South Asian Journal of Management Sciences
Vol. 19(1): 37 – 55, 2025
DOI: 10.21621/sajms.2025191.02

Multiple Mediation Analysis and Binary Mediators: A Methodological Brief

Dr. Faiza Hassan[1]* Dr. Hafsa Hina[2¥] Dr. Arshad Ali[3†]

Abstract: *The paper briefly reviews mediation analysis and provides a detailed review of newly available estimation techniques dealing with categorical variables. The focus of the paper is on dealing with binary mediators. It discusses some recent methods such as marginal mediation analysis in detail highlighting the theoretical underpinnings of these techniques. It also describes the strengths, weaknesses and step by step illustration of each method. The paper is helpful in understanding the mediation analysis, the issues involved in estimation while dealing with binary mediators. The paper is pedagogical in nature and give quick and clear understanding of some recent mediation techniques.*

Keywords: *Mediation Analysis; Binary Mediators; Marginal Mediation; Methodological brief*

Introduction

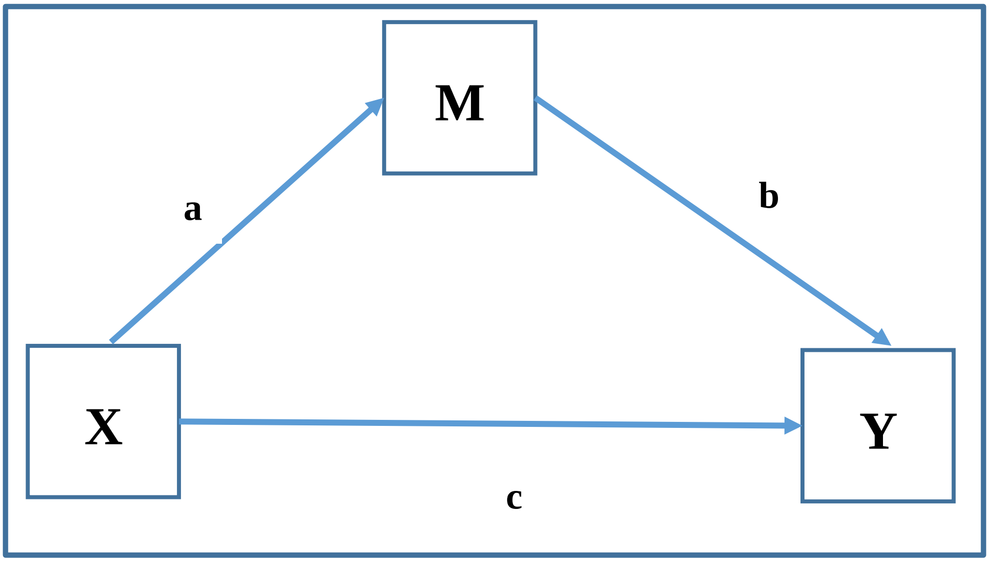
Mediation Analysis is a widely used statistical technique in variety of academic disciplines including psychology, management, economics and health sciences. It is a powerful tool to analyze the complex relationships and to quantify the indirect effects. However, inclusion of binary mediators in mediation models poses significant challenges both for estimation and interpretations. This paper aims to discuss and to provide comparative analysis of some recent techniques available to deal with binary mediators. Specifically, this methodological review synthesizes and compares various approaches, offering researchers a practical guide for selecting the most appropriate method for their specific research context. Each method is discussed in detail for the underlying principles, methods, strengths and weaknesses. This paper begins with a brief review of mediation analysis, followed by a discussion of traditional estimation methods. Subsequently, recent techniques are discussed individually, along with their methodologies in detail. Towards the end of the paper, a summary table is provided to offer a concise overview of the discussed methods.

