m_2} \mathrm{E}\left[u\left(I_Y, S_Y\right)|x_1, x_2, m_1, m_2, c\right] \left\{\Pr\left(m_1, m_2|x_1, x_2, c\right) - \Pr\left(m_1, m_2|x_1^*, x_2^*, c\right)\right\} \Pr(c)$$

With the model shown in Figure 2.2 we have that $Y = u\left(I_Y, S_Y\right) = I_Y + S_Y t + \varepsilon_Y$. Thus given (2.16) and (2.17) we have

$$\mathrm{E}\left[u\left(I_Y, S_Y\right)|x_1, x_2, m_1, m_2, c\right] = \varphi_0 + \varphi_1 x_1 + \varphi_2 x_2 + \varphi_3 m_1 + \varphi_4 m_2 + \varphi_5 x_1 m_1 + \varphi_6 x_1 m_2 + \varphi_7 x_2 m_1 + \varphi_8 x_2 m_2$$
$$+ \varphi_9' c + (\gamma_0 + \gamma_1 x_1 + \gamma_2 x_2 + \gamma_3 m_1 + \gamma_4 m_2 + \gamma_5 x_1 m_1 + \gamma_6 x_1 m_2 + \gamma_7 + x_2 m_1 + \gamma_8 x_2 m_2 + \gamma_9' c) t$$

and

$$\mathrm{E}\left[u\left(I_Y, S_Y\right)|x_1^*, x_2^*, m_1, m_2, c\right] = \varphi_0 + \varphi_1 x_1^* + \varphi_2 x_2^* + \varphi_3 m_1 + \varphi_4 m_2 + \varphi_5 x_1^* m_1 + \varphi_6 x_1^* m_2 + \varphi_7 x_2^* m_1 + \varphi_8 x_2^* m_2$$
$$+ \varphi_9' c + (\gamma_0 + \gamma_1 x_1^* + \gamma_2 x_2^* + \gamma_3 m_1 + \gamma_4 m_2 + \gamma_5 x_1^* m_1 + \gamma_6 x_1^* m_2 + \gamma_7 + x_2^* m_1 + \gamma_8 x_2^* m_2 + \gamma_9' c)t$$

Therefore the average natural direct effect is

$$\sum_{c, m_1, m_2} \Big\{(\varphi_1 + \gamma_1 t + (\varphi_5 + \gamma_5 t)m_1 + (\varphi_6 + \gamma_6 t)m_2)(x_1 - x_1^*)$$
$$+(\varphi_2 + \gamma_2 t + (\varphi_7 + \gamma_7 t)m_1 + (\varphi_8 + \gamma_8 t)m_2)(x_2 - x_2^*)\Big\}$$
$$= (\varphi_1 + \gamma_1 t + (\varphi_5 + \gamma_5 t)\mathrm{E}\left[M_1|x_1^*, x_2^*, c\right] + (\varphi_6 + \gamma_6 t)\mathrm{E}\left[M_2|x_1^*, x_2^*, c\right])(x_1 - x_1^*) +$$
$$(\varphi_2 + \gamma_2 t + (\varphi_7 + \gamma_7 t)\mathrm{E}\left[M_1|x_1^*, x_2^*, c\right] + (\varphi_8 + \gamma_8 t)\mathrm{E}\left[M_2|x_1^*, x_2^*, c\right])(x_2 - x_2^*)$$
$$= (\varphi_1 + \gamma_1 t + (\varphi_5 + \gamma_5 t)(\delta_0 + \delta_1 x_1^* + \delta_2 x_2^* + \delta_3' c) + (\varphi_6 + \gamma_6 t)(\beta_0 + \beta_1 x_1^* + \beta_2 x_2^* + \beta_3' c))(x_1 - x_1^*)$$
$$+ (\varphi_2 + \gamma_2 t + (\varphi_7 + \gamma_7 t)(\delta_0 + \delta_1 x_1^* + \delta_2 x_2^* + \delta_3' c) + (\varphi_8 + \gamma_8 t)(\beta_0 + \beta_1 x_1^* + \beta_2 x_2^* + \beta_3' c))(x_2 - x_2^*)$$

Given (2.13), (2.14), (2.15) and (2.16) we have

$$\mathrm{E}\left[u\left(I_{Y_{x_1x_2}I_{M_{x_1x_2}}S_{M_{x_1x_2}}}, S_{Y_{x_1x_2}I_{M_{x_1x_2}}S_{M_{x_1x_2}}}\right)\right] = \varphi_0 + \gamma_0 t + (\varphi_1 + \gamma_1 t)x_1 + (\varphi_2 + \gamma_2 t)x_2$$
$$+ (\varphi_3 + \gamma_3 t)(\delta_0 + \delta_3' c) + (\varphi_4 + \gamma_4 t)(\beta_0 + \beta_3' c) + ((\varphi_5 + \gamma_5 t)(\delta_0 + \delta_3' c) + (\varphi_6 + \gamma_6 t)(\beta_0 + \beta_3' c))x_1$$
$$+ ((\varphi_7 + \gamma_7 t)(\delta_0 + \delta_3' c) + (\varphi_8 + \gamma_8 t)(\beta_0 + \beta_3' c))x_2 + ((\varphi_3 + \gamma_3 t)\delta_1 + (\varphi_4 + \gamma_4 t)\beta_1)x_1$$
$$+ ((\varphi_3 + \gamma_3 t)\delta_2 + (\varphi_4 + \gamma_4 t)\beta_2)x_2 + ((\varphi_5 + \gamma_5 t)\delta_1 + (\varphi_6 + \gamma_6 t)\beta_1)x_1x_1$$
$$+ ((\varphi_5 + \gamma_5 t)\delta_2 + (\varphi_6 + \gamma_6 t)\beta_2)x_1x_2 + ((\varphi_7 + \gamma_7 t)\delta_1 + (\varphi_8 + \gamma_8 t)\beta_1)x_2x_1$$
$$+ ((\varphi_7 + \gamma_7 t)\delta_2 + (\varphi_8 + \gamma_8 t)\beta_2)x_2x_2$$

$$
\begin{aligned}
\mathrm{E}\left[u\left(I_{Y x_1 x_2 I_{M x_1 x_2}} S_{M x_1 x_2}, S_{Y x_1 x_2 I_{M x_1 x_2}} S_{M x_1 x_2}\right)\right] &= \varphi_0 + \gamma_0 t + (\varphi_1 + \gamma_1 t) x_1 + (\varphi_2 + \gamma_2 t) x_2 \\
&+ (\varphi_3 + \gamma_3 t)(\delta_0 + \delta_3' c) + (\varphi_4 + \gamma_4 t)(\beta_0 + \beta_3' c) + ((\varphi_5 + \gamma_5 t)(\delta_0 + \delta_3' c) + (\varphi_6 + \gamma_6 t)(\beta_0 + \beta_3' c)) x_1 \\
&+ ((\varphi_7 + \gamma_7 t)(\delta_0 + \delta_3' c) + (\varphi_8 + \gamma_8 t)(\beta_0 + \beta_3' c)) x_2 + ((\varphi_3 + \gamma_3 t)\delta_1 + (\varphi_4 + \gamma_4 t)\beta_1) x_1^* \\
&+ ((\varphi_3 + \gamma_3 t)\delta_2 + (\varphi_4 + \gamma_4 t)\beta_2) x_2^* + ((\varphi_5 + \gamma_5 t)\delta_1 + (\varphi_6 + \gamma_6 t)\beta_1) x_1 x_1^* \\
&+ ((\varphi_5 + \gamma_5 t)\delta_2 + (\varphi_6 + \gamma_6 t)\beta_2) x_1 x_2^* + ((\varphi_7 + \gamma_7 t)\delta_1 + (\varphi_8 + \gamma_8 t)\beta_1) x_2 x_1^* \\
&+ ((\varphi_7 + \gamma_7 t)\delta_2 + (\varphi_8 + \gamma_8 t)\beta_2) x_2 x_2^*
\end{aligned}
$$

Therefore the average natural indirect effect is:

$$
\begin{aligned}
&\sum_{c, m_1, m_2} \mathrm{E}\left[u\left(I_Y, S_Y\right) | x_1, x_2, m_1, m_2, c\right] \Pr\left(m_1, m_2 | x_1, x_2, c\right) \Pr(c) - \mathrm{E}\left[u\left(I_Y, S_Y\right) | x_1, x_2, m_1, m_2, c\right] \Pr\left(m_1, m_2 | x_1^*, x_2^*, c\right) \Pr(c) \\
&= ((\varphi_3 + \gamma_3 t)\delta_1 + (\varphi_4 + \gamma_4 t)\beta_1) x_1 + ((\varphi_3 + \gamma_3 t)\delta_2 + (\varphi_4 + \gamma_4 t)\beta_2) x_2 + ((\varphi_5 + \gamma_5 t)\delta_1 + (\varphi_6 + \gamma_6 t)\beta_1) x_1 x_1 \\
&\quad + ((\varphi_5 + \gamma_5 t)\delta_2 + (\varphi_6 + \gamma_6 t)\beta_2) x_1 x_2 + ((\varphi_7 + \gamma_7 t)\delta_1 + (\varphi_8 + \gamma_8 t)\beta_1) x_2 x_1 + ((\varphi_7 + \gamma_7 t)\delta_2 + (\varphi_8 + \gamma_8 t)\beta_2) x_2 x_2 \\
&\quad - ((\varphi_3 + \gamma_3 t)\delta_1 + (\varphi_4 + \gamma_4 t)\beta_1) x_1^* - ((\varphi_3 + \gamma_3 t)\delta_2 + (\varphi_4 + \gamma_4 t)\beta_2) x_2^* - ((\varphi_5 + \gamma_5 t)\delta_1 + (\varphi_6 + \gamma_6 t)\beta_1) x_1 x_1^* \\
&\quad - ((\varphi_5 + \gamma_5 t)\delta_2 + (\varphi_6 + \gamma_6 t)\beta_2) x_1 x_2^* - ((\varphi_7 + \gamma_7 t)\delta_1 + (\varphi_8 + \gamma_8 t)\beta_1) x_2 x_1^* - ((\varphi_7 + \gamma_7 t)\delta_2 + (\varphi_8 + \gamma_8 t)\beta_2) x_2 x_2^* \\
&= ((\varphi_3 + \gamma_3 t)\delta_1 + (\varphi_4 + \gamma_4 t)\beta_1)(x_1 - x_1^*) + ((\varphi_3 + \gamma_3 t)\delta_2 + (\varphi_4 + \gamma_4 t)\beta_2)(x_2 - x_2^*) \\
&\quad + ((\varphi_5 + \gamma_5 t)\delta_1 + (\varphi_6 + \gamma_6 t)\beta_1)(x_1 x_1 - x_1 x_1^*) + ((\varphi_5 + \gamma_5 t)\delta_2 + (\varphi_6 + \gamma_6 t)\beta_2)(x_1 x_2 - x_1 x_2^*) \\
&\quad + ((\varphi_7 + \gamma_7 t)\delta_1 + (\varphi_8 + \gamma_8 t)\beta_1)(x_2 x_1 - x_2 x_1^*) + ((\varphi_7 + \gamma_7 t)\delta_2 + (\varphi_8 + \gamma_8 t)\beta_2)(x_2 x_2 - x_2 x_2^*)
\end{aligned}
$$

