

# The Intermediate Endpoint Effect in Logistic and Probit Regression

MacKinnon, D. P. \*  
Lockwood, C. M. \*  
Brown, C. H. †  
Wang, W. ‡  
Hoffman, J. M. §

April 26, 2007

We thank our colleagues in the Prevention Science and Methodology Group (PSMG) as well as anonymous reviewers for many helpful comments on this paper. We thank Ghulam Warsi and Myeongsun Yoon for assistance with this research and Christopher Winship for comments on the manuscript. We thank Ellen Laing and L. Terri Singer for their help in producing this manuscript. We also thank Mary Ann Pentz for the use of data from the Students Taught Awareness and Resistance study. Part of this work was presented at the 1991 Psychometric Society Meeting. This research was supported by U. S. Public Health Service Grants R01-DA09757, R01-MH40859, R01-MH42968, and P30-MH068685.

RUNNING HEAD: INTERMEDIATE ENDPOINT EFFECT

Address correspondence to:  
David P. MacKinnon, Ph.D.  
Professor  
Dept. of Psychology  
Arizona State University  
Tempe, AZ 85287-1104  
480 727-6120  
480 727-6176 FAX  
Email: davidpm@asu.edu

---

\*Department of Psychology, Arizona State University, Tempe, AZ

†Department of Epidemiology and Biostatistics, University of South Florida and Center for Health Statistics, University of Illinois at Chicago

‡Department of Epidemiology and Biostatistics, University of South Florida

§Department of Rehabilitative Medicine, University of Washington, Seattle, WA

## Abstract

**Background** An intermediate endpoint is hypothesized to be in the middle of the causal sequence relating an independent variable to a dependent variable. The intermediate variable is also called a surrogate or mediating variable and the corresponding effect is called the mediated, surrogate endpoint, or intermediate endpoint effect. Clinical studies are often designed to change an intermediate or surrogate endpoint and through this intermediate change influence the ultimate endpoint. In many intermediate endpoint clinical studies the dependent variable is binary, and logistic or probit regression is used.

**Purpose** The purpose of this study is to describe a limitation of a widely used approach to assessing intermediate endpoint effects and to propose an alternative method, based on products of coefficients, that yields more accurate results.

**Methods** The intermediate endpoint model for a binary outcome is described for a true binary outcome and for a dichotomization of a latent continuous outcome. Plots of true values and a simulation study are used to evaluate the different methods.

**Results** Distorted estimates of the intermediate endpoint effect and incorrect conclusions can result from the application of widely used methods to assess the intermediate endpoint effect. The same problem occurs for the proportion of an effect explained by an intermediate endpoint, which has been suggested as a useful measure for identifying intermediate endpoints. A solution to this problem is given based on the relationship between latent variable modeling and logistic or probit regression.

**Limitations** More complicated intermediate variable models are not addressed in the study, although the methods described in the article can be extended to these more complicated models.

**Conclusions** Researchers are encouraged to use an intermediate endpoint method based on the product of regression coefficients. A common method based on difference in coefficient methods can lead to distorted conclusions regarding the intermediate effect.

*Keywords:* Mediation, surrogate endpoint, logistic regression, probit regression, latent variables.

## 1 Introduction

The purpose of this paper is to describe why standard statistical approaches for examining the intermediate endpoint or mediated effect using probit and logistic regression can be inaccurate and how to correct them. The focus of this paper is the case in which the dependent variable is binary such as dead/alive, presence/absence of disease, or use/non-use of drugs. This paper examines methodology for modeling how an active intervention condition, compared to control, affects this outcome through a continuous intermediate or mediating variable. Logistic or probit regression are traditionally used to examine this mediation effect. Such models are ubiquitous in clinical trials where the effect of a dichotomous treatment variable on a diagnosis or other dichotomous outcome is mediated by an intermediate variable.

The notion of an intervening or mediating variable is well established in clinical studies of a

binary disease outcome [1], [2] where the mediated effect is termed the surrogate or intermediate endpoint effect. Surrogate endpoints are a subset of mediator variables where the theoretical and statistical requirements for a variable to be a surrogate endpoint are more stringent than those for mediators in general. Surrogates are expected to explain the full impact on a distal outcome; a mediator may explain part of this effect. One important use of an intermediate variable is in situations where the ultimate outcome is difficult or costly to obtain. In both prevention and treatment trials, the extended length of time for a disease or death to occur and the low incidence rates often require exorbitantly large sample sizes to study correlates of the disease. In this situation, researchers advocate the use of surrogate or intermediate endpoints [3], [4]. Surrogate endpoints are more frequent or more proximate to the prevention strategy and are therefore easier to study. Examples of surrogate endpoints are serum-cholesterol levels for the ultimate outcome of coronary heart disease [5], measures of immune system response for the ultimate outcome of death in HIV infected individuals [6], and presence of polyps for the ultimate outcome of colon cancer [7]. The use of surrogate endpoints rests on the mediation assumption that the independent variable causes the surrogate endpoint, which, in turn, causes the ultimate outcome [8], [9]. As a result, decomposing the effects by which a surrogate explains impact on a distal outcome is critically important because the surrogate endpoint is considered to be causally related to the outcome. The importance of surrogate endpoints was recently summarized by Begg and Leung [8] (p.27), "Above all else, we believe that the issue of when and how to use surrogate endpoints is probably the pre-eminent contemporary problem in clinical trials methodology, so it merits much extensive scrutiny."

Prentice [9](p.432) defined a surrogate endpoint as a response variable for which a test of the null hypothesis of no relationship to the treatment groups under comparison is also a valid test of the corresponding null hypothesis based on the true endpoint. Freedman et al. [10] proposed

the proportion of the treatment effect explained by an intermediate variable to test this surrogate endpoint effect. A value of 100% indicates that the surrogate endpoint explains all of the relation between the treatment and the dependent variable, therefore satisfying Prentice's definition. If the proportion mediated is less than 100%, it indicates that some of the relation may not be explained by this intermediate variable and other causal mechanisms may be neglected [11]. The proportion measure reflects the size of the surrogate endpoint (i.e., mediated) effect as well as the amount of the treatment effect explained by the surrogate endpoint. The proportion mediated has not been accepted without criticism [12], namely that accurate identification of surrogate endpoints requires measurement of the ultimate outcome [8], that the proportion measure can be negative or greater than one [13], and that values of the proportion mediated are often very small. Furthermore, other research has shown that the proportion mediated tends to be unstable unless the sample size is large or the effect size is very large [14, 15]. In response to the limitations of the proportion measure, Buyse and colleagues proposed two criteria for a surrogate endpoint (see [16, 17] for more specific information on these criteria). The first criteria is that the total effect relating the treatment to the dependent variable divided by the effect of the treatment on the mediator or surrogate variable, is close to 1. The logic here is that an intermediate endpoint should be affected by a treatment to the same magnitude as the treatment affects the dependent variable. A second criterion requires that the relation between the mediator and the dependent variable is also substantial. Both the proportion mediated and the measures proposed by Buyse and colleagues are important because they combine a measure of effect size along with a measure of the mediated effect. These Buyse et al. criteria do not include an estimator of the intermediate endpoint effect so they are not addressed in this article. Other approaches to identification of surrogate endpoints focus on the importance of meta analytical combination of results from individual studies [18, 19] and more detailed approaches

to causal inference from one study [20]. The present article focuses on estimators of the intermediate endpoint effect from a single study.

In addition to replacing costly outcome measures, intermediate variables can also be used to elucidate how an intervention leads to changes in the outcome variable. A common example is in the field of prevention research where programs are designed to change intermediate variables that prevent negative outcomes. An example of testing mediation with binary outcomes was described by Foshee and colleagues [21] in which a program to prevent adolescent dating violence was hypothesized to work by changing norms, decreasing gender stereotyping, improving conflict management, changing beliefs about the need for help, increasing awareness of services for victims and perpetrators, and increasing help-seeking behavior. The researchers compared the values of the logistic regression coefficient of treatment predicting the binary dating violence outcome with and without controlling for the proposed mediators. The authors considered evidence of mediation to occur when the difference between the two values was 20% of the value of the unadjusted coefficient (i.e., 20% of the program effect was explained by the mediators). Given this criterion, results indicated that the treatment effect on sexual violence was mediated by changes in norms, gender stereotyping, and awareness of victim services.

As described by several researchers, the examination of mediation in the evaluation of prevention programs and clinical trials is important for at least two major reasons [22, 23, 24]. First, most trials are based on changing mediating variables hypothesized to be related to the outcome measure. It is important to assess whether the program successfully changed the mediating variables it was designed to change. If the program was not able to change the mediating variable, then it would not be surprising that effects on outcomes were not observed. Second, the extent to which change in one or more mediating variables account for observed program effects sheds light on

the theoretical basis of the program. This information has practical implications as well, because ineffective intervention components may be dropped if they do not contribute to intervention success or intervention components may be added if the original intervention only works through some but not all hypothesized mediators. Statistical methods for assessing mediation for continuous outcomes have been described [25, 26], as well as methods for specific statistical models such as covariance structure models [27] and multilevel models [28]. Issues related to mediation analysis for categorical outcomes have not received much research attention.