[1] University of Malakand, Khyber Pakhtunkhwa, Pakistan. Email: faizahassan@uom.edu.pk
[2] Institute of Development Economics, Islamabad, Pakistan. Email: hafsahina@pide.org.pk
[3] University of Malakand, Khyber Pakhtunkhwa, Pakistan. Email: arshaduom@gmail.com

Mediation Analysis

Mediation analysis investigates the relationship between the exogenous and dependent variable keeping in consideration the third variable i.e. the mediator. Hence, a mediator is an intervening variable that is expected to affect the dependent variable and get affected by changes in the independent variable. M variable is said to be a Mediator in X and Y relationship if the impact of X on Y variable get affected by its introduction in the model as well as if it is established that X significantly affects M and M considerably impact Y variable in the presence of X variable (Judd & Kenny, 1981; Barron & Kenny,1986). The following figure shows a simple mediation model.

Figure 1



Simple Mediation Model

Introduction of intervening variable splits the effect of X on Y into two parts the direct and indirect effect. The indirect effect is the effect of X variable which is transferred into Y variable through the intermediating variable M (X → M → Y). As changes in X variable brings changes into variable M and when M changes it impact Y variable. So M provides a path through which X variable transfer its impact to variable Y.

Mediation model is called a simple mediation model when there is a single mediator in the model. While it is called a multiple mediation model when there are more than one intermediating factors in the model. Full mediation is when the introduction of mediating factor or factors makes the direct path insignificant. It means that in full mediation model all of the effect of X to Y is through mediator M. Partial mediation is when the mediating factor(s) reduces the direct impact of X on Y but the direct path remains non-zero and significant.

Preacher and Hayes, (2008) recommended using a multiple mediation model when there are more than one mediating factor between X and Y variable instead of using a simple mediation model for each intermediate factor separately. It helps in reducing omitted variable bias, comparison of different indirect effects becomes possible, testing the significance of total indirect effect helps in concluding whether the mediators included in the model are jointly mediating the effect of X on Y or not.

Methods of Estimation

This section will illustrate different techniques available for estimating and testing mediation along with their pros and cons. The focus will be on strategies available for performing multiple mediation analysis. The following general model and coefficient names are used to discuss the methods of estimation.

$$\hat{Y} = k_1 + c\,X \tag{a}$$
$$\hat{M}_i = k_2 + a_i X_i \tag{b}$$
$$\hat{Y} = k_3 + c'\,X + \Sigma\, b_i M_i \tag{c}$$

Where X is the explanatory variable, Y is the dependent variable and M_i symbolizes mediators, subscript i is indicating number of mediators e.g. (M_1 is the first mediator, M_2 is a second mediator and so on). Similarly, a_1 a_2 a_3 are representing the paths from X to respective mediators while b_1 b_2 b_3 are denoting coefficients of respective mediators in an equation containing all mediators and X variable as predictors of Y.

The causal Steps Approach

Causal steps approach follows the steps prescribed by (Baron & Kenny, 1986; Judd & Kenny, 1981) that are the basis of mediation analysis. This method derives information from the estimation of equations mentioned at the start of this section. This criterion requires for a variable to be the mediator if the path coefficients a, b and c are all significant and introduction of the mediator(s) significantly reduces the direct effect of X on Y. So this approach basically investigates whether the change between c and c' is significant after the introduction of the mediator(s) holding that paths from X to M and M to Y are significant. The flaw with casual step approach is that it does not directly test the mediation effect. It cannot test how stronger the impact of the mediator is or how large the indirect effect is. And whether the indirect effect is significantly different from zero or not. It is discussed in (MacKinnon et al., 2002) that this approach is not powerful enough to discover an indirect effect. While (Holmbeck, 2002) reported that the causal approach may have inconsistencies in its results. For example, 'a'and 'b' paths may be significant in a model but still, it does not reduce the direct effect of X to Y, which will lead to an inconclusive result. Moreover, this approach also does not provide standard errors and confidence intervals for indirect effect and in the case of multiple mediators, it could not split the effects of mediators. (Preacher & Hayes, 2008) concluded regarding casual step approach that it is of less utility, especially in multiple mediation analysis.