As discussed previously in the absence of interaction we specify

$\varphi_5 = \varphi_6 = \varphi_7 = \varphi_8 = \gamma_5 = \gamma_6 = \gamma_7 = \gamma_8 = 0$. This leads to the following direct effect:

$$\sum_{c,m_1,m_2} \left\{(\varphi_1 + \gamma_1 t)(x_1 - x_1^*) + (\varphi_2 + \gamma_2 t)(x_2 - x_2^*)\right\} \Pr\left(m_1, m_2 | x^*, c\right) \Pr(c)$$

$$= \left\{(\varphi_1 + \gamma_1 t)(x_1 - x_1^*) + (\varphi_2 + \gamma_2 t)(x_2 - x_2^*)\right\} \sum_{c,m_1,m_2} \Pr\left(m_1, m_2 | x^*, c\right) \Pr(c)$$

$$= (\varphi_1 + \gamma_1 t)(x_1 - x_1^*) + (\varphi_2 + \gamma_2 t)(x_2 - x_2^*)$$

and the following indirect effect:

$$\sum_{c,m_1,m_2} \mathrm{E}\left[u\left(I_Y, S_Y\right) | x_1, x_2, m_1, m_2, c\right] \Pr\left(m_1, m_2 | x_1, x_2, c\right) \Pr(c) - \mathrm{E}\left[u\left(I_Y, S_Y\right) | x_1, x_2, m_1, m_2, c\right] \Pr\left(m_1, m_2 | x_1^*, x_2^*, c\right) \Pr(c)$$

$$= ((\varphi_3 + \gamma_3 t)\delta_1 + (\varphi_4 + \gamma_4 t)\beta_1)(x_1 - x_1^*) + ((\varphi_3 + \gamma_3 t)\delta_2 + (\varphi_4 + \gamma_4 t)\beta_2)(x_2 - x_2^*)$$

2.4 Standard Errors of Direct and Indirect Effects

When considering the direct and indirect effects it is important to be able to test the statistical significance of these effects. Folmer[24], Sobel[61, 62], Bollen[5] and Bollen & Stine[7] suggest applying the delta method to estimate the asymptotic variances of the indirect and total effect. We suggest that the delta method is used in this case as well. In the latent growth mediation context both the direct and indirect effects are nonlinear functions of several model coefficient estimators. We then use the first order multivariate delta method in order to approximate the standard errors:

$$g\left(\hat{\theta}\right) \approx g(\theta) + \frac{\partial g(\theta)}{\partial \theta} \tag{2.18}$$

Considering equation 2.18 we see that $g(\hat{\theta})$ is approximately equal to a linear function of θ. We have from large sample theory that $g\left(\hat{\theta}\right)$ is approximately normal. Given that $g(\theta)$ is a constant we have a constant plus a multiple of a normally distributed variable so in large samples $g\left(\hat{\theta}\right)$ is approximately normal[7].

$$g\left(\theta'\right) \sim N\left(g(\theta), \nabla g(\theta)' Var(\theta) \nabla g(\theta)\right)$$

This means that we can use the normal distribution to create confidence intervals as well as perform hypothesis tests on the direct and indirect effects of models 1 and 2.

2.4.1 Standard Errors for Model 1

Using standard SEM software to fit model 1 results in estimates $\hat{\delta}$ of $\delta \equiv (\delta_0, \delta_1, \delta_2')'$, $\hat{\beta}$ of $\beta \equiv (\beta_0, \beta_1, \beta_2')'$, $\hat{\varphi}$ of $\varphi \equiv (\varphi_0, \varphi_1, \varphi_2, \varphi_3, \varphi_4, \varphi_5, \varphi_6')'$ and $\hat{\gamma}$ of $\gamma \equiv (\gamma_0, \gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5, \gamma_6')'$. Using these we take

$$\theta \equiv (\delta, \beta, \varphi, \gamma) \equiv (\delta_0, \delta_1, \delta_2', \beta_0, \beta_1, \beta_2', \varphi_0, \varphi_1, \varphi_2, \varphi_3, \varphi_4, \varphi_5, \varphi_6', \gamma_0, \gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5, \gamma_6').$$

Given the direct effect for model 1 in Section 2.2 we have

$$g(\theta) = (\varphi_1 + \varphi_4(\delta_0 + \delta_1 x^* + \delta_2' c) + \varphi_5(\beta_0 + \beta_1 x^* + \beta_2' c) + \gamma_1 t + \gamma_4(\delta_0 + \delta_1 x^* + \delta_2' c)t + \gamma_5(\beta_0 + \beta_1 x^* + \beta_2' c)t)$$

Thus we have

$$\nabla g(\theta) = \Big(\varphi_4 + \gamma_4 t, (\varphi_4 + \gamma_4 t)x^*, (\varphi_4 + \gamma_4 t)c', \varphi_5 + \gamma_5 t, (\varphi_5 + \gamma_5 t)x^*, (\varphi_5 + \gamma_5 t)c', 0, 1$$
$$, 0, 0, \delta_0 + \delta_1 x^* + \delta_2' c, \beta_0 + \beta_1 x^* + \beta_2' c, 0', 0, t, 0, 0, (\delta_0 + \delta_1 x^* + \delta_2' c)t, (\beta_0 + \beta_1 x^* + \beta_2' c)t, 0'\Big)'$$

Thus $SE\,(g(\theta)) = \sqrt{\nabla g(\theta)' Var(\theta) \nabla g(\theta)}$. This leads to the standard error of the direct effect in model 1:

$$\sqrt{\nabla g(\theta)' Var(\theta) \nabla g(\theta)} |x - x^*|$$

Given the indirect effect for model 1 in Section 2.2 we have

$$g(\theta) = \Big[(\varphi_2 + \gamma_2 t)\delta_1 + (\varphi_3 + \gamma_3 t)\beta_1\Big](x - x^*) + \Big[(\varphi_4 + \gamma_4 t)\delta_1 + (\varphi_5 + \gamma_5 t)\beta_1 (x^2 - xx^*)\Big]$$

Thus we have

$$\nabla g(\theta) = \Big(0, (\varphi_2 + \gamma_2 t)(x - x^*) + (\varphi_4 + \gamma_4 t)(x^2 - xx^*), 0', 0, (\varphi_3 + \gamma_3 t)(x - x^*) + (\varphi_3 + \gamma_5 t)(x^2 - xx^*)$$
$$, 0', 0, 0, \delta_1 (x - x^*), \beta_1 (x - x^*), \delta_1 (x^2 - xx^*), \beta_1 (x^2 - xx^*), 0', 0, \delta_1 t(x - x^*), \beta_1 t(x - x^*)$$
$$, \delta_1 t(x^2 - xx^*), \beta_1 t(x^2 - xx^*), 0'\Big)$$

Thus the standard error of the indirect effect in model 1:

$$\sqrt{\nabla g(\theta)' Var(\theta) \nabla g(\theta)}$$

2.4.2 Standard Errors for Model 2

Using standard SEM software to fit model 2 results in estimates $\hat{\rho_0}$ of ρ_0, $\hat{\lambda}_0$ of λ_0, $\hat{\delta}$ of $\delta \equiv (\delta_0, \delta_1, \delta_2, \delta_3')'$, $\hat{\beta}$ of $\beta \equiv (\beta_0, \beta_1, \beta_2, \beta_3')'$, $\hat{\varphi}$ of $\varphi \equiv (\varphi_0, \varphi_1, \varphi_2, \varphi_3, \varphi_4, \varphi_5, \varphi_6, \varphi_7, \varphi_8, \varphi_9')'$ and $\hat{\gamma}$ of $\gamma \equiv (\gamma_0, \gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5, \gamma_6, \gamma_7, \gamma_8, \gamma_9')'$.

Using these we take

$$\theta \equiv (\rho_0, \lambda_0, \delta, \beta, \varphi, \gamma) \equiv (\rho_0, \lambda_0, \delta_0, \delta_1, \delta_2, \delta_3', \beta_0, \beta_1, \beta_2, \beta_3', \varphi_0, \varphi_1, \varphi_2, \varphi_3, \varphi_4, \varphi_5, \varphi_6, \varphi_7$$
$$, \varphi_8, \varphi_9', \gamma_0, \gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5, \gamma_6, \gamma_7, \gamma_8, \gamma_9').$$

Given the direct effect for model 2 in Section 2.3 we have

$$g(\theta) = (\varphi_1 + \gamma_1 t + (\varphi_5 + \gamma_5 t)(\delta_0 + \delta_1 x_1^* + \delta_2 x_2^* + \delta_3' c) + (\varphi_6 + \gamma_6 t)(\beta_0 + \beta_1 x_1^* + \beta_2 x_2^* + \beta_3' c))(x_1 - x_1^*)$$
$$+ (\varphi_2 + \gamma_2 t + (\varphi_7 + \gamma_7 t)(\delta_0 + \delta_1 x_1^* + \delta_2 x_2^* + \delta_3' c) + (\varphi_8 + \gamma_8 t)(\beta_0 + \beta_1 x_1^* + \beta_2 x_2^* + \beta_3' c)(x_2 - x_2^*)$$

Thus we have

$$\nabla g(\theta) = \Big(0, 0, (\varphi_5 + \gamma_5 t)(x_1 - x_1^*) + (\varphi_7 + \gamma_7 t)(x_2 - x_2^*), (\varphi_5 + \gamma_5 t)x_1^*(x_1 - x_1^*) + (\varphi_7 + \gamma_7 t)x_1^*(x_2 - x_2^*)$$
$$, (\varphi_5 + \gamma_5 t)x_2^*(x_1 - x_1^*) + (\varphi_7 + \gamma_7 t)x_2^*(x_2 - x_2^*), (\varphi_5 + \gamma_5 t)c'(x_1 - x_1^*) + (\varphi_7 + \gamma_7 t)c'(x_2 - x_2^*)$$
$$, (\varphi_6 + \gamma_6 t)(x_1 - x_1^*) + (\varphi_8 + \gamma_8 t)(x_2 - x_2^*), (\varphi_6 + \gamma_6 t)x_1^*(x_1 - x_1^*) + (\varphi_8 + \gamma_8 t)x_1^*(x_2 - x_2^*)$$
$$, (\varphi_6 + \gamma_6 t)x_2^*(x_1 - x_1^*) + (\varphi_8 + \gamma_8 t)x_2^*(x_2 - x_2^*), (\varphi_6 + \gamma_6 t)c'(x_1 - x_1^*) + (\varphi_8 + \gamma_8 t)c'(x_2 - x_2^*)$$
$$, 0, (x_1 - x_1^*), (x_2 - x_2^*), 0, 0, (\delta_0 + \delta_1 x_1^* + \delta_2 x_2^* + \delta_3' c)(x_1 - x_1^*), (\beta_0 + \beta_1 x_1^* + \beta_2 x_2^* + \beta_3' c)(x_1 - x_1^*)$$
$$, (\delta_0 + \delta_1 x_1^* + \delta_2 x_2^* + \delta_3' c)(x_2 - x_2^*), (\beta_0 + \beta_1 x_1^* + \beta_2 x_2^* + \beta_3' c)(x_2 - x_2^*), 0', 0, (x_1 - x_1^*)t, (x_2 - x_2^*)t$$
$$, 0, 0, (\delta_0 + \delta_1 x_1^* + \delta_2 x_2^* + \delta_3' c)(x_1 - x_1^*), (\beta_0 + \beta_1 x_1^* + \beta_2 x_2^* + \beta_3' c)(x_1 - x_1^*)$$
$$, (\delta_0 + \delta_1 x_1^* + \delta_2 x_2^* + \delta_3' c)(x_2 - x_2^*), (\beta_0 + \beta_1 x_1^* + \beta_2 x_2^* + \beta_3' c)(x_2 - x_2^*), 0' \Big)$$

Thus the standard error of the direct effect in model 2:

$$\sqrt{\nabla g(\theta)' Var(\theta) \nabla g(\theta)}$$

Given the indirect effect for model 2 in Section 2.3 we have

$$g(\theta) = ((\varphi_3 + \gamma_3 t)\delta_1 + (\varphi_4 + \gamma_4 t)\beta_1)(x_1 - x_1^*) + ((\varphi_3 + \gamma_3 t)\delta_2 + (\varphi_4 + \gamma_4 t)\beta_2)(x_2 - x_2^*)$$
$$+ ((\varphi_5 + \gamma_5 t)\delta_1 + (\varphi_6 + \gamma_6 t)\beta_1)(x_1 x_1 - x_1 x_1^*) + ((\varphi_5 + \gamma_5 t)\delta_2 + (\varphi_6 + \gamma_6 t)\beta_2)(x_1 x_2 - x_1 x_2^*)$$
$$+ ((\varphi_7 + \gamma_7 t)\delta_1 + (\varphi_8 + \gamma_8 t)\beta_1)(x_2 x_1 - x_2 x_1^*) + ((\varphi_7 + \gamma_7 t)\delta_2 + (\varphi_8 + \gamma_8 t)\beta_2)(x_2 x_2 - x_2 x_2^*)$$