Two general approaches have been offered to quantify mediation [23, 24, 29]. In the situation where the dependent and mediating variables are continuous and normal-based models are used, one way to quantify mediation is to compare the regression coefficient of the outcome on the independent variable both without adjustment for the mediator ( $\beta_z$  in Equation 5 below) and with adjustment for the mediator ( $\gamma_z$  in Equation 1 below). If the unadjusted independent variable coefficient  $\beta_z$  is nonzero and the adjusted coefficient  $\gamma_z$  is zero when the mediator is included in the model, the effect of the independent variable is entirely mediated by the mediating variable. Standard errors for the mediated effect are available [26] and are generally accurate [15]. This is called the difference in coefficients method.

A second general method, based on path analysis and more commonly used in the social sciences for continuous outcomes, treats mediation as the product of two regression coefficients, the regression of the mediator on the independent variable ( $\alpha_z$  in Equation 2 below) and the partial regression coefficient of the outcome on the mediator, adjusted for the independent variable ( $\gamma_m$  in Equation 1 below) [30]. This method is called the product of coefficients method [26].

For standard least squares regression models without missing data, the difference in coefficients and the product of coefficients estimators are identical [15]. For binary response variables, the two

estimators are not identical and can differ dramatically, as shown below. The most widely used estimator of the standard error of this mediated effect was derived by Sobel [27] using the multivariate delta method under the assumption of normality (see MacKinnon et al. [26] for examples of other standard errors).

The proportion mediated measure,  $1 - \gamma_z/\beta_z$  (from Equations below), was suggested by Freedman et al. [10]; we note that this measure is not necessarily between 0 and 1. Other possible definitions of proportion mediated measures are  $\gamma_m\alpha_z/\beta_z$  and  $\gamma_m\alpha_z/(\gamma_z + \gamma_m\alpha_z)$  (from Equations described below). In the continuous dependent variable case with no missing data, all three proportion measures for the mediated effect are equal. Because logistic and probit regression are nonlinear models, the point estimates of these two quantities are not equal. In this paper we show some of the estimates have low bias while others require standardization procedures in order to be appropriate.

The mediated effect along with its standard error can be used to create confidence limits for the mediated effect [15]. For dichotomous outcomes, Freedman et al. [10] proposed a variance estimator of the difference in coefficients method. The results of this paper suggest a preference for the product of coefficients method.

## **1.1 Complications with the Estimation of the Intermediate Variable Effect with a Binary Dependent Variable**

When the dependent variable is binary and logistic regression is used, the difference in coefficients and product of coefficients methods for calculating the mediated effect are not equal as MacKinnon and Dwyer [24] demonstrated with a simulation study. This article will examine estimation methods for mediation analysis with binary outcomes, describe a problem that arises with the estimation, and propose a solution. The solution is demonstrated with analytical methods and a simulation

study that investigates both small sample properties and robustness to distributional assumptions. Finally, we present a case study of a randomized preventive trial to prevent later cigarette use in adolescence.

## 2 Estimation of the Mediated Effect with Binary Outcomes

Let  $Z$  be a binary variable representing intervention status (1 for active intervention and 0 for control), and let the binary  $Y$  be the major outcome variable (diagnosis, death, or improvement). Define  $\mathbf{X}$  to be  $p$ -dimensional covariates measured at baseline. Let  $M$  be the hypothesized surrogate variable. In most trials,  $M$  is measured with a continuous variable, rather than a dichotomous one, since the former provides much greater statistical power. For both practical and statistical reasons, we consider the case of a continuous  $M$  variable, with normal or non-normal errors conditional on intervention condition.

As in the case of a continuous dependent variable, the mediated effect can be calculated in two ways when the dependent variable is binary [24]. To calculate the mediated effect both methods use information from two of the three equations listed below. In this model we assume no treatment by baseline interaction or moderation.

First, Equation 1 shows a logistic regression model where  $Y$  depends on  $M$  as well as on baseline



covariates  $\mathbf{X}$  and intervention condition  $Z$ <sup>1</sup>

$$\text{logit } Pr\{Y = 1 | \mathbf{X} = \mathbf{x}, M = m, Z = z\} = \gamma_0 + \gamma'_x \mathbf{x} + \gamma_m m + \gamma_z z, \quad z = 1, 0. \quad (1)$$

where the  $\gamma$  coefficients refer to intercept ( $\gamma_0$ ), covariates ( $\gamma'_x$ ), and intervention ( $\gamma_z$ ) after adjustment for the mediator. For probit regression, a similar expression exists, with logit replaced by the inverse distribution function for a standard normal.

Second, a linear relationship for the relation between  $Z$  and  $M$  is assumed.

$$M = \alpha_0 + \alpha'_x \mathbf{x} + \alpha_z z + \varepsilon_m \quad (2)$$

$$E(\varepsilon_m | \mathbf{x}, z) = 0, \quad z = 0, 1 \quad (3)$$

The  $\alpha$  coefficients refer to the prediction of  $M$  (ignoring  $Y$  completely). Like the previous model, this model assumes no baseline by intervention interaction. Also, most analytic models impose a distributional assumption on the error, such as normality with a homogeneous variance. This distributional assumption is more restrictive than the conditional mean specifications given in Equations 2 and 3. To specify this exact distribution,  $f_{M|XZ}(m|\mathbf{x}, z)$  refers to the conditional density of  $M$  given covariates  $\mathbf{X} = \mathbf{x}$  and intervention condition  $Z = z$ .

The subscript  $i$  is used to indicate subject level covariates ( $\mathbf{X}_i$ ), intervention assignment ( $Z_i$ ), mediator ( $M_i$ ), and distal outcome ( $Y_i$ ), for subjects  $i = 1, \dots, n$ . It is assumed that each mediator

---

<sup>1</sup>In the social science literature four symbols are often used to express the mediation equations. Our notation in this paper allows other covariates and is more in line with notation used in standard logistic regression modeling. This footnote shows their comparability. The symbol  $a$ , the regression of the mediator on the intervention condition, is given by  $\alpha_z$  in this paper's notation. The symbol  $b$ , the partial regression of the outcome on the mediator adjusted for the intervention condition, is given by  $\gamma_m$  in our notation. The symbol  $c$  for the regression of the outcome on the intervention condition without adjusting for the mediator is given by  $\beta_z$  in our notation. Finally,  $c'$ , the partial regression of the outcome on the intervention condition adjusted for the mediator, is given by  $\gamma_z$ .

is independent of all other mediator outcomes and is given by Equation 2 – 3 conditional on all covariates and intervention assignments. Also, given all covariates, intervention assignments, and mediators, the distal outcomes  $Y_1, \dots, Y_n$  are all independent and follow the distribution specified in Equation 1.

Under the above assumptions, as sample size increases, asymptotically unbiased estimates of  $\alpha$ 's and  $\gamma$ 's can be estimated through standard least squares and logistic regression respectively. Further, the marginal relationship between  $Z$  and  $Y$ , conditional on  $\mathbf{X}$  but unconditional on  $M$ , is given by

$$Pr\{Y = 1 | \mathbf{X} = \mathbf{x}, Z = z\} = \int_{-\infty}^{\infty} (1 + e^{-\gamma_0 - \gamma'_x \mathbf{x} - \gamma_m m - \gamma_z z})^{-1} f_{M|Xz}(m | \mathbf{x}, z) dm, \quad z = 1, 0. \quad (4)$$

Equation 4 shows the true relation between  $Z$  and  $Y$ , conditioning on multiple covariates  $\mathbf{X}$  but not depending on  $M$ . This relation is called the marginal relation between  $Z$  and  $Y$  in this paper. Without knowing the explicit form of  $f_{M|Xz}$ , the above integral cannot be evaluated explicitly, and regardless of the form of this error distribution, there is no simple form for this integral. Instead, researchers often introduce a standard logit model for this marginal relationship between  $Y$  and  $\mathbf{X}$  and  $Z$ . In particular, consider the following model:

$$\text{logit } Pr\{Y = 1 | \mathbf{X} = \mathbf{x}, Z = z\} = \beta_0 + \beta'_x \mathbf{x} + \beta_z z, \quad z = 1, 0. \quad (5)$$

Equation 4 reduces to the logit form given in Equation 5 only if there are no covariates [31]. When there are no covariates, the binary nature of  $Z$  allows specification of these parameters in a standard logit model. Even in this case of no covariates, however, there is no analytic way to obtain the exact value of  $\beta_z$  from the  $\alpha$  and  $\gamma$  coefficients; the integral needs to be evaluated numerically or approximated.

## 2.1 Quantifying mediation using the difference of coefficients method

There are two equivalent methods for examining the degree of mediation, difference in coefficients and product of coefficients methods. As will be shown below, these two methods are not the same for binary outcome models. A natural method for quantifying mediation is to compare the coefficients for the marginal relationship between  $Z$  and  $Y$ , to that adjusting for  $M$ . No difference in these coefficients implies that knowledge of  $M$  plays no role in predicting how intervention affects the distal outcome. This method has been proposed for dichotomous outcomes as well, but applied to the approximate marginal model Equation 5 rather than the exact marginal model given by Equation 4. Thus this straightforward (but as we will see inaccurate) difference in coefficients method is based on the parameter difference,  $\beta_z - \gamma_z$ . It is estimated by fitting two separate logistic regression models, one using covariates and intervention status as predictors, the other including the mediator as well. We know of no applied research examples that use the difference in coefficients method with the correct marginal relationship, Equation 4.