Product of Coefficients Method

Product of coefficient approach is an extension and substantial improvement of the causal step approach nullifying most of the flaws in it. This method ascertained that the product of paths from X to M and then from M to Y represents, the indirect effect. In terms of notations in the above-mentioned model, the indirect effect is the product of 'a' and 'b' in a simple mediation model. While in case of multiple mediators the specific indirect effects are represented by respective a and b paths and total indirect effect is $\Sigma a_i b_i$. This method also provides tools for testing these mediation effects, in case of simple mediation the most popular method for testing the significance of the indirect effect is the Sobel test. **Sobel test** assumes that ab follows the normal distribution and the significance of this indirect effect can be tested by the following formula

$$z = \frac{\mathrm{a} * \mathrm{b}}{\sqrt{b^2\,\sigma_a{}^2 \,+\, a^2\,\sigma_b{}^2}} \qquad \text{(a)}$$

Where denominator shows the standard error of indirect effect (ab) which is based on the work of (Aroian, 1947) and also illustrated by (Mood et al., 1974) using the 2nd order Taylor approximation of product of coefficients ab. The formula of Standard error of ab by Taylor approximation also includes the cross product of variance of 'a' and variance of 'b'. It is dropped in Sobel test formula knowing that this cross product is too small to be included (MacKinnon & Dwyer, 1993; Sobel, 1982) mentioned that this cross product of variances are not required to be included because it does not appear if standard error of ab is derived through delta method. (Preacher & Hayes, 2008) discuss the Sobel test and delta method for specific indirect effect in case of multiple mediator models. It is proposed that these methods can be extended to find indirect effects and the standard errors to test the mediation effect of each mediator. For instance the indirect effect of the third mediator in the model M_3 will be equals to a_3b_3. Test of significance will be performed by the formula given below where the term in the denominator is a standard error of indirect effect.

$$z = \frac{a_3b_3}{\sqrt{{b_3}^2\, {\sigma_{a_3}}^2 + {a_3}^2\, {\sigma_{b_3}}^2}} \quad \text{(b)}$$

The main shortcoming of the product of coefficient method is that it assumes the product of coefficients distribution to be normally distributed which is contradicted by studies of (Aroian, 1947; Goodman, 1960; Mackinnon et al., 2007) stating that the paths a and b even if individually follow normal distribution their products are not normal and are skewed.

Use of Structural Equation Modelling (SEM) for Mediation Analysis

Structural Equation modelling for estimation of parameters in the mediation model is usually preferred for a variety of reasons. SEM is theoretically stronger than OLS technique because it estimates all the related variable coefficients simultaneously. It makes the estimation and interpretation of complex models easy. It deals easily the multiple exogenous variables, mediators and outcome variable. Handling of missing data is also easy in SEM. It is also preferred for its information given on model fit. It is especially more useable in complex mediation models where there are multilevel mediation processes, when the effect of X transmits to Y through many in-between nodes, when there is a mix of latent and observed variables or when all or most of variables are latent variables.

However, SEM although preferable for all the reasons mentioned above requires specialized techniques while dealing with binary mediators as it assumes that all the variables are continuous and have a multivariate normal distribution. Binary variables often violate this assumption.

Bootstrapping Technique

Bootstrapping technique is preferable and recommended particularly in small and moderate samples because it does not assume the distribution of indirect effects (ab) to be normal; these methods do not require to make any assumptions about the distribution of parameter estimates to be tested (Lockwood & MacKinnon,1998; Shrout & Bolger, 2002; Preacher & Hayes, 2008). (Bollen & Stine, 1990; Stone & Sobel, 1990) using bootstrapping and simulation studies respectively found that distribution of the product of ‘a’ and ‘b’ paths i.e. of indirect effect is skewed, they used small sample sizes of less than 400 observations. So it is more appropriate to utilize non-parametric test instead of the conventional tests for testing mediation e.g. Sobel test which assumes the distribution of indirect effect to be normal. As not only it does not make assumptions about the distribution of the test but also has a higher power of the test.