Thus we have

$$
\begin{aligned}
\nabla g(\theta) = (\ & 0, 0, 0, (\varphi_3 + \gamma_3 t)(x_1 - x_1^*) + (\varphi_5 + \gamma_5 t)(x_1 x_1 - x_1 x_1^*) + (\varphi_7 + \gamma_7 t)(x_2 x_1 - x_2 x_1^*) \\
& , (\varphi_3 + \gamma_3 t)(x_2 - x_2^*) + (\varphi_5 + \gamma_5 t)(x_1 x_2 - x_1 x_2^*) + (\varphi_7 + \gamma_7 t)(x_2 x_2 - x_2 x_2^*), 0', 0 \\
& , (\varphi_4 + \gamma_4 t)(x_1 - x_1^*) + (\varphi_6 + \gamma_6 t)(x_1 x_1 - x_1 x_1^*) + (\varphi_8 + \gamma_8 t)(x_2 x_1 - x_2 x_1^*) \\
& , (\varphi_4 + \gamma_4 t)(x_2 - x_2^*) + (\varphi_6 + \gamma_6 t)(x_1 x_2 - x_1 x_2^*) + (\varphi_8 + \gamma_8 t)(x_2 x_2 - x_2 x_2^*) \\
& , 0', 0, 0, 0, \delta_1(x_1 - x_1^*) + \delta_2(x_2 - x_2^*), \beta_1(x_1 - x_1^*) + \beta_2(x_2 - x_2^*) \\
& , \delta_1(x_1 x_1 - x_1 x_1^*) + \delta_2(x_1 x_2 - x_1 x_2^*), \beta_1(x_1 x_1 - x_1 x_1^*) + \beta_2(x_1 x_2 - x_1 x_2^*) \\
& , \delta_1(x_2 x_1 - x_2 x_1^*) + \delta_2(x_2 x_2 - x_2 x_2^*), \beta_1(x_2 x_1 - x_2 x_1^*) + \beta_2(x_2 x_2 - x_2 x_2^*) \\
& , 0', 0, 0, 0, \delta_1 t(x_1 - x_1^*) + \delta_2 t(x_2 - x_2^*), \beta_1 t(x_1 - x_1^*) + \beta_2 t(x_2 - x_2^*) \\
& , \delta_1 t(x_1 x_1 - x_1 x_1^*) + \delta_2 t(x_1 x_2 - x_1 x_2^*), \beta_1 t(x_1 x_1 - x_1 x_1^*) + \beta_2 t(x_1 x_2 - x_1 x_2^*) \\
& , \delta_1 t(x_2 x_1 - x_2 x_1^*) + \delta_2 t(x_2 x_2 - x_2 x_2^*), \beta_1 t(x_2 x_1 - x_2 x_1^*) + \beta_2 t(x_2 x_2 - x_2 x_2^*), 0'
\end{aligned}
$$

Thus the standard error of the indirect effect in model 2:

$$\sqrt{\nabla g(\theta)' Var(\theta) \nabla g(\theta)}$$

2.5 An Example

In this section we give an example of a longitudinal mediation analysis using latent growth curve models and the definition of the direct and indirect effects shown in section 2.2. The data and motivation of this example comes from Gunzler et

al.[30]. Their goal was to develop an adjusted screening tool to better assess depressive symptoms in Multiple Sclerosis (MS) patients. Screening for depression in this population can be challenging due to the overlap of MS symptoms with symptoms of depression. One mechanism by which MS may affect depression is through physical limitations as measured, for example, by a timed 25-foot walk. Disentangling these relationships can be key for treatment as depression is the most frequent psychiatric diagnosis in MS patients [26].

Consider the latent growth curve model as shown in figure 2.3. We are interested in how MS type ($0 \rightarrow$ relapsing, $-1 \rightarrow$ progessive) affects self-reported depression screening (PHQ-9) directly and indirectly through a timed 25-foot walk. As noted in Figure 2.3, log timed walk (ltw) and PHQ-9 (PHQ) are measured at 6 different time points. These time points vary between subjects.

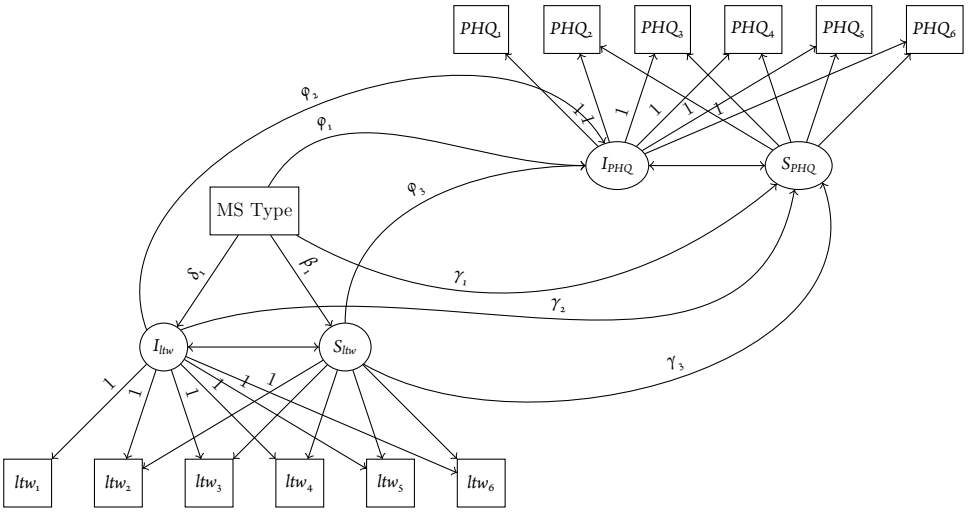


Figure 2.3: MS and Depression Example

PHQ-9 is used both in screening and monitoring of depression in patients.

Patients respond to a likert scale from 0 (not at all) to 3 (every day) about 9 different symptoms over the prior 2 weeks before their appointment[30]. This leads to a total score with a range from 0 to 27. The 25-foot timed walk is a quantitative test of mobility and leg function. Gunzler et al used an additional peg test to quantitatively assess arm and hand function as well as using each symptom of PHQ-9 as an outcome. Here we focus on the total PHQ-9 score as the outcome and only the log timed walk for the mediator.

Gunzler et al.[30] used data from the Knowledge Program developed at Cleveland Clinic's Neurological Institute [2] which links PHQ-9 data to its EPIC electronic health record. The Mellen Center[3] for MS manages more than 20,000 visits and 1,000 new patients every year for MS treatment. The Knowledge Program tracks illness severity and treatment efficacy over time across the Mellen Center. This data comes from a retrospective cohort containing patients with measurements of PHQ-9 and a 25-foot timed walk data available. Table 2.1 displays the demographic information of the 3,507 patients in the sample from 2008 - 2011. In the table the patients are split by a PHQ-9 score of < 10 and ≥ 10, where 10 is a validated threshold for moderate depression[35].

For this example we fit the following mediator process

$$ltw_{it} = I_{ltw_i} + S_{ltw_i} + \varepsilon_{ltw_{it}}$$

$$I_{ltw_i} = \delta_0 + \delta_1 CCLB1_i + \nu_{I_{ltw_i}}$$

$$S_{ltw_i} = \beta_0 + \beta_1 CCLB1_i + \nu_{S_{ltw_i}}$$

	$PHQ-9<10$ $n=2,502$	$PHQ-9\geq 10$ $n=1,005$	P-value
PHQ-9	3.64 ± 2.75	15.26 ± 4.40	< 0.001
MSPS fatigue	1.62 ± 1.25	3.35 ± 1.12	< 0.001
MSPS cognitive	0.86 ± 0.96	2.23 ± 1.30	< 0.001
MSPS mobility	1.37 ± 1.58	2.39 ± 1.48	< 0.001
MSPS hand function	0.77 ± 0.94	1.79 ± 1.27	< 0.001
25-Foot time walk	7.85 ± 10.56	8.83 ± 7.61	0.002
9-hole peg test	23.68 ± 10.66	26.82 ± 12.48	< 0.001
Age	46.12 ± 11.88	44.47 ± 11.20	< 0.001
Baseline time since diagnosis	11.80 ± 10.00	10.89 ± 9.37	0.016
Female $n(\%)$	$1,836(74)$	$740(74)$	0.879
Race, $n(\%)$			0.07
Caucasian	2,112 (85)	821 (82)	
African-American	225 (9)	114 (11)	
Other	144 (6)	65 (7)	
MS type, $n(\%)$			0.067
Relapsing	2,045 (84)	787 (82)	
Progressive	383 (16)	177 (18)	

Table 2.1: Demographics of the 3,507 Patients in Sample

and the following outcome process

$$PHQ_{it} = I_{PHQ_i} + S_{PHQ_i} + \varepsilon_{PHQ_{it}}$$

$$I_{PHQ_i} = \varphi_0 + \varphi_1 CCLB1_i + \varphi_2 I_{ltw_i} + \varphi_3 I_{Stw_i} + \nu_{I_{PHQ_i}}$$

$$S_{PHQ_i} = \gamma_0 + \gamma_1 CCLB1_i + \varphi_2 I_{ltw_i} + \varphi_3 I_{Stw_i} + \nu_{S_{ltw_i}}.$$

Where $\mathrm{E}\left[\varepsilon_{ltw_{it}}\right] = \mathrm{E}\left[\varepsilon_{PHQ_{it}}\right] = \mathrm{E}\left[\nu_{I_{ltw_i}}\right] = \mathrm{E}\left[\nu_{S_{ltw_i}}\right] = \mathrm{E}\left[\nu_{I_{PHQ_i}}\right] = \mathrm{E}\left[\nu_{S_{PHQ_i}}\right] = 0$ and where $\varepsilon_{ltw_{it}}, \varepsilon_{PHQ_{it}}, (\nu_{I_{ltw_i}}, \nu_{S_{ltw_i}}) and (\nu_{I_{PHQ_i}}, \nu_{S_{PHQ_i}})$ are mutually independent. The example here is included for the purposes of illustration only and no covariates are

adjusted for in the analysis; future work will carefully assess what covariates need to be adjusted for to make the confounding assumptions plausible.