## 2.2 Quantifying mediation using the product of coefficients method

The second method for assessing mediation in continuous outcome models relies on a path diagram approach. Mediation depends on the product of the relation between intervention and mediator, and the relation between mediator and distal outcome, adjusted for intervention. A direct application of this method to the regression model of  $M$  on  $Z$  and the logistic regression model of  $Y$  on both  $M$  and  $Z$  leads to examination of  $\alpha_z \gamma_m$ .

### 2.3 Mediation using latent variables

Since there is equivalence between the difference in coefficient and product of coefficient methods with linear, continuous models, it is useful to embed the problem within a continuous model. Investigation of the nonequivalence of the methods clarifies how the two approaches differ statistically and conceptually. The equivalence of the two methods is demonstrated using an alternative representation of these logistic regression models in terms of an underlying latent variable, an approach that goes back at least to Finney [32].

Equation 1 can be expressed equivalently by introducing the unobserved latent variable  $Y^*$  that is linearly related to  $Z$  and  $M$  as well as covariates  $\mathbf{X}$ .

$$Y^* = \gamma_0 + \boldsymbol{\gamma}'_x \mathbf{x} + \gamma_m m + \gamma_z z + \varepsilon_y, \quad z = 1, 0. \quad (6)$$

In this model we let  $\varepsilon_y$  represent the residual variability having a standard logistic distribution, that is,

$$Pr\{\varepsilon_y < u\} = \frac{e^u}{1 + e^u}. \quad (7)$$

We also assume that this error in predicting  $Y^*$  is independent of the error predicting  $M$  from  $Z$  in Equation 2. The dichotomous, observable  $Y$  is derived from  $Y^*$  through the relation  $Y = 1$  if and only if  $Y^* > 0$ . This definition of  $Y$  in terms of  $Y^*$  and the logit error distribution together produce a model that is identical to that of Equation 1.

In Equation 6, we can substitute the expression for  $M$  in terms of  $Z$  that is explicit in Equation

2, obtaining

$$\begin{aligned}
Y^* &= \gamma_0 + \gamma'_x \mathbf{x} + \gamma_m(\alpha_0 + \alpha'_x \mathbf{x} + \alpha_z z + \varepsilon_m) + \gamma_z z + \varepsilon_y \\
&= \gamma_0 + \gamma_m \alpha_0 + (\gamma_x + \gamma_m \alpha_x)' \mathbf{x} +
\end{aligned} \tag{8}$$

$$(\gamma_z + \gamma_m \alpha_z) z + \tag{9}$$

$$\gamma_m \varepsilon_m + \varepsilon_y \tag{10}$$

The next to last line above shows that the mean difference on  $Y^*$  for active intervention versus control when not adjusting for  $M$  is  $(\gamma_z + \gamma_m \alpha_z)$ . The mean difference on  $Y^*$  when adjusting for  $M$  is  $\gamma_z$ . Thus the product of coefficients,  $\gamma_m \alpha_z$  is identical to the difference in coefficients on the  $Y^*$  scale. This underlying equivalence of the two methods when applied to the mean of  $Y^*$  holds regardless of the distribution of  $\varepsilon_m$  as long as these errors have zero conditional means as in Equation 3.

## 2.4 Estimation and variance of mediated effects with binary outcomes

This section shows that the estimates based on the product of coefficients method produce asymptotically unbiased estimates of the mediated effect while those derived from the difference in coefficients method must be scaled properly in order to reduce zero-order bias. We also show that variance estimates from the General Estimating Equation (GEE, see [33]) approach can be used to form confidence intervals and testing. Appendix A shows the consistency of the product of coefficients method for logistic regression when  $Y|\mathbf{X}MZ$  is given by Model 1 and the conditional mean of  $M|\mathbf{X}Z$  is linear as in Equations 2 and 3. Appendix A also shows the asymptotic independence of  $\hat{\gamma}$  and  $\hat{\alpha}_z$ .

### 2.4.1 Standardizing mediated effects based on the difference of coefficients method

In this section, it is shown that the ordinary estimates obtained from a logistic regression of  $Y$  on  $\mathbf{X}$  and  $Z$  based on Equation 5 are not consistent for  $\gamma_z + \gamma_m \alpha_z$ . We start from the true latent variable representation in Equations 8 - 10. Then

$$Pr\{Y = 1 | \mathbf{X} = \mathbf{x}, Z = z\} = Pr\{\gamma_m \varepsilon_m + \varepsilon_y > -(\gamma_0 + \gamma_m \alpha_0) - (\gamma_x + \gamma_m \alpha_x)' \mathbf{x} - (\gamma_z + \gamma_m \alpha_z) z\} \quad (11)$$

Unless  $\gamma_m = 0$ , this probability function will not have a logistic form. More importantly, because the variance of  $\gamma_m \varepsilon_m + \varepsilon_y$  is larger than that of a standard logistic error, all of the logistic regression estimates of the intercept and slopes for  $\mathbf{x}$  and  $z$  will be biased. It is possible to make adjustments for this higher variance, noting that a standard logistic regression error has variance  $Var(\varepsilon_y) = \pi^2/3$  [35] while the marginal variance is  $Var(\gamma_m \varepsilon_m + \varepsilon_y) = \gamma_m^2 Var(\varepsilon_m) + \pi^2/3$ . Thus if we first perform logistic regression of  $Y$  on  $\mathbf{x}$  and  $z$ , then multiply all the logistic regression estimates, including  $\hat{\beta}_z$  by this factor,

$$\hat{\beta}_z^{logit.standard} = \hat{\beta}_z \sqrt{3(\hat{\gamma}_m^2 \widehat{Var}(\varepsilon_m) + \pi^2/3)/\pi^2}, \quad (12)$$

the coefficient will at least be scaled properly. In the tables below we refer to this as the standardized logistic solution and investigate how this simple adjustment compares with the unstandardized solutions. Similarly, the proportion mediated can then be estimated by  $1 - \hat{\gamma}_z / \hat{\beta}_z^{logit.standard}$ .

For probit regression, the residual variance is fixed to one while  $Var(\gamma_m \varepsilon_m + \varepsilon_y) = \gamma_m^2 Var(\varepsilon_m) + 1$ . Thus the standardized probit solution is to adjust the routine probit regression estimate of  $\hat{\beta}_z$  by

$$\hat{\beta}_z^{probit.standard} = \hat{\beta}_z \sqrt{1 + \hat{\gamma}_m^2 \widehat{Var}(\varepsilon_m)}. \quad (13)$$

When using the standardized estimate of  $\beta_z$ , the standard error of the estimate needs to be

scaled as well. Bootstrap resampling can also be used to obtain the standard error.

### 3 Simulation Study

A simulation study was conducted to demonstrate the discrepancy between  $\beta_z - \gamma_z$  and  $\gamma_m \alpha_z$  and to investigate the performance of different estimators for the proportion mediated. The mean and variance of different estimators of the mediated effect and the proportion mediated were investigated as a function of sample size and size of relations among variables. In addition, the simulation allows the exploration of robustness to both the logistic assumption in Equation 4 and to the normality assumption for the distribution of  $M|Z$ .

The  $Z$  variable was always dichotomous with equal numbers in each group, and the mediator  $M$  was either normally distributed, distributed as a t-distribution with 3 degrees of freedom, or having large skewness in the error; these were generated using a chi-square distribution on 3 degrees of freedom. This latter situation was chosen to mimic the primary departure from normality that we observed in our actual data example below. For all of these cases we still apply least squares so that we can examine the impact of misspecifying the error distribution for  $M|Z$ .

The outcome variable  $Y^*|MZ$  was generated to have either an underlying logistic or a normal error distribution. Thus with the assumption of a logit error,  $Y|MX$  follows Equation 1 with a similar probit model for normal errors. Estimates were compared for both logistic and probit regression with underlying probit errors so that we can assess the impact of misspecifying the conditional model for  $Y|MX$ . Simulated sample sizes were 50, 100, 200, 500, 1000, and 5000. Parameter values for  $\alpha_z$ ,  $\gamma_m$ , and  $\gamma_z$  were chosen to correspond to four different effect sizes ([36], p. 412-414) of small (2% of the variance), medium (13% of the variance), large (26% of the variance), and very large (40% of the variance). These parameter values correspond to .14, .39, .59, and

1, respectively in a population mediation model with continuous variables. For each of these 64 combinations of parameter values and all six sample sizes used in the simulation, there were 500 replications. In addition, for every combination of parameter values we computed one replication for a sample size of 20,000, using this as the true expected value for infinite sample size.

The accuracy of the point estimates were compared to the parameter values generating the latent  $Y^*$  data. Also, by comparing the means across different estimators for the largest sample size (20,000), we can examine the variation in population averages of the different parameterizations. Analysis of Variance was used to estimate effects of parameter value and sample size in order to summarize results. To conserve space, details of the analysis of variance are not presented.