Variable	Estimate	Std. Err.	t	$\Pr > \lvert t \rvert$	95% CI	
δ_0	2.361	0.029	81.106	< 0.001	2.30416	2.41784
δ_1	0.59	0.031	19.057	< 0.001	0.52924	0.65076
β_0	0.066	0.012	5.695	< 0.001	0.04248	0.08952
β_1	0.049	0.012	4.018	< 0.001	0.02548	0.07252
φ_0	-0.825	0.996	-0.828	0.408	-2.77716	1.12716
φ_1	-1.238	0.377	-3.288	0.001	-1.97692	-0.49908
φ_2	3.677	0.499	7.365	< 0.001	2.69896	4.65504
φ_3	-8.897	4.343	-2.049	0.04	-17.4093	-0.38472
γ_0	1.643	0.463	3.552	< 0.001	0.73552	2.55048
γ_1	0.396	0.18	2.205	0.027	0.0432	0.7488
γ_2	-0.904	0.233	-3.875	< 0.001	-1.36068	-0.44732
γ_3	8.185	2.096	3.905	< 0.001	4.07684	12.29316

Table 2.2: Estimates from Model shown in Figure 2.3. Obtained using Mplus version 7.2[49]

Table 2.2 displays the results estimated by fitting the above model in Mplus. Recall from Section 2.2 that for this model without interaction the direct effect is $(\varphi_1 + \gamma_1 t)(x - x^*)$ and the indirect effect is $((\varphi_2 + \gamma_2 t)\delta_1 + (\varphi_3 + \gamma_3 t)\beta_1)(x - x^*)$. We let $x = 0$ and $x^* = -1$ to reflect a change in MS status from relapsing to progressive such that the direct effect is

$$(\varphi_1 + \gamma_1 t)(x - x^*) = -1.238 + 0.396t$$

and the indirect effect is

$$((\varphi_2+\gamma_2 t)\delta_1+(\varphi_3+\gamma_3 t)\beta_1)(x-x^*) = ((3.677-0.904t)0.59+(-8.897+8.185t)0.049) = 1.733-0.132t$$

Recall from Section 2.4.1 that the standard error of the direct effect is $\sqrt{\nabla g(\theta_d)' Var(\theta) \nabla g(\theta_d)} |x - x^*|$ where $\nabla g(\theta_d)' = (0, 0, 0, 0, 0, 1, 0, 0, 0, t, 0, 0)$ and the standard error for the indirect effect is $\sqrt{\nabla g(\theta_{id})' Var(\theta) \nabla g(\theta_{id})}$ where $\nabla g(\theta_{id})' = ((0, (\varphi_2 + \gamma_2 * t)(x - x^*), 0, (\varphi_3 + \gamma_3 * t)(x - x^*), 0, 0, (\delta_1, \beta_1)(x - x^*), 0, 0, \delta_1 * t(x - x^*), \beta_1 * t(x - x^*))$, where $Var(\theta)$ is estimated by Mplus.

Table 2.3 displays the direct and indirect effects at various time points as well as a 95% confidence interval at each time point and p-value of the effect at that particular time point. We see that initially direct effect is negative however it becomes statistically insignificant sometime between 1 and 2 years. However the indirect effect is positive with decreasing effect size yet remains statistically significant throughout the duration of this study.

Time (years)	Direct Effect	95%CI		p-value	Indirect Effect	95%CI		
		Lower	Upper			Lower	Upper	
0	-1.238	-1.97659	-0.49942	0.001	1.733477	1.270811	2.196144	< 0.001
1	-0.842	-1.50667	-0.17733	0.013	1.601182	1.245434	1.95693	< 0.001
2	-0.446	-1.21115	0.319154	0.253	1.468887	1.065199	1.872575	< 0.001
3	-0.05	-1.0382	0.938198	0.921	1.336592	0.768012	1.905172	< 0.001

Table 2.3: Direct and Indirect Effects of Model in Figure 2.3

MacKinnon[41] estimates as the direct effect $\gamma_1 = 0.396$ 95% CI (0.043, 0.749) and an indirect effect of $\beta_1 \gamma_3 = (0.049)(8.185) = 0.401$ 95% CI (0.122, 0.680), where the standard error of the indirect effect is $\sqrt{\beta_1^2 \sigma_{\gamma_3}^2 + \gamma_3^2 \sigma_{\beta_1}^2} = 0.142$. This makes the further assumption that the direct and indirect effect remain constant throughout time as opposed to the methods in this paper which allow for the direct and indirect effect to change with time.

2.6 Discussion

This paper mathematically defines the direct and indirect effects of longitudinal mediation with latent growth curve models using counterfactuals. We build upon the models considered by MacKinnon[14, 41] but allowed for the presence of treatment/exposure-mediator interaction. We then considered the assumptions needed for identifiability of these direct and indirect effects. Those assumptions are:

1. No unmeasured confounding of the exposure-outcome relationship
2. No unmeasured confounding of the mediator-outcome relationship
3. No unmeasured confounding of the exposure-mediator relationship
4. No mediator-outcome confounders which are affected by the exposure

We consider these effects using the above assumptions first with a model in which the treatment/exposure is binary, followed by a model in which the treatment/exposure itself changes with time. We find that latent growth mediation models in current literature allow for the intercept of the outcome to be a cause for the slope of the mediator. This violates assumption 4 since the intercept of the outcome would become a mediator-outcome confounder which itself was affected by exposure. We also find that with many models currently used in the literature it is assumed that the direct and indirect effects remain constant while these methods allow them to vary with time. With the direct and indirect effects defined we consider the delta method for estimating the standard error of those effects.

The methodology developed in this paper thus formalizes mediation analysis with latent growth curve models using counterfactuals, makes clear the assumptions and extends these methods to allow for exposure mediator interactions.

3

The Results of Blended Instruction in Quantitative Methods in Public Health: A Pilot Study

Adam J Sullivan, Jenny Bergeron, & Marcello Pagano

3.1 Introduction

Technology advances continuously. It is our responsibility as educators to carefully evaluate when these advances have the potential to impact our teaching, and consider how best to adopt or possibly ignore them.

As with most change, it is advisable to introduce slowly, drawing on parallels with established methods of teaching. This will typically facilitate the comparison of the new with the old, in order to evaluate whether the new methods improve

the learning experience. For example, McGready and Brookmeyer [47], analyze data from a study comparing online biostatistics courses with the traditional course. In their study they found no statistical evidence of a difference in student achievement when comparing the two styles of teaching. There is a wide and growing body of literature on evaluating blended learning at the university level [8, 13, 15, 19, 20, 38, 42, 65, 79]. Zhao and Breslow [78] compiled an extensive literature review on blended learning and found that there were various studies which showed statistically significant differences between blended and traditional learning styles. However many of these studies adequately control for confounders and they report that there is still no strong evidence that one blended method is better than another at this time.

We are reporting on a pilot study we have carried out over the last two years at the Harvard School of Public Health that was guided by the above considerations. The focus of this report is on evaluating the impact these innovations have had on the learning of biostatistics. We began designing 2 courses that allowed us to make the comparisons/contrasts. These courses are:

1. Bio 200: Principles of Biostatistics (5 credits), an introductory biostatistics course that has existed for almost 100 years at the Harvard School of Public Health (although the name of the course has changed) and newly blended for the first time in fall of 2013;

2. ID 200: Principles of Biostatistics and Epidemiology (7.5 credits), a new course pioneered in the fall of 2013 and offered again in summer of 2014. ID 200 combines the materials of BIO 200 and EPI 500: Fundamentals of Epidemiology (2.5 credits).

Both of these courses were taught in a blended format, which consisted of online materials as well as traditional classroom meetings. The online materials were produced for the edX course PH207x: Health in Numbers: Quantitative Methods in Clinical & Public Health Research. PH207x's online material consisted of videos covering the material listed in section 3.2. In between videos were a series of multiple choice questions, discussion questions and applets to help further student understanding. The traditional classroom meetings consisted of weekly discussion and labs.

With this study we compare BIO 200 from fall of 2013 and ID 200 from fall of 2013 and summer of 2014 to BIO 200 classes from 1992 and 1993, as well as BIO 201: Introduction to Statistical Methods from the fall of 2013. BIO 201 is similar in content to BIO 200 but the primary audience is for students with higher mathematical and statistical backgrounds than BIO 200.

We first discuss the way in which our blended courses were designed. We then analyze the delivery of this course from the perspectives of the students and an informal comparison to both current and past courses. We then suggest changes to consider when offering a course of this format. Finally we consider how implementing these changes effected the assessment of ID 200 in the summer 2014 semester.

3.2 Design and Implementation of Courses

3.2.1 BIO 200

The blended format of BIO 200 was offered for the first time during the 2013 fall semester. We used the edX platform with the biostatistics material loaded into

weekly modules. These modules covered the material which is classically required for accreditation in the MPH degree [1] and are covered in the book, Principle of Biostatistics [51]:

Data Presentation	Numerical Summary Measures
Rates and Standardization	Life Tables[1]
Probability	Theoretical Probability Distributions
Sampling Distribution of the means	Confidence Intervals
Hypothesis Testing	Comparison of Two Means
Analysis of Variance	Non-parametric Methods
Inference on Proportions	Contingency Tables
Correlation[2]	Linear Regression [2]
Logistic Regression[1]	Sampling Theory[1]

These topics were covered over 15 weeks. Each week consisted of at least one lesson module on the topics above. Modules were constructed with a series of short video lessons, quick check questions, discussions, applets and online problem sets. Many of the videos were 10 minutes or less and were utilized as a video "textbook". There were also instructional videos on how to utilize the associated statistical computer program package, Stata [64], in order to evaluate the statistical methods taught in each module. Many of the videos including the Stata instructional videos as well as problem sets utilized a sample of data from the Framingham Heart Study including 4,434 observations and 74 variables. This sample allowed the students to become familiar with a real data set as they worked throughout the course. The quick check questions are multiple choice questions that followed a

[1]This topic was not covered in the previous brick and mortar offering of BIO 200

[2]This topic was only briefly covered in the previous brick and mortar offering of BIO 200

video. They were ungraded and could be answered as many times as the student desired. The students were encouraged to watch all or part of a video over again if they could not answer them correctly. A number of applets were used to help students understand concepts (such as the quincunx applet [55] to help students visualize the binomial distribution). For students who did wish to view the videos or had difficulty in learning from videos it was suggested that they read from the suggested course textbook [51] or the "jotter notes." The jotter notes are edited scripts of the video lessons with slides from the videos filled in where they would have appeared in the video. Additional online material consisted of discussion boards and problem sets. The discussion boards contained guided discussion topics, module material discussion and discussion on the problem set material. These boards were moderated by several teaching assistants to maintain accuracy of information given. All students who emailed questions were instructed to place their questions on the discussion board where both the question and responses would be available to the entire class. The online problem sets consisted of multiple choice and fill in answers based on conceptual as well as data analysis questions.

Prior to a weekly class, students were expected to complete the assigned modules (Either by watching all of the videos or reading the textbook or "jotter notes"). After completing this material they were asked to respond to an online survey, called "Muddy Points Surveys", on how comfortable they were with the major concepts in the videos and practice problems. Concepts were rated on a four point Likert scale ranging from "Very Unclear" to "Very Clear." These surveys helped to form the basis of the 80 minute class discussion with the professor each week. These discussions contained review of that week's particular modules as well as more examples from new data sets of the Framingham Heart Study sample. In

order to encourage more discussion the students were split into two groups of 65 students. In addition to the in class session students were assigned to a 2-hour lab run by a teaching assistant. The lab sessions consisted of problem solving and data analysis using Stata. Labs alternated between the teaching assistant teaching how to solve a particular problem and individual or group work to solve the problems. Problems were designed to walk the students through how a statistician would approach and solve them.

3.2.2 ID 200

The blended format of ID 200 was first offered during the fall 2013 semester then again in the summer 2014 semester. The fall semester spanned 15 weeks and the summer semester spanned 6 weeks. With the fall semester the biostatistics material covered was the same as in the fall 2013 BIO 200. For the summer Life Tables and Sampling theory were removed due to the time constraints and course load of the students (17.5 credits for their summer semester). Along with the biostatistics component there was the addition of epidemiology component. For the epidemiology component of ID 200 the topics covered were the same as PH207X:

Prevalence	Incidence
Measures of Association	Case Reports
Experimental Studies	Causal Inference
Cohort Study Design	Confounding
Case Control Studies	Stratification
Mantel-Haenszel Estimation	Matching
Regression Coefficients	Propensity Scores
Screening	

In ID 200 the procedure was similar to BIO 200 with the exception that they met once a week for 4 hours at a time. During this block of time the students had about 80 minutes for biostatistics discussion, 25 minutes for epidemiology discussion, 60 minutes for lab time and the rest of the time was devoted to a term long project. The project for the fall semester was completed in groups where the groups each determined a topic of interest and posited a series of questions about the chosen topic. In some cases IRB approval was sought in order to survey students at HSPH. Other groups utilized publicly available data for analysis. Once the data was collected students analyzed their data and used various biostatistical and epidemiological tests to best assess their posited questions. The project culminated in a presentation given to the class, teaching assistants and faculty. For the summer semester the project was again completed in groups but all groups were required to design a survey, obtain IRB approval and then collect the data from the summer students at the school. Once they analyzed this data they were required to write a paper and give a presentation to the class and instructors.

3.2.3 GAISE Guidelines

The Guidelines for Assessment and Instruction in Statistics Education (GAISE) College Report gives six recommendations for an introductory course in statistics

1. Emphasize statistical literacy and develop statistical thinking.
2. Use real data.
3. Stress conceptual understanding, rather than mere knowledge of procedures.
4. Foster active learning in the classroom.
5. Use technology for developing conceptual understanding and analyzing data.

Both Bio 200 and ID 200 meet all of the proposed guidelines. Throughout the course topics were tied together in order to boost statistical literacy and thinking. The Framingham Heart Study data was utilized in videos, online discussion questions, in-class discussions as well as problem sets gave students the opportunity to work with real data which is not perfect (i.e. it contains missing data as well non-normal data). The use of Stata throughout the course allowed for actual data analysis to be performed. Labs and in-class discussions were designed to get students engaged and to be active in their learning. Students performed data analysis as a class, in small groups or individually. Finally the exam questions were not written to test students ability to reproduce calculations or multiple formulas, rather there were about the overarching concepts and core knowledge of statistics. We consider it more valuable that a student would understand what a concept was and why a particular test was used rather than testing to see if they can compute a test statistic by hand. These guidelines allow us to utilize blended learning and technology while maintaining the best current standards of teaching in statistics education.

3.3 Evaluation of 2013 Fall Semester Courses

3.3.1 Student's Evaluation of Courses

We employed 3 different surveys to allow students to evaluate the course. The first survey was a midterm survey given to the students using Qualtrics (Qualtrics, Provo, UT). We chose to leave the results of this survey out of our analysis due to complications with the survey. In order to maintain anonymity of the survey no record of students names were kept. Initial review of the mid-term survey provided sufficient evidence that some students answered multiple times. In one simple example a student wrote that they had filled it out earlier but forgot one comment. The second survey employed was the standard end of course evaluation which is given by the Office of Education at Harvard T.H. Chan School of Public Health. Given the complications that arose with the mid-term survey we had a person experienced in evaluation from the Bok Center at Harvard carry out the survey for us. Before we received the results of the survey the responses were anonymized. Students were informed about the design of the survey and had the option of taking this survey. Results for selected questions of both end of year surveys can be seen in Table 3.1 and Tables 3.3, 3.4 and 3.5.

3.3.1.1 Survey 2 - End of Course survey Harvard T.H. Chan School of Public Health

In the Harvard Chan School survey for ID 200, 29 out of 30 students (97%) responded and for BIO 200, 114 out of 130 (88%) responded. In the Bok Center survey for ID 200, 22 out of 30 students (73%) responded and for BIO 200, 68 out

of 130 (52%) responded. Students were asked if they would recommend this course to another student. For ID 200 58.6% of the respondents answered that they probably or definitely would whereas 46.8% of students in BIO 200 answered the same. This gives a combined 47.6% of the respondents in the two courses would probably or definitely recommend this course to another. In Fall 2012 semester, with the traditional teaching only, 95.5% of the respondents answered that they would probably or definitely recommend this course to another student. There was a statistically significant difference between the responses of the three courses. When combining the data for the fall 2013 blended courses and comparing it to 2012 BIO 200 the p-value is still highly significant. This survey did not allow the students to explain why they chose their responses.

When asked about the amount of time spent outside of class for the courses, 65.5% of ID 200 respondents answered that they spend less than or equal to 6 hours a week, 81.6% of BIO 200 respondents and 84.3% of 2012 BIO 200 respondents answered the same. There is a difference between the courses however this is somewhat concerning as BIO 200 is a 5 credit course and ID 200 is a 7.5 credit course. Students should be spending at least 50% more time on ID 200. They not only had the same biostatistics videos but had additional epidemiology videos. ID 200 students also had additional problem set questions and a semester project to work on.

Finally students were asked about formal training in the content area prior to taking this course. Table 3.1 displays the results and we find that there are significant differences. Students in Fall 2013 ID 200 had more experience than that of BIO 200. Students who self selected to take ID 200 in the Fall 2013 semester may have chosen to do so due to having more exposure to statistics than BIO 200 students.

Would you recommend this course to another? [n(%)]					
	Fall 2013		Fall 2012	Summer 2014	
	ID 200	BIO 200	BIO 200	ID 200	p-value
Definitely Would	6 (20.7)	17 (14.9)	91 (84.3)	11 (73.3)	< 0.001
Probably Would	11 (37.9)	34 (29.8)	12 (11.1)	4 (26.7)	
Probably Would Not	10 (34.5)	62 (54.4)	2 (1.9)	0	
No Response	2 (6.9)	1 (0.9)	3 (2.8)	0	
On average, how many hours per week outside of class did you dedicate to this course? [n(					
	Fall 2013		Fall 2012	Summer 2014	
	ID 200	BIO 200	BIO 200	ID 200	p-value
<2 hours	0	7 (6.1)	2 (1.9)	0	0.092
2-3 hours	2 (6.9)	34 (29.8)	38 (35.2)	2 (13.3)	
4-6 hours	17 (58.6)	52 (45.7)	51 (47.2)	8 (15.3)	
7-12 hours	8 (27.6)	17 (14.9)	14 (13)	4 (26.7)	
>12 hours	1 (3.4)	4 (3.5)	2 (1.9)	1 (6.7)	
No Response	1 (3.4)	0	1 (0.9)	0	
Did you have formal training in the content area prior to taking this course? [n(%)]					
	Fall 2013		Fall 2012	Summer 2014	
	ID 200	BIO 200	BIO 200	ID 200	p-value
None	3 (10.3)	41 (36)	50 (37)	0	< 0.001
Some	22 (75.9)	66 (57.9)	61 (56.5)	10 (66.7)	
Considerable	2 (6.9)	6 (5.3)	6 (5.6)	5 (33.3)	
No Response	2 (6.9)	1 (0.9)	1 (0.9)	0	

Table 3.1: Student responses to Harvard T.H. Chan School of Public Health end of course survey

The responses from this survey suggest that students are not as comfortable with this course as with the prior brick and mortar format. Some blended research shows that students' dislike of blended learning is from the extra workload [79], given the student responses shown here it is not evident to us that students are working more outside the classroom. They may believe that they are working more and longer than they would have in a traditional course however this is not upheld by our survey findings.

3.3.1.2 Survey 3 - Bok Center Survey

In the Bok Center survey 22 out of 30 (73%) ID 200 students responded and 68 out of 130 (52%) BIO 200 students responded. Students were asked questions about in class meetings, problem sets as well as online instruction and watching behavior. They were also allowed to give extensive comments and written critiques that we discuss as well. Table 3.2 shows some of the demographic information which we use to compare the survey sample to the overall class. We note that in the overall class data there is some missing data. One student selected FERPA so their information is not shared. Other students with missing data are from other Harvard schools outside of Harvard T.H. Chan School of Public Health and we were unable to obtain access to their demographic data. Based on table 3.2 the available demographic data suggests that there is no difference between the sample for the survey and the overall class.

	ID 200			BIO 200		
	Sample	Overall	p-value	Sample	Overall	p-value
Gender, count (%)			0.573			0.592
Female	13 (59.1)	14 (46.7)		47 (69.1)	83 (63.8)	
Male	9 (40.9)	15 (50)		18 (26.5)	26 (20)	
Missing	0	1 (3.3)		3 (4.4)	21 (16.2)	
Concentration, Count (%)			1			0.393
GHP	2 (9)	2 (6.7)		10 (14.7)	15 (11.5)	
HPM	1 (5)	2 (6.7)		2 (2.9)	12 (9.2)	
MPH	18 (81)	23 (76.7)		38 (55.9)	61 (46.9)	
SBS	1 (5)	2 (6.7)		13 (19.1)	18 (13.8)	
SHH	0	0		1 (1.5)	1 (0.8)	
Missing	0	1 (3.3)		3 (4.4)	23 (17.7)	

Table 3.2: Demographic Information for Survey Sample

Table 3.3 displays the results of the Bok Center survey pertaining to statements regarding in class meetings. The first statement that students responded to was that the in class meetings were well organized. In ID 200 48% of the respondents

agree or strongly agree versus 34% who disagree or strongly disagree. In BIO 200 29% of the respondents agree or strongly agree versus 44% who disagree or strongly disagree. In some comments students felt that the in class meetings did not clarify the confusing portion of the videos or highlight the main points. The second statement responded to was that the in class meetings expanded upon what was learned online. In ID 200 29% of respondents agree or strongly agree versus 28% who disagree or strongly disagree. In BIO 200 31% of the respondents agree or strongly disagree versus 42% who disagree or strongly disagree. The last statement was that students were actively engaged in the in-class exercises and discussion. In ID 200 64% of the respondents agreed or strongly agreed versus BIO 200 where only 35% of the respondents agreed or strongly agreed. This suggests that the class size of 30 in ID 200 allowed for for engagement than the class size of greater than 60 for BIO 200.

Table 3.4 displays the results of the Bok Center survey pertaining to statements regarding problem sets and assignments.The first statement was that the weekly problem sets helped the students think critically about the course material. In ID 200 81% of the respondents agreed or strongly agreed versus BIO 200 where 77% agreed or strongly agreed. These high responses were as we had hoped when designing the problem sets for the course. The second response was the muddiest points survey helped students think critically about the material. In ID 200 10% of the respondents agreed versus in BIO 200 where 9% agreed or strongly agreed. We hoped the muddiest points surveys would have students think about all of the concepts that the videos covered and pick which ones they most wanted to learn more about in class. In order to ensure this we used these surveys to help with

	In Class Meetings [%]					
	Response 1: Were well-organized			Response 2: Expanded upon what I learned online		
	Fall 2013		Summer 2014	Fall 2013		Summer 2014
	ID 200	BIO 200	ID 200	ID 200	BIO 200	ID 200
Strongly	10	8	0	14	13	0
Disagree	24	36	0	14	29	0
Neutral	19	27	11	43	27	22
Agree	38	24	33	24	28	67
Strongly	10	5	56	5	3	11
	Response 3: I was actively engaged in in-class exercises and discussion					
	Fall 2013		Summer 2014			
	ID 200	BIO 200	ID 200			
Strongly	5	5	0			
Disagree	5	26	0			
Neutral	27	35	11			
Agree	59	33	67			
Strongly	5	2	22			

Table 3.3: Bok Center Survey Results: In Class Meetings

participation grade. Students mentioned that many times they blindly filled the survey out just for participation points but did not put much into it. The last statement on projects was only answered by ID 200 students, 52% of the respondents agreed or strongly agreed that semester long project helped them to think critically about the material. Students stated that they felt smaller groups would have been better to assure all students worked on the project. They also stated that a more strict time line would have kept them more focused on it during the semester rather than rushed at the end.

Table 3.5 displays the results of the Bok Center survey pertaining to statements regarding online instruction and watching behavior. The first statement was that the online lessons were interesting and engaging. In ID 200 70% of the respondents agree with this versus in BI0 200 where only 34% of the students agree or strongly

	Homework and Assignments [%]					
	Response 4: The weekly homework assignments helped me think critically about the course material			Response 5: The muddiest points surveys helped me think critically about the material		
	Fall 2013		Summer 2014	Fall 2013		Summer 2014
	ID 200	BIO 200	ID 200	ID 200	BIO 200	
Strongly	5	3	0	19	39	
Disagree	10	9	0	22	34	
Neutral	5	11	0	28	17	
Agree	52	50	11	10	6	
Strongly	29	27	89	0	3	

	Response 6: The semester long project helped me think critically about the material	
	Fall 2013	Summer 2014
	ID 200	ID 200
Strongly	14	0
Disagree	10	0
Neutral	14	0
Agree	33	44
Strongly	29	56

Table 3.4: Bok Center Survey Results: Problem Sets and Assignments

agree with this. We find this intriguing as the biostatistics video material was exactly the same for both courses. These responses may be in part due to the students in ID 200 being self selected into this blended format. The second statement was that online lessons were divided into manageable segments, 45% in ID 200 versus 47% in BIO 200 agree or strongly agree. This left 35% in ID 200 versus 28% in BIO 200 who disagree or strongly disagree. Some students commented that some videos were too long and they would lose attention on them. This could simply be remedied by breaking longer videos up into shorter segments.