### 3.1 Graphical Comparison of the difference in coefficient and product of coefficient methods

We begin with a graphical comparison of the product of coefficients with both standardized and unstandardized difference in coefficient methods, all derived under a correctly specified model. For logistic regression, a comparison of the unstandardized  $\hat{\beta}_z - \hat{\gamma}_z$ , the standardized  $\hat{\beta}_z^{logit.standard} - \hat{\gamma}_z$ , and the product of coefficients  $\hat{\gamma}_m \hat{\alpha}_z$  estimates of mediation effects is shown in Figure 1 for the case where  $M|Z$  is normally distributed. In this figure, the large sample expected values, obtained through 500 simulations of the three estimators of the mediated effect with  $n = 20,000$ , are plotted as a function of increasing values of the  $\gamma_m$  coefficient, averaging across values for the relation between the independent variable and the mediator ( $\alpha_z$ ). As the relation between the mediator and dependent variable gets larger, the size of the expected mediated effect should increase linearly because this relation is one of the components of the mediated effect. As shown in Figure 1, with increasing values of the  $\gamma_m$  parameter, all three estimators increase at different rates. While the



mean of the estimated product of coefficients measure,  $\gamma_m \alpha_z$ , has the highest rate, the difference in coefficients measure,  $\beta_z - \gamma_z$ , tends to flatten out at larger values of  $\gamma_m$ . Also shown in this figure, the standardized measure for  $\hat{\beta}_z^{logit.standard} - \hat{\gamma}_z$  is very close to the  $\hat{\gamma}_m \hat{\alpha}_z$  values. This result demonstrates the distortions possible when using the unstandardized difference in coefficients method. A similar pattern in the three estimates exists for probit regression and for other combinations of logit and probit error and logistic and probit regression.

Insert Figure 1 here

Figure 2 and 3 demonstrate problems with the  $\beta_z - \gamma_z$  method of estimating mediation. Figure 2 shows the value of  $\beta_z - \gamma_z$  as function of  $\gamma_z$  for the case where the true model holds  $\alpha_z$  at zero and allows  $\gamma_z$  to vary. The true mediated effect is zero for all conditions in the plot and the  $\alpha_z \gamma_m$  estimator of the mediated effect correctly has a value of zero for all values while the  $\beta_z - \gamma_z$  estimate departs from zero and becomes more negative as  $\gamma_z$  increases. Figure 3 shows similar information for the two estimators with  $\alpha_z$  held at .14 and  $\gamma_m$  varied so that the mediated effect should increase as  $\gamma_z$  is increased. The value of  $\beta_z - \gamma_z$  varies as function of the value of  $\gamma_m$  and for  $\gamma_z$  equal to 1,  $\beta_z - \gamma_z$  increases slightly and then declines even though the true mediated effect is increasing.

Insert Figures 2 and 3 here

### 3.2 Numerical Comparison of Mediation Measures under the Correct and Incorrect Models for the Dependent Variable

The first two tables are based on logistic and probit models with each fit using the appropriate model. The third table fits a logistic model to data generated from a probit model. For tables 1 to 6 below, we provide three rows for comparison. The first two rows compare the replication averages

to population quantities whereas the last row examines variability due to replications. The first row, labeled Mean ( $\mu_n$ ), provides a mean value of the estimates for each sample size ( $n$ ), averaged over the 64 combinations of parameter values. Since all of the parameter values chosen are positive, a mean value that is smaller than those used to generate the data, labeled in the table as  $\mu$ , indicates attenuation. Changes across this row reflect sample size bias. The second row, labeled Standard Deviation ( $\sigma_n$ ), measures the average standard deviation of the replication-averaged means over the 64 different parameter value simulations at that sample size. These values can be compared to the true standard deviation across the 64 sets of parameter values ( $\sigma$ ). The last row, labeled Average Squared Distance, measures the average squared distance of the estimate from the expected value of the estimate. The expected value of the estimate was obtained through 500 simulations with  $n = 20,000$ .

Insert Table 1 here

Table 1 compares the three estimators of mediated effect when the underlying data are both generated and modeled with a logistic distribution. The distribution of  $M|Z$  is generated from the normal and modeled as normal. For the product of coefficients method which is listed first, the first row shows there is very little overall bias for large samples (0.281) and very modest bias for small samples (0.314), compared to the true population mean (0.281). In terms of variability of these estimates across the sets of parameters, the second row shows substantial variation at sample size of 50 ( $\sigma_n = 0.439$  relative to  $\sigma = 0.258$ ), but minimal difference in these estimates with sample sizes of 500 or above. The last row indicates that the replication variances for the product of coefficients method decrease smoothly to zero.

The unadjusted difference in coefficients method exhibits several unfavorable behaviors. The first row shows it produces systematic attenuation, with  $\mu_{5000} = 0.213$  compared to  $\mu = 0.281$ .

Also, these means are nearly constant over the different sample sizes. Variability across these 64 replications in the average difference in coefficients estimates shows a similar pattern of decrease with sample size compared to the product of coefficients method.

Standardizing the difference in coefficients under a logistic model does substantially reduce the attenuation, as indicated by the first row in the third part of this table. Also,  $\sigma_n$  is quite appropriate given the 8% attenuation that still occurs for a sample size of 5,000.

Insert Table 2 here

Table 2 shows the same comparisons of mediated effect when the underlying data are both generated and modeled with a probit distribution. Again, the distribution of  $M|Z$  is taken from the normal and modeled as such. As shown in the three parts in Table 2, the  $\gamma_m\alpha_z$  and  $\beta_z - \gamma_z$  estimators are not equal but standardization makes  $\beta_z - \gamma_z$  very close to  $\gamma_m\alpha_z$ .

The product of coefficients method behaves just as well under a probit model as it does under a logit model. There is very little bias (row 1), and variability in these estimates behaves quite well for large and small sample sizes (row 2).

Examination of the unadjusted difference in coefficients method in Table 1, leads to a very different conclusion. First, there is large attenuation of this method ( $\mu_{5000} = 0.160$  compared to  $\mu = 0.281$ ), even more than under a logistic model. Also, compared to the product of coefficients, there is larger variance as a function of sample size ( $\sigma_{50}/\sigma_{5000} = 1.76$ ) for this difference in coefficients method. However, standardization succeeds very well, making each estimate of  $\beta_z^{logit.standard} - \gamma_z$  nearly identical to those of  $\gamma_m\alpha_z$ .

Insert Table 3 here

The sensitivity of these estimates to the assumption of a logistic distribution is investigated in this section. Table 3 shows the behavior of the different estimators for the mediated effect when the

underlying data are generated with a probit distribution but analyzed with logistic regression. We note that these two distributional assumptions are difficult to distinguish in practice unless large amounts of data are available. Thus methods that behave differently under these logistic and probit assumptions are not robust against this fairly innocuous change in models. Again, the distribution of  $M|Z$  is taken from the normal and modeled as such.

Unlike the two previous tables where definitions of mediation could be computed directly from the parameters themselves, this table's true mediated effects had to be defined numerically based on finding a best-fitting linear logistic model applied to data generated by a linear probit model. We found  $\mu = 0.475$  and  $\sigma = 0.442$ .

As shown in Table 3, the three methods produce quite different results, with the difference in coefficients having the smallest values, the product of coefficients method having the largest values, and the standardized method having a mean 13% smaller than the first. For large sample sizes, the product of coefficients methods agrees very closely with the mean over the 64 sets of parameter values ( $\mu_{5000} = 0.475$ ). We do, however, observe in row 1 of this table a relatively large bias of order  $\sqrt{n-1}$ . This overestimation in small samples under an incorrect logit assumption is somewhat larger than the overestimation when the true probit model is used. In contrast, the unadjusted difference in coefficients method shows severe attenuation at large samples, even more than it did under a correctly modeled probit. The unadjusted difference in coefficients method produces a modest negative  $\sqrt{n-1}$  bias making attenuation larger in smaller sample sizes.

### 3.3 Simulation for the Proportion Mediated Measures

We next compared five different estimators of the proportion mediated, using the product of coefficients and unstandardized and standardized estimates of difference in coefficients. Note that the

first estimator is based on coefficients that need no standardization; because the second and third estimators use  $\beta_z$  in the denominator, we anticipate that they will contain bias. The fourth and fifth estimators use the standardized  $\beta_z$  in the denominators and because of this we expect a priori that they will do nearly as well as the first. For Table 4 the dependent variable is generated from a logit and modeled using logistic regression, with the mediator having a normal distribution and modeled as such. The true mean of the 64 simulations for the proportion mediated is 0.343 and  $\sigma = 0.227$ . Both of these agree well with the  $\mu_{5000}$  and  $\sigma_{5000}$  for the first estimator,  $\hat{\alpha}_z \hat{\gamma}_m / (\hat{\alpha}_z \hat{\gamma}_m + \hat{\gamma}_z)$ . However, this estimator performs extremely poorly for samples of 500 or smaller since at these sample sizes the variability in the estimates is extreme. This is caused by the denominator occasionally being near zero. For the second and third estimators, which are both based on unstandardized  $\beta_z$ , both are severely biased in large samples and are therefore not recommended. The last two estimators, which are based on denominators with standardized  $\beta_z$  both have modest bias in large samples ( $\mu_{5000} = 0.360$  and  $0.335$  respectively versus a true value of 0.343) and agree well across all 64 simulations, as indicated by the fact that  $\sigma_{5000}$  is only modestly larger than  $\sigma = 0.227$ . Somewhat surprisingly, both of these estimators perform better than the first in small samples, although there is still substantial variability across replications.