The last set of questions were about the watching behavior of the students. We first asked them how often they multi-tasked while watching the online lessons. In ID 200, 50% of the respondents said they did fairly often to very often. In Bio 200, only 28% of the respondents did fairly often to very often. 10% of respondents in ID 200 versus 26% of respondents in Bio 200 said they never multi-tasked while

	Online Instruction and Watching Behavior [%]					
	Response 7: Online lessons were interesting and engaging			Response 8: Online lessons were divided into manageable segments		
	Fall 2013		Summer 2014	Fall 2013		Summer 2014
	ID 200	BIO 200	ID 200	ID 200	BIO 200	ID 200
Strongly	10	14	0	15	8	0
Disagree	15	36	14	20	20	0
Neutral	5	16	43	20	25	14
Agree	70	31	29	40	44	57
Strongly	0	3	14	5	3	29
	Response 9: How often did you multitask while watching online lessons?			Response 10: How often did you skip part of the video?		
	Fall 2013		Summer 2014	Fall 2013		Summer 2014
	ID 200	BIO 200	ID 200	ID 200	BIO 200	ID 200
Never	10	26	14	10	13	0
Almost Never	15	17	29	15	36	43
Sometimes	25	28	29	25	22	0
Fairly Often	35	17	29	25	16	43
Very Often	15	11	0	25	14	14
	Response 11: How often did you rewind/repeat parts of the video					
	Fall 2013		Summer 2014			
	ID 200	BIO 200	ID 200			
Never	5	3	0			
Almost Never	5	17	14			
Sometimes	45	41	57			
Fairly Often	40	27	14			
Very Often	5	13	14			

Table 3.5: Bok Center Survey Results: Online Instruction and Watching Behavior

watching lessons. This seems to contradict the above statement about online lessons being interesting and engaging. BIO 200 respondents self reported being less engaged or interested but were more likely to focus on the online lessons alone. When asked how often they skipped part of a video 50% of ID 200 respondents versus 30% of BIO 200 respondents answered fairly often to very often. Finally when asked about how often they rewound or repeated parts of the video only 10% of ID 200 respondents said that they almost never or never did in contrast to 20% of BIO 200 respondents that answered the same.

3.3.2 Evaluation of the Students

In this section we compare students in the Fall 2013 ID 200 and BIO 200 offerings to each other, to students in the Fall of 2013 offering of BIO 201 andto students from the Fall of 1992 and 1993 semesters. We use the course exams to compare the 2 courses and to test for differences in exam scores. We then compare the students in ID 200 and BIO 200 to the traditional class, BIO 201. In order to achieve this comparison we utilized the same exact exam question on all three of the final exams. Finally we compare the exam scores of the current blended courses to that of the traditional courses of the past.

3.3.2.1 Students in ID 200 versus students in BIO 200

The students self-selected to take ID 200 (30 students) versus BIO 200 (130 students) plus some epidemiology course to replace the epidemiology component of ID 200, such as EPI500. We first consider whether we can detect any difference between how these two groups of students handled the material.

The overlapping material was the biostatistics covered in both courses which was identical. The online problem sets were identical with additional epidemiology problems added to the problem sets of ID 200. Problem sets were assigned to be completed individually. We did a problem by problem comparison on the home works (roughly 10 problems a week, comparisons not reported here) and could not detect a difference between the two groups of students. Overall, the ID 200 students had a mean grade of 95% (median 96%) and the BIO 200 students the same mean (median 97%). Thus there was no discernible difference in how well the two groups did in the problem sets they submitted.

Secondly, we gave three (equi-spaced in time) exams to the students. Unbeknownst to the students, we gave the two groups the same biostatistical component of the exam for the first and third exam. Table 3.6 shows the percentages achieved by the students for all 3 exams (Note: Exam 2 Historical differs in score as BIO 200 was matched with students in BIO 200 Fall 1993 and ID 200 was matched with students in BIO 200 Fall of 1992).

	BIO 200		ID 200	
Exam	Mean	Median	Mean	Median
1	96	98	94	95
Historical	84	86	84	86
2	89	91	88	89
Historical	84	87	96	98
3	92	94	91	93
Historical	78	79	78	79

Table 3.6: Comparison of Exam Scores Between Blended Courses and Historical Courses

We were unable to discern between the exam scores between these two groups. Together with the problem set results, it seems, to the accuracy of our measuring method, that there was no difference between the two groups' performance.

3.3.2.2 Blended students versus current other students

One other introductory biostatistics class is also taught at Harvard T.H. Chan School of Public Health, BIO 201 Introduction to Statistical Methods, was taught in the traditional lecture/lab/problem set format. Incoming students typically choose between BIO 200 and BIO 201 depending on their areas of concentration and with those that are more mathematically inclined taking BIO201. The

coverage is broader in BIO 200.

In order to try and compare students in the class we took a question from the BIO 201 final and placed it on the 3rd exam for both BIO 200 and ID 200. This question covered about one quarter of the points and was a topic that in previous years has not been covered in the traditional BIO 200. Table 3.7 displays the results of the problem scores.

Semester	Course	Students	Mean	Median
Fall 2013	BIO 201	99	21.7	23
Fall 2013	ID 200	30	22.2	23.5
Fall 2013	BIO 200	130	22.1	23
Summer 2014	ID 200	15	19.1	21

Table 3.7: Comparison of scores for Blended courses vs BIO 201 (out of 24 points)

We see that all of the grades were too close for statistical significance but the trend would suggest that students in the blended courses did better than the students in BIO 201. This is surprising since it is generally believed that the BIO 201 students are more quantitatively adept, the coverage in BIO 201 is much less than in the blended course, and on a minor note, the exam question was written by the BIO 201 instructor and one would expect the BIO201 students to be more attuned to his wording of questions.

3.3.2.3 Historical Comparisons

To enhance the evaluations we also incorporated old exams. The current Exam 1 was also administered in Fall 1992 offering of BIO 200. The intent was to provide a comparison between students taking the "same" course from the same instructor

but 20 years later. We did something similar with Exam 3 by using the final exam from Fall 1992 offering of BIO 200. For Exam 2 we gave different exams to the two groups; for the ID 200 students we used the Fall 1992 BIO 200 exam 2, and for the BIO 200 students we used the Fall 1993 BIO 200 exam. The results for all the exams taken are displayed in Table 3.6.

Students in BIO 200 did considerably better on all the exams compared to their historical counterparts. Except for the second exam for the ID 200 group, the current group of students did considerably better than their counterparts from 20 years ago. It should also be noted that at the time of exam entry this type of analysis was not planned and the scores for exam 2 report the overall grade including biostatistics and epidemiology questions. We recognize that education has changed greatly over the course of 20 years. Students now are more likely to have seen statistics in high school and undergraduate programs than 20 years ago. This is possibly an explanation for some of the grade improvement but this shows promise for the blended teaching method as well.

The mean and median of course are summary statistics that by their very summary definition may hide complexities that make the comparisons more nuanced. Figure 3.1 displays the histograms of the different exam scores for BIO 200, ID 200 and the historical courses. For exams 1 and we can see that the histograms for Fall 2013 BIO 200 and ID 200 are much more negatively skewed than the historical comparison of Fall 1992. The histograms for these exams also show the difficulty of comparing the exam scores. With the upper limit of 100% on the exam we can see the distributions of the Fall 2013 classes becoming more weighted on grades closer to that upper limit. For exam 2 we find that for Fall 2013 BIO 200 the histogram again shows a distribution more weighted towards

100% however exam 2 for Fall 1992 BIO 200's distribution of scores appears more weight towards 100% than that of Fall 2013 ID 200.

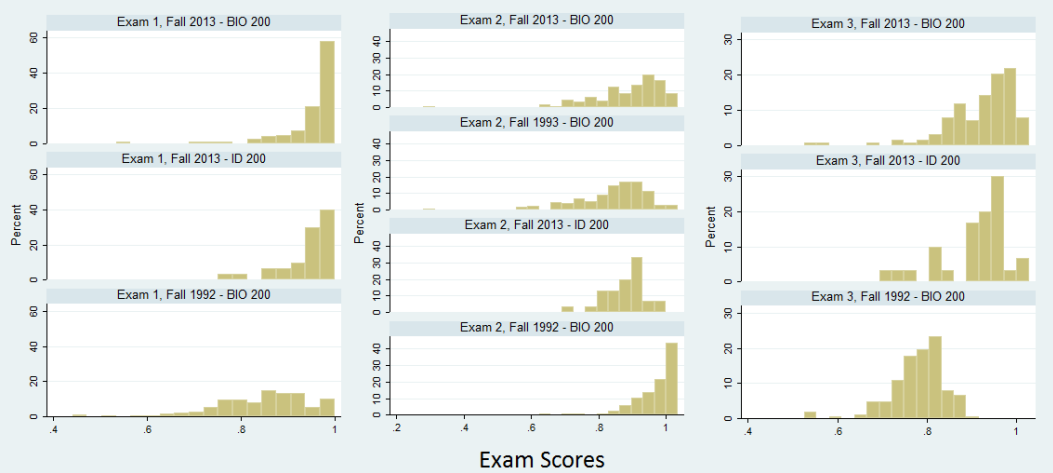


Figure 3.1: Histograms of Exam Scores.

3.3.3 Discussion of Results and Suggested Changes

From the students' perspective of the course there are four major areas of comments: online videos, classroom, problem sets and labs. These are suggestions for improvements to our course and future courses.

For the online component:

- Divide videos into manageable segments.
- Publish the video times so that students are better able to plan their week.
- Establish the purpose of each lesson and preview the organization.
- Improve the discussion format as it currently was not very helpful. Students would prefer more real time interaction and collaboration then the discussion board allowed for.

For the in class component:

- Students would prefer meeting more than one session a week.
- Students felt the class was not integrated with videos, problem sets and labs.
- Students would like class to consist of:
  - Question and Answer time to resolve misunderstandings.
  - Review of main points of the videos and concepts.
  - Time to practice problems and material in collaboration with their classmates.

For the problem sets component:

- Align problem sets with videos and exams.
- Utilize more problems that address conceptual understanding to allow students to check understanding of the course material.

For the lab component, students only suggestion was to move lab to earlier in week to allow them more time to complete assignments after having lab.

Students in ID 200 found the project to be very useful. Some suggested smaller group sizes and more structure in order to make sure they are working on it throughout the semester.

Students mentioned that many videos were too long and a struggle to adequately plan their week. This seems contradictory to the amount of time spent outside of class on the two courses though. we do suggest keeping the videos to under 10 minutes each and to post video times in order to give students more information. We suggest that rather than meeting with the students once a week for 80 minutes that meeting two times a week for 50 minutes a time may be better

suited to their needs. Smaller class sizes would allow for adjustment of the in class meetings. Students in ID 200 (class size 30) felt more engaged in the class than students in BIO 200 (class size 65). Instead of lecturing or reviewing for the in class time we suggest running the in class session similar to the labs. This would be real life biostatistics problems that allow students to work together or individually using Stata to aid in finding solutions. Within these problems reviewing important concepts and answering questions.

For the problem set component we suggest making sure if the problem set is in an online format to make sure the grading has a working error variance as well as making sure there are as few bugs as possible.

Most importantly we suggest instructing the students on how to take a course such as this. In table 3.5 many students reported multitasking while watching videos. We suggest that they reserve breaks to in between videos and we believe that having videos under 10 minutes would minimize the amount of multitasking that was being done. Instructing the students on how to spread the video materials over the course of a week would also be helpful to them.

Overall the majority of the students preferred having more interaction with their classmates and professor than this format allowed for. They desired real time feedback that the course did not offer. This could be changed with video office hours or other times during the week where video conferencing could be utilized to help students engage more.