Insert Table 4 here

The results for probit regression under a true probit model mirror these results for a correct logistic regression and therefore are not presented in tabular form but only described. Again, the large sample fits using  $\hat{\alpha}_z \hat{\gamma}_m / (\hat{\alpha}_z \hat{\gamma}_m + \hat{\gamma}_z)$  are the best, followed by the last two estimators that use standardized  $\beta_z$ . The two nonstandardized models have very large bias in large samples. Also as in the previous case with logistic regression modeling a true logit model, there is extremely high instability of the first estimator for sample sizes of 500 or below. We conclude that the measure of

the proportion mediated is inherently difficult to estimate in small to moderate sized samples.

We now turn to estimation of the proportion mediated when the data are generated under a probit but analyzed incorrectly with logistic regression. As the results are similar to the two correctly analyzed logistic and probit regressions just described, no table is shown. Again, large sample bias for  $\hat{\alpha}_z \hat{\gamma}_m / (\hat{\alpha}_z \hat{\gamma}_m + \hat{\gamma}_z)$  is negligible, but this estimator is still subject to extreme variance in small to moderate samples. The last two estimators with standardized  $\beta_z$  in the denominators perform much better in small samples and show only a small amount of bias with a sample size of 5,000. However, both of these estimators that rely on unstandardized  $\beta_z$  show substantial bias in large samples and cannot be recommended.

### 3.4 Simulation results when both conditional distributions are incorrectly specified

To investigate how sensitive results are to the underlying distributional assumptions, we investigated the behavior of the three estimators of mediation effect in two situations where the conditional distributions of  $Y$  and  $M$  were both misspecified. Again, for both of these simulations we assumed a true probit model but used logistic regression to model how  $Y$  depends on  $M$  and  $Z$ . In the first simulation described below, we assumed that  $M|Z$  was normal, but the true error distribution for the mediator was given a t-distribution with 3 degrees of freedom. That is,  $M = \alpha_z(Z - 3/2) + T_3/\sqrt{3}$ ,  $z = 1, 2$  with the scale factor chosen so that the variance of the error term is one. This symmetric error distribution with large but finite variance is often used in studies of robustness to a small proportion of large errors. The second simulation was generated to resemble the data in our example described below. These data had substantial skewness ( $\kappa_1 \sim 2$ ). In our simulation, the mediator was generated as  $M = \alpha_z(Z - 3/2) + (\chi_2^2 - E(\chi_2^2))/2$  so that the skewness in the data

were exactly 2. The errors were weighted so that the standardized regression coefficients for the mediator were again small (0.14), medium (0.39), and large (0.59) to match the earlier simulations.

In Table 5, where the t-distribution is incorrectly modeled as normal, the product of coefficients method performs quite well. In fact, the simulation results for this estimator are nearly identical to the ones obtained in Table 3 where conditional normality for  $M$  was correctly assumed. In fact, the means and variances are almost identical for the product of coefficients method (0.475 versus 0.475 for  $\mu_{5000}$  and 0.443 versus 0.444 for  $\sigma_{5000}$ ). There is also very good agreement between the product of coefficients method and the standardized difference in coefficients method across all sample sizes, much better than that in Table 3. Both of these methods show positive bias at smaller sample sizes. Just as before, the unstandardized difference in coefficients method shows substantial bias in large samples and little change in bias as a function of sample size.

Insert Table 5 here

In Table 6, with an asymmetric error distribution, we anticipate some instability of these estimators because they are computed using two wrong distributional assumptions. Only the product of coefficients method provides very little overall bias. In particular,  $\mu_{5000} = 0.487$ , which agrees very well with value of  $\mu_{5000} = 0.475$  from Table 3 and 0.475 from Table 5, the two previous models with the same incorrectly specified logistic model. There is somewhat more variation in these estimates with a asymmetric error distribution than in the previous two cases ( $\sigma_{5000} = 0.579$  compared to 0.444 when normality is correctly assumed). These differences, however, are quite minimal compared to the instability of both the unstandardized and standardized difference in coefficients methods for the normality, long-tailed, and asymmetric error distributions. In particular,  $\mu_{5000}$  varies from 0.235 to 0.365 for the unstandardized difference in coefficients and from 0.410 to 0.582 even with standardization.

Insert Table 6 here

## 4 Case Study with Intentions as a Mediator for Cigarette Use

Mediation for a binary dependent variable is illustrated here with a data set from a school-based drug prevention program. The Midwestern Prevention Project (MPP) was a longitudinal school- and community-based drug prevention program. The intervention program was introduced to 6th and 7th grade students in Kansas City and consisted of school and parent programs, health policy changes and mass media coverage (for full details see Pentz et al.[37]). Schools were randomly assigned to treatment or control conditions. In this example we posit that behaviors are preceded by behavioral intentions, and test whether intention to use cigarettes is a mediator of the program effect on cigarette use. The data set consists of 864 students with complete data on three variables:  $Z$  treatment condition;  $M$  intention to use tobacco in the following two month period, measured approximately six months after baseline; and  $Y$  report of tobacco use at follow-up in the prior one month period. Table 7 contains the cross-tabulations of these three variables, as well as the coding schemes.

Insert Table 7 here

The mediated effect was estimated using Equation 11 and least squares regression of  $M$  on  $Z$ . When calculated as  $\hat{\alpha}_z \hat{\gamma}_m$ , the mediated effect was -0.171 ( $\hat{\sigma}_{\hat{\alpha}_z \hat{\gamma}_m}^2 = .0648$ , CI= -0.298, -0.044). When calculated as  $\hat{\beta}_z - \hat{\gamma}_z$  the mediated effect was -0.129 ( $\hat{\sigma}_{\hat{\beta}_z - \hat{\gamma}_z} = 0.089$ , CI= -0.303, 0.046), a discrepancy of 25% and a formal difference in whether a significant mediation was found. This discrepancy is also present in the various proportion mediated measures, with  $1 - \hat{\gamma}_z / \hat{\beta}_z = 0.254$ ,  $\hat{\alpha}_z \hat{\gamma}_m / \hat{\beta}_z = 0.338$ , and  $\hat{\alpha}_z \hat{\gamma}_m / \hat{\gamma}_z + \hat{\alpha}_z \hat{\gamma}_m = 0.312$ . Although the proportion mediated estimates in



this example are all greater than .20, the proportion mediated estimates using the  $\hat{\alpha}_z\hat{\gamma}_m$  estimate are 20 to 30% greater than the size of the estimate using  $\hat{\beta}_z$  and  $\hat{\gamma}_z$ .

Equation 12 was used to standardize the  $\beta_z$  logistic regression estimate. After standardization, the difference in coefficient mediated effect and the two proportion measures involving proportions were recalculated. The standardized estimate of the difference in coefficients estimate was -0.197, much closer to the product of coefficients estimate. The standardized proportion measures are  $1 - \hat{\gamma}_z/\hat{\beta}_z = 0.343$  and  $\hat{\alpha}_z\hat{\gamma}_m/\hat{\beta}_z = 0.297$ . These results are also consistent with the simulation study; standardization brings the estimates much closer together. In this example, the use of the unstandardized estimates would lead to the conclusion that the mediated effect was smaller than it truly was.

## 5 Discussion

The purpose of this paper was to examine two different approaches to estimating the intermediate endpoint or mediated effect when the dependent variable is binary and logistic or probit regression is used. Unlike the linear model situation, the difference in coefficients and product of coefficients estimators can lead to substantially different estimates and inferential conclusions. We recommend using the product of coefficients methods for the following reasons. It generally has less bias than the difference in coefficients method, it reflects the population mediated effect, and in our simulations it was quite robust against departures from the logistic or probit assumption as well as the normality assumption for the distribution of the mediator. This estimator behaved well under both a symmetric error distribution with heavy tails as well as an asymmetric error distribution. Another advantage of this method is that it is the easiest to compute.

We do not recommend the use of the difference in coefficients method. The unstandardized

procedure can indicate no change in the mediated effect when it is actually increasing. This effect is due to the difference in scales of the logistic regression coefficients of  $Y$  on  $Z$  adjusting or not adjusting for  $M$  as these coefficients come from separate logistic regression equations. There may be situations where the mediated effect is defined in terms of a difference in coefficients so the method may be appropriate. The same problem also affects the proportion mediated measures based on this difference in coefficients method. Standardizing coefficients prior to calculating these quantities does improve the performance of the difference in coefficients method.

None of the five estimators for the proportion mediated effect were fully satisfactory. The method based on the product of coefficients performed well in samples over 500 both in terms of bias and robustness to the probit assumption, but this estimator showed a high degree of variability in small samples. The two estimators based on the unstandardized difference in coefficients method showed marked bias even in large samples. By standardizing the denominator for these two estimators, there was a marked decrease in bias compared to the unstandardized version, and the results were somewhat more stable across different sample sizes. We recommend not using any of these estimators for samples below a few hundred and in large samples recommend estimating the proportion mediated based on the product of coefficients method.

Although the results discussed in this article are for the binary  $X$  case, the results were very similar when the study was conducted with a continuous  $X$ . The only important difference in the results was related to the  $\hat{\alpha}_z \hat{\gamma}_m / \hat{\gamma}_z + \hat{\alpha}_z \hat{\gamma}_m$  proportion mediated estimator, which performs much better when  $Z$  is continuous.

A limitation of the current study is that it does not discuss multiple mediator models. More complicated models would have the same issues with mediation estimation as described in this article. Similarly, it was assumed that the form of the true mediation model was known and no

moderation existed. With actual data, the model is not known and issues related to the true relations among  $Z$ ,  $M$ , and  $Y$  may be more difficult to disentangle.

## References

- [1] Susser M. Causal Thinking in the Health Sciences: Concepts and Strategies of Epidemiology. New York: Oxford University Press, 1973.
- [2] Susser M. What is a cause and how do we know one? A grammar for pragmatic epidemiology. *Am J Epidemiol* 1991; **133**: 635-648.
- [3] Statistical and mathematical modeling of the AIDS epidemic. Papers from a conference. Baltimore, Maryland, 17-18 November, 1987. *Stat Med* 1989; **8**: 1-139.
- [4] Henney JE. Remarks by: Jane E. Henney, M.D. Commissioner of Food and Drugs. International Conference on Surrogate Endpoints and Biomarkers - National Institutes of Health, 1999. <http://www.fda.gov/oc/speeches/surrogates8.html> [Accessed 11/21/01].
- [5] Criqui MH, Cowan LD, Tyroler HA, et al. Lipoproteins as mediators for the effects of alcohol consumption and cigarette smoking on cardiovascular mortality: Results from the lipid research clinics follow-up study. *Am J Epidemiol* 1987; **126**: 629-637.
- [6] Lin DY, Fleming TR, De Gruttola V. Estimating the proportion of treatment effect explained by a surrogate marker. *Stat Med* 1997; **16**: 1515-1527.
- [7] Freedman LS, Schatzkin A. Sample size for studying intermediate endpoints within intervention trials or observational studies. *Am J Epidemiol* 1992; **136**: 1148-1159.
- [8] Begg CB, Leung DHY. On the use of surrogate end points in randomized trials. *J R Stat Soc A* 2000; **163**: 15-28.
- [9] Prentice RL. Surrogate endpoints in clinical trials: Definition and operational criteria. *Stat Med* 1989; **8**: 431-440.

- [10] Freedman LS, Graubard BI, Schatzkin A. Statistical validation of intermediate endpoints for chronic diseases. *Stat Med* 1992; **11**: 167-178.
- [11] Fleming TR, DeMets DL. Surrogate end points in clinical trials: Are we being misled? *Ann Intern Med* 1996; **125**: 605-613.
- [12] Molenberghs G, Buyse M, Geys H, Renard D, Burzykowski T, Alonso, A. Statistical challenges in the evaluation of surrogate endpoints in randomized trials. *Controlled Clinical Trials* 2002; **23**: 607-625.
- [13] Verbeke G, Molenberghs G. Linear Mixed Models for Longitudinal Data. New York: Springer, 2000.
- [14] Freedman LS. Confidence intervals and statistical power of the "Validation" ratio for surrogate or intermediate endpoints. *J Stat Plan Infer* 2001; **96**: 143-153.
- [15] MacKinnon DP, Warsi G, Dwyer JH. A simulation study of mediated effect measures. *Multivar Behav Res* 1995; **30**: 41-62.
- [16] Buyse M, Molenberghs G. Criteria for the validation of surrogate endpoints in randomized experiments. *Biometrics* 1998; **54**: 1014-1029.
- [17] Buyse M, Molenberghs G, Burzykowski T, Renard D, Geys H. The validation of surrogate endpoints in meta-analyses of randomized experiments. *Biostatistics* 2000; **1**: 49-67.
- [18] Burzykowski T, Molenberghs G, Buyse M. The Evaluation of Surrogate Endpoints. New York: Springer, 2005.
- [19] Gail MH, Pfeiffer R, van Houwelingen HC, Carroll RJ On meta-analytic assessment of surrogate outcomes. *Biostatistics* 2000; **1**, 231-246.

- [20] Frangakis CE, Rubin, DB. Principal stratification in causal inference. *Biometrics* 2002, **58**, 21-29.
- [21] Foshee VA, Bauman KE, Arriaga XB, et al. An evaluation of Safe Dates, an adolescent dating violence prevention program. *Am J Public Health* 1998; **88**: 45-50.
- [22] Botvin GJ. Preventing drug abuse in schools: Social and competence enhancement approaches targeting individual-level etiologic factors. *Addict Behav* 2000; **25**: 887-897.
- [23] Judd CM, Kenny DA. Process analysis: Estimating mediation in treatment evaluations. *Evaluation Rev* 1981; **5**: 602-619.
- [24] MacKinnon DP, Dwyer JH. Estimating mediated effects in prevention studies. *Evaluation Rev* 1993; **17**: 144-158.
- [25] Kenny DA, Kashy DA, Bolger N. Data analysis in social psychology. In: Gilbert DT, Fiske ST, Lindzey G, editors. *The Handbook of Social Psychology*. Boston: McGraw-Hill, 1998; 233-265.
- [26] MacKinnon DP, Lockwood CM, Hoffman JM, West SG, Sheets V. A comparison of methods to test mediation and other intervening variable effects. *Psychol Methods* 2002; **7**: 83-104.
- [27] Sobel ME. Asymptotic confidence intervals for indirect effects in structural equation models. *Sociological Methodology* 1982; **13**: 290-312.
- [28] Krull JL, MacKinnon DP. Multilevel mediation modeling in group-based intervention studies. *Evaluation Rev* 1999; **23**: 418-444.
- [29] Baron RM, Kenny DA. The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *J Pers Soc Psych* 1986; **51**: 1173-1182.
- [30] Bollen KA, *Structural Equation with Latent Variables*. Wiley. NY: New York; 1989.

- [31] Hosmer DW, Lemeshow, S. Applied Logistic Regression. New York: Wiley, 2000.
- [32] Finney DJ. Probit Analysis: Statistical Treatment of the Sigmoid Response Curve. Cambridge, England: Cambridge University Press; 1964, ©1956
- [33] Zeger SL, Liang KY, Albert PS. Models for longitudinal data: a generalized estimating equation approach. *Biometrics*, 1989; **45**: 1049-60.
- [34] Draper MP, Smith H. Applied Regression Analysis. 3rd ed. New York: Wiley, 1998.
- [35] McCullagh P, Nelder JA. Generalized Linear Models. London: Chapman Hall, 1989.
- [36] Cohen J. Statistical Power Analysis for the Behavioral Sciences. Hillsdale, NJ: Erlbaum; 1988.
- [37] Pentz MA, Dwyer JH, MacKinnon DP, et al. A multicomunity trial for primary prevention of adolescent drug abuse: Effects on drug use prevalence. *JAMA* 1989; **261**: 3259-3266.

Figure 1: Average values of mediated effect estimates: Logistic regression

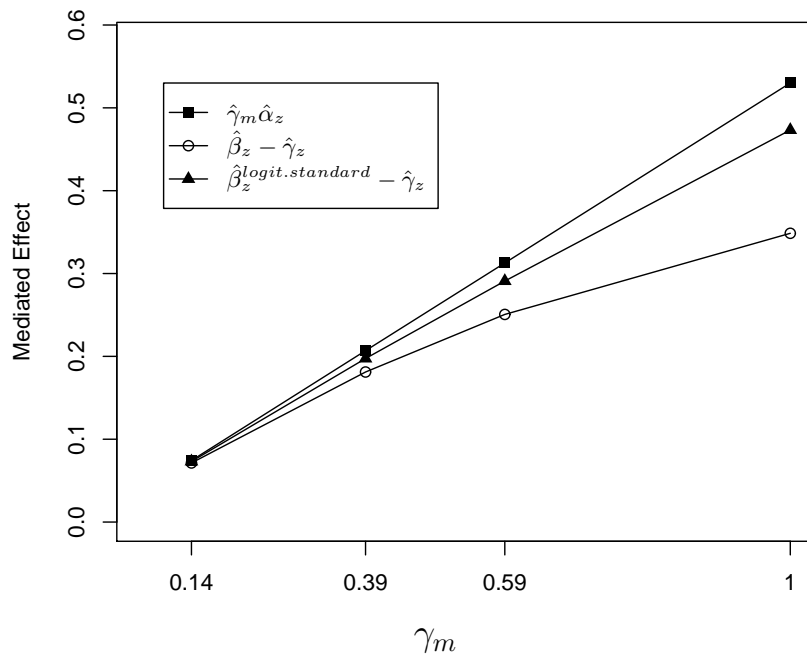




Figure 2: Estimated mediated effects as function of  $\gamma_z$  when  $\alpha_z$  is held at 0: Logistic regression