Things we have learned from our research and informal comparison of the courses:

- Inform the students that the videos are not replacements for the lectures.

Rather, they are replacement for the book, if anything. The class discussion is a more apt replacement.

- Allow students to annotate the videos to better find points difficult to understand.
- Cold call in order to help push students to watch the videos and/or read the "jotter note" or textbook and come to class prepared.

We were not able to show that the ID 200 students learned biostatistics better than the BIO 200 students. Part of this was that the means and medians were too close to 100% so there was not much space to see a difference. This was because they did so much better than in the past that the exams which had some discriminatory power in the past, no longer do! It is also possible that there was no difference. We had hoped that learning introductory epidemiology and biostatistics simultaneously would have a positive synergistic effect on the learning of both. We hope that this will allow for longevity of understanding the concepts taught in the course as they mover further from the course.

In summary, this is not by any means a stringent study to evaluate the students. It is important to consider how this blended format will compare over time. It would be beneficial to follow-up with these students in the future to test whether blended learning leads to increased retention over that of a traditional class. The existing evidence is that the students performed better, or at least as well, past students. Presumably, as we become more adept at teaching a blended course, we should improve. After all, this was the first time we taught in this manner.

3.4 Changes Made to Summer 2014 ID 200 Course

With the summer course being over 7 weeks we took great care in utilizing the suggested changes that students in the Fall 2013 courses suggested. We addressed all concerns in the online, classroom, problem set and lab components of the course. For the online component we divided many of the longer videos so that all videos were at most 10 minutes or less. With the course management system each module was listed with objectives, video times and respective jotter notes for the module. All of these things were utilized to better explain what we expected the students to learn as well as how much time they needed to plan for with the videos. The new management system for the course allowed easier use of the discussion board as well as instant alerts to the instructors so that we were aware of questions immediately. This provided a simple and real time interaction for the students. For the in class component we met twice a week for 4 hours each time. Each class was started off with time for questions on any of the materials. Then within the lab setting there was review of formulas and concepts before working together on problems or paper reviews. The problem set component was revamped and instead of being graded automatically the problem set was designed with data analysis problems similar to the lab material. This provided a seamless interaction between in-class work and problem sets. With each class being taught in a lab or seminar format they essentially had problem solving and collaboration with classmates each and every class. Most classes have 45 minutes to an hour left over for working on the group projects.

In order to address other concerns of navigating this new style of the course we created a couple new videos for them to watch before starting class material. The

first video was designed to show them the course management system and to take them through how to utilize it and access all the material. The next video discussed strategy for how to take a blended course. It was suggested that they watched a video with no distractions while taking notes. It was also suggested that they could take breaks between the short videos in order to help them focus better during the videos.

The last major changes for the summer course was the assignment of pre-course work. The following topics were covered in the month of June before the course started:

- Data Presentation
- Numerical Summary Measures
- Rates and Standardization
- Probability
- Prevalence
- Incidence

A comprehensive problem set was given to cover all of these concepts in order to make sure the students understood before the first day of class.

3.5 Evaluation of 2104 Summer Semester Course

We wished to discover whether the changes noted in section 3.4 made an impact on both the students knowledge gained as well as how students perceived the course. In order to gauge student perception we use the end of the year survey from Harvard T.H. Chan School of Public Health (the same as the one used in

section 3.3.1.1 and a survey similar to the one conducted by the Bok center. In order to gauge the knowledge of students we were unable to compare the 3 exams as in fall 2013 course due to the time constraint of the summer semester. We did however use the same exam question that was used in the fall 2013 final exam of BIO 200, ID 200 and BIO 201.

3.5.1 Student's Evaluation of Course

In the Harvard T.H. Chan School of Public Health survey, 15 out of the 15 students responded. In the other survey 9 out of 15 students responded. This is similar to the fall 2013 survey from the Bok Center where 58.6% of ID 200 and 46.8% of the BIO 200 students responded. However in this instance both surveys allowed for students to give further details regarding their perception of the course.

When asked about whether they would recommend this course to another student 100% of the students responded that they probably or definitely would. This is much different outcome than in the fall courses where combined only 47.6% said the probably or definitely would. The students commented that this course allowed them to work at their own pace. A number of students said that the labs run in class provided an opportunity to expand on what they learned from the online material. When asked about the amount of time spent outside of class for the course 66.6% of the students responded that they spent less than or equal to 6 hours a week outside of class. This is very similar to the responses in the fall courses but this time is concerning. The average time per week of the videos is 4 hours 43 minutes and 36 seconds. If each student watched the videos the 66.6% would have had less than an hour and a half to do their problem set as well as group project. It should be noted here that during these 6 weeks in the summer

these particular students were taking 17.5 credits. This is likely to cause their time available to spend on ID 200 to drop significantly. Finally when students were asked about formal training in the content area prior to taking this course 33.3% of the students reported that they had considerable training. This is different from the fall courses but is expected given that the 15 students were enrolled in the DrPH program as opposed to the fall where most students were enrolled in a an MPH program.

For the in class meetings the first statement that students responded to was that the in class meetings were well organized. For the summer 88.9% of the students agreed or strongly agreed versus 0% disagreement. This is much different from the fall evaluations where in ID 200 34% disagreed or strongly disagreed with this statement. Students felt that having most of the time spent in class working on problems was very valuable. The change to spending ore class time in a lab setting seems to have helped students more. The second statement about in class meetings was that they expanded on what the students learned online. In the summer 2014 semester 77.8% of the students agreed or strongly disagreed versus 0% in disagreement. This is once again different from the fall courses. This perhaps explains part of the positive response for promoting this class and how the students felt about the organization. The last statement was that students were actively engaged in the in class exercises and discussion 88.9% of the students agreed or strongly agreed with this versus 0% in disagreement. This is consistent with our conclusion on the fall courses that having smaller groups tends to help students feel more engaged.

The first statement about problem sets was that the weekly problem sets helped students to think critically about the course material. 100% of the students agreed

or strongly disagreed with this. This is higher agreement than the semester and may be attributed to the problem set being submitted in written format rather than through an online grading system. The muddiest points surveys were removed for the summer and the students were not asked about them. When asked if the semester long project helped the students think critically about the material, 77.8% agreed or strongly agreed. This is a larger percent of the students than in the fall course.

The first statement about the online instruction was that the online lessons were interesting and engaging. In the summer 42.9% of the students agreed or strongly agreed versus 14.3% which disagreed. This again is different from the fall semester and we do not understand the differences. The online materials were the same with the exception that some longer videos were split into multiple shorter videos. The statement of the online lessons being divided into manageable segments, 85.7% agreed or strongly agreed versus 0% in disagreement. This suggests that the splitting of longer videos was helpful for students.

The last set of questions was on the online watching behavior of the students. We first asked how often they multitasked while watching the online lessons; 28.6% of the responded fairly often with 14.3% responding that they never did. This was disappointing given that we spent more time teaching the students how to watch videos and even split them into shorter videos to maintain their attention span. However given that less than half agreed that they were engaging this seems expected. When asked how often they skipped part of a video, 57.1% of the students responded fairly often or very often. Finally when asked about how often they rewound or repeated parts of the video 28.6% said that they did fairly often or very often. This is an improvement over the fall courses and likely due to the

time spent teaching the students how to watch the videos.

3.5.2 Evaluation of the Students

Given the time restraint of the summer course we did not give the 3 exams as is traditionally done in the fall. With that said we have no way to compare the exams as we did with the fall course. We did however give the students the same final exam question as was given in the fall courses BIO 200, ID 200 and BIO 201. Table 3.7 shows the mean and median score for this problem. We note that ID 200 in the summer has a lower mean and median though not statistically significant. We feel much of this has to do with the students working their way from summary statistics to logistic regression in only 6 weeks. With this time constraint we expect the scores on a regression question to be slightly lower as they were here.

3.6 Discussion

With these 3 courses we were able to look at students perspectives as well as an evaluation of the students compared to current and past students. The suggestions made in section 3.3.3 were made for the summer course and seemed to have made a positive impact on the results. With the changes made to video times students felt they were much more manageable. Many students also commented that having the video times for each week helped in planning out their time.

For the in class component the students in the fall felt the course was not integrated. This was addressed and the students felt it was much more congruent. The majority of class time being spent on lab type problems gave the students more time to practice data analysis and Stata while under the guidance of the

instructors. The problem sets were changed to follow more closely with lab problems. This gave us more opportunity to see the thought process as well as the work that the students used to arrive at their answers. This is very important to understanding how they grasp the material.

We feel that our main take away points from these three courses is that the students perform as well as the traditional classroom setting. When making changes directed by the students evaluation of the course we can see a dramatic change in the acceptance of the course. We suggest that when implementing courses in this manner that surveying the students during and after the course will allow the instructor to best meet the needs of their particular students.

We feel that it is good practice to keep the videos short or possibly under 10 minutes and to publish the times so that students can adequately plan. We also have seen that students do not inherently know how to take a course with online videos. More needs to be done to instruct them on the importance of note taking and not multitasking while watching the videos.

For the in class component of the class it is important to give the students practical problems to solve. Keeping the review to a minimum places more emphasis on the importance of watching the videos before class as well as giving them more time to work on real problems. It appears that smaller class sizes seem to help students feel more engaged. If this is not possible we suggest having a room where the students can be in groups and where instructors and teaching assistants are making sure to make contact with each and every group.

Utilizing problem sets that give similar types of data analysis questions allows for the instructor to gauge student's grasp of the material. We maintain the suggestion to understand the error variance of an automatic style of grading. If

possible we suggest that written assignments be utilized so that there is better understanding of students processes as they answer problems.

Blended learning classes can be useful and enjoyed by the students. More work is needed to be done in order to evaluate the quality of learning done. Our study is limited in that we compared students to students from 20 years prior and to students in a different but similar course in the same semester. If possible randomizing students to blended vs traditional style of teaching would allow us to evaluate blended learning further.

References

[1] Accredidation criteria: public health programs. http://ceph.org/assets/PHP-Criteria-2011.pdf. Accessed: 2014-04-25.

[2] Knowledge program developed at cleveland clinic's neurological institute. http://my.clevelandclinic.org/neurological_institute/about/default.aspx, 2008-2013.

[3] Mellen center for multiple sclerosis treatment and research, cleveland clinic, neurological institute. http://my.clevelandclinic.org/neurological_institute/mellen-center-multiple-sclerosis/default.aspx, 2013.

[4] Onyebuchi A Arah, Yasutaka Chiba, and Sander Greenland. Bias formulas for external adjustment and sensitivity analysis of unmeasured confounders. Annals of epidemiology, 18(8):637–646, 2008.

[5] Kenneth A Bollen. Total, direct, and indirect effects in structural equation models. Sociological methodology, 17(1):37–69, 1987.

[6] Kenneth A. Bollen. Structural Equations with Latent Variables. John Wiley & Sons, New York, New York, 1989.

[7] Kenneth A Bollen and Robert Stine. Direct and indirect effects: Classical and bootstrap estimates of variability. Sociological methodology, 20(1):15–140, 1990.

[8] William G Bowen, Matthew M Chingos, Kelly A Lack, and Thomas I Nygren. Interactive learning online at public universities: Evidence from a six-campus randomized trial. Journal of Policy Analysis and Management, 33(1):94–111, 2014.

[9] N. E. Breslow and N. E. Day. Statistical methods in cancer research. Volume I - The analysis of case-control studies. IARC Sci. Publ., (32):5–338, 1980.

[10] I. D. Bross. Spurious effects from an extraneous variable. J Chronic Dis, 19 (6):637–647, Jun 1966.

[11] Babette A Brumback, Miguel A Hernán, Sebastien JPA Haneuse, and James M Robins. Sensitivity analyses for unmeasured confounding assuming a marginal structural model for repeated measures. Statistics in medicine, 23 (5):749–767, 2004.