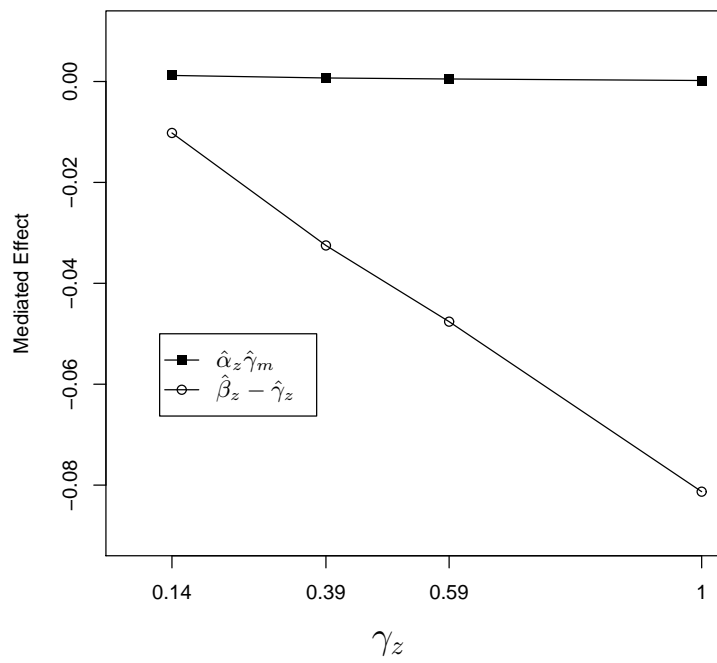


Figure 3: Estimated mediated effect  $\hat{\beta}_z - \hat{\gamma}_z$  as function of  $\gamma_m$  and  $\gamma_z$  with  $\alpha_z = .14$ : Logistic regression

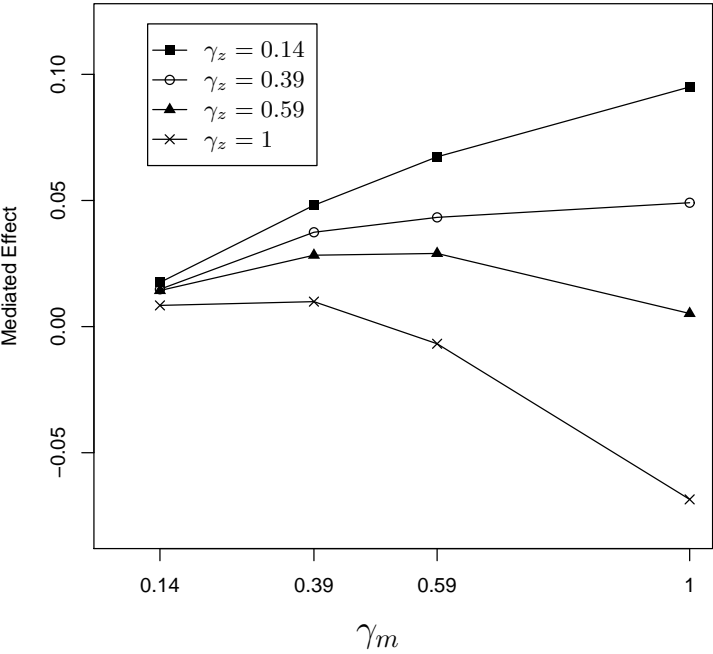


Table 1: Comparison of the product of coefficients and difference in coefficients estimators of the mediated effect under a correct logistic and correct normality assumption

		Sample Size					
		50	100	200	500	1000	5000
$\hat{\gamma}_m \hat{\alpha}_z$	$\mu = 0.281, \sigma = 0.258$						
	Mean ( $\mu_n$ )	0.314	0.295	0.289	0.285	0.283	0.281
	Standard Deviation ( $\sigma_n$ )	0.439	0.342	0.298	0.272	0.265	0.258
	Average Squared Distance	0.111	0.045	0.020	0.007	0.004	0.001
$\hat{\beta}_z - \hat{\gamma}_z$							
	Mean ( $\mu_n$ )	0.214	0.213	0.214	0.214	0.213	0.213
	Standard Deviation ( $\sigma_n$ )	0.356	0.270	0.234	0.213	0.206	0.201
	Average Squared Distance	0.080	0.031	0.014	0.005	0.003	0.001
$\hat{\beta}_z^{logit.standard} - \hat{\gamma}_z$							
	Mean ( $\mu_n$ )	0.289	0.271	0.266	0.262	0.260	0.259
	Standard Deviation ( $\sigma_n$ )	0.420	0.320	0.277	0.253	0.245	0.239
	Average Squared Distance	0.105	0.040	0.018	0.007	0.003	0.001

Table 2: Comparison of the product of coefficients and difference in coefficients estimators of the mediated effect under a correct probit and correct normality assumption

		Sample Size					
		50	100	200	500	1000	5000
$\hat{\gamma}_m \hat{\alpha}_z$ $\mu = 0.281, \sigma = 0.258$							
	Mean ( $\mu_n$ )	0.321	0.295	0.289	0.285	0.283	0.281
	Standard Deviation ( $\sigma_n$ )	0.428	0.323	0.288	0.269	0.263	0.257
	Average Squared Distance	0.096	0.031	0.014	0.005	0.003	0.001
$\hat{\beta}_z - \hat{\gamma}_z$							
	Mean ( $\mu_n$ )	0.149	0.156	0.159	0.161	0.160	0.160
	Standard Deviation ( $\sigma_n$ )	0.301	0.222	0.194	0.180	0.174	0.170
	Average Squared Distance	0.057	0.019	0.008	0.003	0.002	0.000
$\hat{\beta}_z^{probit.standard} - \hat{\gamma}_z$							
	Mean ( $\mu_n$ )	0.326	0.297	0.290	0.285	0.283	0.281
	Standard Deviation ( $\sigma_n$ )	0.445	0.326	0.290	0.269	0.263	0.257
	Average Squared Distance	0.108	0.033	0.015	0.005	0.003	0.001

Table 3: Comparison of the product of coefficients and difference in coefficients estimators of the mediated effect under an incorrect logistic and correct normality assumption

		Sample Size					
		50	100	200	500	1000	5000
$\hat{\gamma}_m \hat{\alpha}_z$	$\mu = .475, \sigma = .442$						
	Mean ( $\mu_n$ )	0.548	0.501	0.490	0.482	0.478	0.475
	Standard Deviation ( $\sigma_n$ )	0.741	0.560	0.499	0.464	0.454	0.444
	Average Squared Distance	0.288	0.095	0.042	0.016	0.008	0.002
$\hat{\beta}_z - \hat{\gamma}_z$							
	Mean ( $\mu_n$ )	0.218	0.229	0.234	0.236	0.235	0.235
	Standard Deviation ( $\sigma_n$ )	0.512	0.369	0.320	0.294	0.285	0.278
	Average Squared Distance	0.168	0.055	0.024	0.009	0.004	0.001
$\hat{\beta}_z^{logit.standard} - \hat{\gamma}_z$							
	Mean ( $\mu_n$ )	0.483	0.435	0.424	0.416	0.413	0.410
	Standard Deviation ( $\sigma_n$ )	0.707	0.508	0.448	0.414	0.403	0.394
	Average Squared Distance	0.284	0.082	0.037	0.013	0.007	0.001

Table 4: Comparison of the product of coefficients and difference in coefficients estimators for the proportion mediated estimators under a correct logistic assumption and correct normal assumption

	Sample Size					
	50	100	200	500	1000	5000
$\hat{\gamma}_m \hat{\alpha}_z / (\hat{\gamma}_z + \hat{\gamma}_m \hat{\alpha}_z)$ $\mu = 0.343, \sigma = 0.227$						
Mean ( $\mu_n$ )	$-2.4 \times 10^{11}$	$3.8 \times 10^{11}$	$1.3 \times 10^{10}$	0.358	0.280	0.349
Standard Deviation ( $\sigma_n$ )	$1.7 \times 10^{13}$	$4.4 \times 10^{13}$	$2.3 \times 10^{12}$	13.508	13.144	0.247
Average Squared Distance	$3.0 \times 10^{26}$	$1.9 \times 10^{27}$	$5.2 \times 10^{24}$	182.396	172.716	0.009
$1 - \hat{\gamma}_z / \hat{\beta}_z$						
Mean ( $\mu_n$ )	0.243	0.311	0.412	0.340	0.320	0.304
Standard Deviation ( $\sigma_n$ )	3.862	6.659	8.494	4.905	1.142	0.251
Average Squared Distance	14.872	44.256	72.060	23.981	1.243	0.010
$\hat{\gamma}_m \hat{\alpha}_z / \hat{\beta}_z$						
Mean ( $\mu_n$ )	0.360	0.407	0.501	0.427	0.403	0.386
Standard Deviation ( $\sigma_n$ )	3.845	6.682	8.428	5.084	1.145	0.284
Average Squared Distance	14.716	44.532	70.926	25.743	1.231	0.011
$1 - \hat{\gamma}_z / \hat{\beta}_z^{logit.standard}$						
Mean ( $\mu_n$ )	0.288	0.335	0.435	0.369	0.350	0.335
Standard Deviation ( $\sigma_n$ )	3.414	6.107	7.776	4.584	1.116	0.248
Average Squared Distance	11.618	37.214	60.387	20.934	1.185	0.010
$\hat{\gamma}_m \hat{\alpha}_z / \hat{\beta}_z^{logit.standard}$						
Mean ( $\mu_n$ )	0.314	0.360	0.460	0.395	0.375	0.360
Standard Deviation ( $\sigma_n$ )	3.403	6.115	7.708	4.715	1.114	0.258
Average Squared Distance	11.535	37.314	59.324	22.151	1.175	0.010

Table 5: Comparison of the product of coefficients and difference in coefficients estimators of the mediated effect under an incorrect logistic and incorrect normality assumption (probit and  $t(3)$ )