[12] R. A. Charlton, S. Landau, F. Schiavone, T. R. Barrick, C. A. Clark, H. S. Markus, and R. G. Morris. A structural equation modeling investigation of age-related variance in executive function and DTI measured white matter damage. Neurobiol. Aging, 29(10):1547–1555, Oct 2008.

[13] Zhongzhou Chen, Timothy Stelzer, and Gary Gladding. Using multimedia modules to better prepare students for introductory physics lecture. Physical Review Special Topics-Physics Education Research, 6(1):010108, 2010.

[14] J. Cheong, D. P. Mackinnon, and S. T. Khoo. Investigation of Mediational Processes Using Parallel Process Latent Growth Curve Modeling. Struct Equ Modeling, 10(2):238, Apr 2003.

[15] Eun Man Choi. Applying inverted classroom to software engineering education. International Journal of e-Education, e-Business, e-Management and e-Learning, 3(2):121–125, 2013.

[16] D. A. Cole and S. E. Maxwell. Testing mediational models with longitudinal data: questions and tips in the use of structural equation modeling. J Abnorm Psychol, 112(4):558–577, Nov 2003.

[17] John B Copas and HG Li. Inference for non-random samples. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 59(1):55–95, 1997.

[18] J. Cornfield, W. Haenszel, E. C. Hammond, A. M. Lilienfeld, M. B. Shimkin, E. L. Wynder, J. Cornfield, W. Haenszel, E. C. Hammond, A. M. Lilienfeld, M. B. Shimkin, and E. L. Wynder. Smoking and lung cancer: recent evidence and a discussion of some questions. 1959. Int J Epidemiol, 38(5):1175–1191, Oct 2009.

[19] Randall S Davies, Douglas L Dean, and Nick Ball. Flipping the classroom and instructional technology integration in a college-level information systems spreadsheet course. Educational Technology Research and Development, 61 (4):563–580, 2013.

[20] J. A. Day and J. D. Foley. Evaluating a web lecture intervention in a human–computer interaction course. IEEE Trans. on Educ., 49(4):

420–431, November 2006. ISSN 0018-9359. doi: 10.1109/TE.2006.879792. URL http://dx.doi.org/10.1109/TE.2006.879792.

[21] Terry E Duncan, Susan C Duncan, and Lisa A Strycker. An introduction to latent variable growth curve modeling: Concepts, issues, and application. Routledge Academic, 2013.

[22] Emilio Ferrer and John McArdle. Alternative structural models for multivariate longitudinal data analysis. Structural Equation Modeling, 10(4): 493–524, 2003.

[23] W Dana Flanders and Mum J Khoury. Indirect assessment of confounding: graphic description and limits on effect of adjusting for covariates. Epidemiology, 1(3):239–246, 1990.

[24] Henk Folmer. Measurement of the effects of regional policy instruments by means of linear structural equation models and panel data. Environment and Planning A, 13(11):1435–1448, 1981.

[25] MH Gail, S Wacholder, and JH Lubin. Indirect corrections for confounding under multiplicative and additive risk models. American journal of industrial medicine, 13(1):119–130, 1988.

[26] Consensus Group Goldman. The goldman consensus statement on depression in multiple sclerosis. Multiple sclerosis (Houndmills, Basingstoke, England), 11(3):328, 2005.

[27] H. F. Gollob and C. S. Reichardt. Taking account of time lags in causal models. Child Dev, 58(1):80–92, Feb 1987.

[28] Sander Greenland. The impact of prior distributions for uncontrolled confounding and response bias: a case study of the relation of wire codes and magnetic fields to childhood leukemia. Journal of the American Statistical Association, 98(461):47–54, 2003.

[29] Sander Greenland. Multiple-bias modelling for analysis of observational data. Journal of the Royal Statistical Society: Series A (Statistics in Society), 168 (2):267–306, 2005.

[30] D. D. Gunzler, A. Perzynski, N. Morris, R. Bermel, S. Lewis, and D. Miller. Disentangling Multiple Sclerosis and depression: an adjusted depression screening score for patient-centered care. J Behav Med, 38(2):237–250, Apr 2015.

[31] Kosuke Imai, Luke Keele, and Dustin Tingley. A general approach to causal mediation analysis. Psychological methods, 15(4):309, 2010.

[32] Kosuke Imai, Luke Keele, and Teppei Yamamoto. Identification, inference, and sensitivity analysis for causal mediation effects. Statistical Science, 25(1): 51–71, 2010. URL http://imai.princeton.edu/research/mediation.html.

[33] Guido W Imbens. Sensitivity to exogeneity assumptions in program evaluation. American Economic Review, pages 126–132, 2003.

[34] E. M. Kitagawa. Components of a difference between two rates. J Am Stat Assoc., 50(272):1168–1194, 1955.

[35] Kurt Kroenke and Robert L Spitzer. The phq-9: a new depression diagnostic and severity measure. Psychiatr Ann, 32(9):1–7, 2002.

[36] T. Lange and J. V. Hansen. Direct and indirect effects in a survival context. Epidemiology, 22(4):575–581, Jul 2011.

[37] Timothy L Lash, Matthew P Fox, and Aliza K Fink. Applying quantitative bias analysis to epidemiologic data. Springer, 2011.

[38] J Scott Lewis and Marissa A Harrison. Online delivery as a course adjunct promotes active learning and student success. Teaching of Psychology, 39(1): 72–76, 2012.

[39] D. Y. Lin, B. M. Psaty, and R. A. Kronmal. Assessing the sensitivity of regression results to unmeasured confounders in observational studies. Biometrics, 54(3):948–963, Sep 1998.

[40] DY Lin, BM Psaty, and RA Kronrnal. Assessing the sensitivity of regression results to unmeasured confounders in observational studies. Biometrics, 54(3): 948–963, 1998.

[41] David Peter MacKinnon. Introduction to statistical mediation analysis. Routledge, 2008.

[42] David J Marcey and Michael E Brint. Transforming an undergraduate introductory biology course through cinematic lectures and inverted classes: A preliminary assessment of the clic model of the flipped classroom. In Biology Education Research Symposium at the meeting of the National Association of Biology Teachers, volume 12, 2012.

[43] Scott E Maxwell and David A Cole. Bias in cross-sectional analyses of longitudinal mediation. Psychological methods, 12(1):23, 2007.

[44] John J McArdle and Fumiaki Hamagami. Latent difference score structural models for linear dynamic analyses with incomplete longitudinal data. 2001.

[45] John J McArdle and John R Nesselroade. Growth curve analysis in contemporary psychological research. Handbook of psychology, 2003.

[46] Lawrence C McCandless, Paul Gustafson, and Adrian Levy. Bayesian sensitivity analysis for unmeasured confounding in observational studies. Statistics in medicine, 26(11):2331–2347, 2007.

[47] John McGready and Ron Brookmeyer. Evaluation of student outcomes in online vs. campus biostatistics education in a graduate school of public health. Preventive medicine, 56(2):142–144, 2013.

[48] William Meredith and John Tisak. Latent curve analysis. Psychometrika, 55 (1):107–122, 1990.

[49] B. O. Muthén and L. K. Muthén. Mplus (Version 7.2) [Computer software], 1998–2014).

[50] Bengt O Muthén and Patrick J Curran. General longitudinal modeling of individual differences in experimental designs: A latent variable framework for analysis and power estimation. Psychological methods, 2(4):371, 1997.

[51] Marcello Pagano, Kimberlee Gauvreau, and Marcello Pagano. Principles of biostatistics. Duxbury Pacific Grove, CA, 2000.

[52] Judea Pearl. Causality: models, reasoning and inference, volume 29. Cambridge Univ Press, 2000.

[53] Judea Pearl. Direct and indirect effects. In Proceedings of the seventeenth conference on uncertainty in artificial intelligence, pages 411–420. Morgan Kaufmann Publishers Inc., 2001.

[54] L. Penke and I. J. Deary. Some guidelines for structural equation modelling in cognitive neuroscience: the case of Charlton et al.'s study on white matter integrity and cognitive ageing. Neurobiol. Aging, 31(9):1656–1660, Sep 2010.

[55] Rod Pierce. "quincunx". http://www.mathsisfun.com/data/quincunx.html. Accessed: 2014-08-21.

[56] J. M. Robins and S. Greenland. Identifiability and exchangeability for direct and indirect effects. Epidemiology, 3(2):143–155, Mar 1992.

[57] James M Robins, Andrea Rotnitzky, and Daniel O Scharfstein. Sensitivity analysis for selection bias and unmeasured confounding in missing data and causal inference models. Springer, 2000.

[58] P. R. Rosenbaum and D. B. Rubin. Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome. J R Stat Soc Series B., 45(2):212–218, 1983.

[59] Paul R Rosenbaum. Observational studies. Springer, 2002.

[60] J. J. Schlesselman. Assessing effects of confounding variables. Am. J. Epidemiol., 108(1):3–8, Jul 1978.

[61] Michael E Sobel. Asymptotic confidence intervals for indirect effects in structural equation models. Sociological methodology, 13(1982):290–312, 1982.

[62] Michael E Sobel. Some new results on indirect effects and their standard errors in covariance structure models. Sociological methodology, 16:159–186, 1986.

[63] Michael E Sobel. Effect analysis and causation in linear structural equation models. Psychometrika, 55(3):495–515, 1990.

[64] StataCorp. Stata statistical software: Release 13, 2013.

[65] Bethany B Stone. Flip your classroom to increase active learning and student engagement. In Proceedings from 28th Annual Conference on Distance Teaching & Learning, Madison, Wisconsin, USA, 2012.

[66] Til Stürmer, Sebastian Schneeweiss, Jerry Avorn, and Robert J Glynn. Adjusting effect estimates for unmeasured confounding with validation data using propensity score calibration. American journal of epidemiology, 162(3): 279–289, 2005.

[67] Eric J Tchetgen Tchetgen, Ilya Shpitser, et al. Semiparametric theory for causal mediation analysis: Efficiency bounds, multiple robustness and sensitivity analysis. The Annals of Statistics, 40(3):1816–1845, 2012.

[68] L. Valeri and T. J. Vanderweele. Mediation analysis allowing for exposure-mediator interactions and causal interpretation: theoretical assumptions and implementation with SAS and SPSS macros. Psychol Methods, 18(2):137–150, Jun 2013.

[69] M. J. van der Laan and M. L. Petersen. Direct effect models. Int J Biostat, 4 (1):Article 23, 2008.

[70] T. J. Vanderweele. Sensitivity analysis: distributional assumptions and confounding assumptions. Biometrics, 64(2):645–649, Jun 2008.

[71] T. J. VanderWeele. Bias formulas for sensitivity analysis for direct and indirect effects. Epidemiology, 21(4):540–551, Jul 2010.

[72] T. J. VanderWeele. Invited commentary: structural equation models and epidemiologic analysis. Am. J. Epidemiol., 176(7):608–612, Oct 2012.

[73] T. J. Vanderweele and O. A. Arah. Bias formulas for sensitivity analysis of unmeasured confounding for general outcomes, treatments, and confounders. Epidemiology, 22(1):42–52, Jan 2011.

[74] Tyler VanderWeele and Stijn Vansteelandt. Conceptual issues concerning mediation, interventions and composition. Statistics and its Interface, 2: 457–468, 2009.

[75] Stijn Vansteelandt, Maarten Bekaert, and Theis Lange. Imputation strategies for the estimation of natural direct and indirect effects. Epidemiologic Methods, 1(1):131–158, 2012.

[76] John B Willett and Aline G Sayer. Using covariance structure analysis to detect correlates and predictors of individual change over time. Psychological Bulletin, 116(2):363, 1994.

[77] Takashi Yanagawa. Case-control studies: assessing the effect of a confounding factor. Biometrika, 71(1):191–194, 1984.

[78] Yiran Zhao and Lori Breslow. Literature review on hybrid/blended learning. 2013.

[79] Yiran Zhao and Andrew Ho. Evaluating the flipped classroom in an undergraduate history course. 2014.