		Sample Size					
		50	100	200	500	1000	5000
$\hat{\gamma}_m \hat{\alpha}_z$ $\mu = .474, \sigma = .440$							
	Mean ( $\mu_n$ )	0.555	0.510	0.492	0.480	0.476	0.475
	Standard Deviation ( $\sigma_n$ )	0.762	0.579	0.506	0.466	0.451	0.443
	Average Squared Distance	0.330	0.116	0.049	0.018	0.009	0.002
$\hat{\beta}_z - \hat{\gamma}_z$							
	Mean ( $\mu_n$ )	0.317	0.314	0.310	0.307	0.306	0.306
	Standard Deviation ( $\sigma_n$ )	0.532	0.398	0.352	0.326	0.317	0.311
	Average Squared Distance	0.168	0.051	0.022	0.008	0.004	0.001
$\hat{\beta}_z^{logit.standard} - \hat{\gamma}_z$							
	Mean ( $\mu_n$ )	0.597	0.538	0.514	0.500	0.496	0.495
	Standard Deviation ( $\sigma_n$ )	0.828	0.618	0.541	0.497	0.483	0.475
	Average Squared Distance	0.386	0.124	0.052	0.019	0.010	0.002

Table 6: Comparison of the product of coefficients and difference in coefficients estimators of the mediated effect under an incorrect logistic and incorrect normality assumption (probit and  $\chi^2(2)$ )

		Sample Size					
		50	100	200	500	1000	5000
$\hat{\gamma}_m \hat{\alpha}_z$ $\mu = .484, \sigma = .447$							
	Mean ( $\mu_n$ )	0.588	0.528	0.508	0.492	0.487	0.484
	Standard Deviation ( $\sigma_n$ )	0.928	0.610	0.529	0.475	0.459	0.449
	Average Squared Distance	0.582	0.139	0.059	0.021	0.010	0.002
$\hat{\beta}_z - \hat{\gamma}_z$							
	Mean ( $\mu_n$ )	0.418	0.386	0.377	0.368	0.365	0.364
	Standard Deviation ( $\sigma_n$ )	0.630	0.460	0.411	0.377	0.368	0.361
	Average Squared Distance	0.215	0.062	0.027	0.010	0.005	0.001
$\hat{\beta}_z^{logit.standard} - \hat{\gamma}_z$							
	Mean ( $\mu_n$ )	0.891	0.674	0.624	0.596	0.587	0.582
	Standard Deviation ( $\sigma_n$ )	2.110	0.885	0.692	0.613	0.590	0.577
	Average Squared Distance	3.780	0.337	0.097	0.031	0.014	0.003



Table 7: Example data set from Midwestern Prevention Project (N=864)

Mediator	Treatment Group (1)		Control Group (0)	
Intention to use cigarettes	Cigarette Use (1)	No Use (0)	Cigarette Use (1)	No Use (0)
Yes (4)	19	9	20	9
Probably (3)	11	15	20	14
Don't think so (2)	11	43	13	36
No (1)	32	353	30	229

# Appendix A: Mediated effects based on the product of coefficients method

This section shows the consistency of the product of coefficients method for logistic regression when  $Y|XZ$  is given by Model 1 and the conditional mean of  $M|XZ$  is linear as in Equations 2 and 3. A formula for the variance of this product of coefficients estimator is then given. Define  $\hat{\gamma}$  as the maximum likelihood solution for the logistic regression coefficients  $\gamma = (\gamma_0, \gamma, \gamma_m, \gamma_z)$  in Equation 1, the standard logistic regression model of  $Y$  predicted by  $M$  as well as  $Z$  and  $\mathbf{X}$ . As the sample size gets large,  $\hat{\gamma}$  are asymptotically unbiased estimates for  $\gamma$ . Similarly for the regression of  $M$  on  $Z$  adjusted for  $\mathbf{X}$  in Equation 2, the least squares estimates  $\hat{\alpha}$  are asymptotically unbiased estimates for  $\alpha$  provided the conditional mean is correctly specified as in Equation 3. Thus the product of these estimates is consistent.

Further, the following sandwich-type variance estimates are consistent as well. First define the  $i^{th}$  observation's score function as  $\mathbf{S}_{\gamma_i} = \partial \log Pr\{Y = y_i | \mathbf{X} = \mathbf{x}_i, M = M_i, Z = z_i\} / \partial \gamma$  evaluated at  $\hat{\gamma}$ . The information matrix for  $\alpha$  is defined as  $\mathcal{I}_{\gamma} = -\sum_{i=1}^n \frac{\partial^2}{\partial \gamma \partial \gamma'} \log Pr\{Y = y_i | \mathbf{X} = \mathbf{x}_i, M = m_i, Z = z_i\}$ . Then

$$Avar(\hat{\gamma}) = \mathcal{I}_{\gamma}^{-1} \left( \sum_{i=1}^n \mathbf{S}_{\gamma_i} \mathbf{S}_{\gamma_i}' \right) \mathcal{I}_{\gamma}^{-1} \quad (14)$$

Turning to the regression of  $M$  on  $Z$  and  $\mathbf{X}$ , if  $f_{M|XZ; \alpha}(m|\mathbf{x}, z)$  is known, it would be used to maximize the corresponding likelihood. For example, if this distribution is normal, define  $\phi(u; \mu, \sigma^2)$  as the normal density. Then the  $i^{th}$  observation's score function, with parameters  $\theta = (\alpha_0, \alpha_x, \alpha_z, \sigma_{m.xz}^2)$ , is given by  $\mathbf{S}_{\theta_i} = \partial \log \phi(m_i; \alpha_0 + \alpha' \mathbf{x}_i + \alpha_z z_i, \sigma_{m.xz}^2) / \partial \theta$  evaluated at the least squares solution  $\hat{\theta}$ . Similarly, the information matrix for this problem is given by

$\mathcal{I}_\theta = -\sum_{i=1}^n \frac{\partial}{\partial \boldsymbol{\theta}'} \mathbf{S}_{\theta_i}$ . Then

$$\text{Avar}(\hat{\boldsymbol{\theta}}) = \mathcal{I}_\theta^{-1} \left( \sum_{i=1}^n \mathbf{S}_{\theta_i} \mathbf{S}'_{\theta_i} \right) \mathcal{I}_\theta^{-1} \quad (15)$$

and the diagonal element corresponding to  $\alpha_z$ ,  $\text{Avar}(\hat{\alpha}_z)$  is therefore a consistent estimate of the variance of  $\hat{\alpha}_z$  provided the mean of  $M$  is linear in  $\mathbf{X}$  and  $Z$ .

If the underlying distribution for  $Y|XZ$  is not known, then consistent estimates of  $\boldsymbol{\alpha}$  can still be obtained as long as the conditional model for the mean is correctly specified. In this case, the least squares approach described above provides consistent estimates of both the regression parameters and their standard errors [34].

Finally, we show that these GEE estimators of  $\hat{\boldsymbol{\gamma}}$  and  $\hat{\alpha}_z$  are asymptotically independent. Using a Taylor expansion on the score function for  $\boldsymbol{\gamma}$ ,

$$\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma} = \mathcal{I}_\gamma^{-1} \left( \sum_{i=1}^n \mathbf{S}_{\gamma_i} \right) + O_p(n^{-1}) \quad (16)$$

where  $O_p(n^{-1})$  is a random term that converges to a distribution with order  $n^{-1}$ . Explicitly, the  $i^{\text{th}}$  observation's contribution to the logistic score function,  $\mathbf{S}_{\gamma_i}$  is

$$\mathbf{S}_{\gamma_i} = \frac{\partial}{\partial \boldsymbol{\gamma}} \log \Pr\{Y = y_i | \mathbf{X} = \mathbf{x}_i, M = M_i, Z = z_i\} = (y_i - p_i)(1, \mathbf{x}_i, M_i, z_i)' \quad (17)$$

where  $p_i = \Pr\{Y = 1 | \mathbf{X} = \mathbf{x}_i, M = M_i, Z = z_i\} = E(Y = 1 | \mathbf{X} = \mathbf{x}_i, M = M_i, Z = z_i)$ .

Also,

$$\mathcal{I}_\gamma = -\sum_{i=1}^n \frac{\partial^2 \log p_i}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}'} = \sum_{i=1}^n p_i(1 - p_i)(1, \mathbf{x}'_i, M_i, z_i)'(1, \mathbf{x}'_i, M_i, z_i) \quad (18)$$

Note that like  $\mathcal{I}_\gamma$ ,  $\hat{\alpha}_z$  depends only on  $\mathbf{T} = \{(\mathbf{x}_i, m_i, z_i), i = 1, \dots, n\}$  and not on  $y$ . Then by

conditioning first on  $\mathbf{T}$ , we have

$$\begin{aligned} \text{Cov}(\hat{\gamma}, \hat{\alpha}_z) &= E(\hat{\gamma} - \gamma)\hat{\alpha}_z \\ &= E^M(E^Y((\hat{\gamma} - \gamma)\hat{\alpha}_z|\mathbf{T})) \\ &= E(\mathcal{I}_\gamma^{-1}(\sum_{i=1}^n E(y_i - p_i|\mathbf{T})(1, \mathbf{x}_i, M_i, z_i)'))\hat{\alpha}_z \\ &= \mathbf{0} \end{aligned}$$

This result can be easily extended to any GEE type of estimator of  $\gamma$ , of which the maximum likelihood estimator from an exponential family distribution is a special case.