Bootstrapping in its general sense means a self-executing process that does not require external input. In a statistical view, it is a resampling method to re-estimate a required summary statistics some large numbers of time (usually thousands). The method of bootstrapping is equally applicable to calculate Standard errors, p-values, and confidence interval of direct, total indirect and specific indirect effects in mediation analysis. Bootstrapping method do resampling, based on the original sample some thousands of time. From each generated sample the required statistics is calculated e.g. specific indirect effect. In this manner, there are thousands of estimated values for the required parameter to be tested. Based on these estimated values then standard errors, p-values and confidence interval are calculated. Furthermore, point estimates can be calculated by averaging all the estimated values of a parameter.

Discussion on some complexities involved in mediation analysis

This section will discuss some of the complexities involved in mediation analysis and the proposed methods in literature to deal with it.

Categorical Variables

Iacobucci, (2012) in the article “mediation analysis and categorical variables: the final frontier” discussed the definition of a categorical variable in the mediation analysis context. It is clearly stated in the study that 5, 7 and 9 point scale variables are treated as continuous and the variables with binary outcomes e.g. yes or no and nominal categories like ethnicity groups and brands like A, B or C will be considered as categorical variables.

Estimation of the mediation model with categorical variables

It is illustrated in (Iacobucci, 2012) that if mediator (M) or outcome Y is binary then it is not optimal to use traditional approaches to estimate mediation model equations. It is stated that “OLS regression is better suited for continuous response variables and logit model are better suited for discrete response variables”. It is also indicated that SEM is difficult to implement in case if the mediator or dependent variable is binary. It is described in the following words “It seems daunting to ask that Eqs. (1)-(3) could be reformulated such that all parameter estimates be estimated simultaneously when they comprise a mix of regression and logistic regressions, with their variant requisite assumptions”. Similarly, it is quoted that “For our purposes, this issue means that we cannot fit the equivalent of a structural equation model wherein one path is a logit link and another is OLS regression".

Method proposed by (Iacobucci, 2012)

It is advocated to use Structural Equation modelling, if all X, M and Y variables are continuous. However, if the Mediator or Y is categorical, after comparing and listing shortcomings of different techniques available e.g. Neural Networks, generalized least squares and using R^2 instead of β, the following procedure is recommended to test the mediation.

Step 1: If the dependent variable Y is continuous then equation 1 presented into mediation model given in the start of this section will be estimated through OLS otherwise logistic regression will be used for estimation.

Step 2: If Mediator is continuous equation 2 will be estimated by OLS and if it is categorical then estimations will be done by logistic. This will give parameter estimate of 'a' and its standard error.

Step 3: If Y is continuous then the 3^{rd} equation will be estimated by OLS otherwise logistic regression will be used. This will give the parameter estimate of 'b path' and its standard error.

Step 4: Using information collected in step 2 and 3, the standardized elements can be found.

$$z_a = \frac{\hat{a}}{\hat{s}_a} \tag{a}$$

$$z_b = \frac{\hat{b}}{\hat{s}_b} \tag{b}$$

their product will be $Z_{a \times b} = Z_a Z_b$ (c)

and standard error is given by

$$\sqrt{z_a^2 + z_b^2 + 1} \tag{d}$$

Step 5: Finally the test of mediation proposed in the study is given by the following formula which is equally applicable to whether a and b both are from OLS or 'a' is from logistic and 'b' is from OLS or whether both are from logistic.

$$z_{Mediation} = \frac{z_a z_b}{\sqrt{z_a^2 + z_b^2 + 1}} = \frac{\frac{a}{s_a} \times \frac{b}{s_b}}{\sqrt{z_a^2 + z_b^2 + 1}} \tag{e}$$

A mediator is significant at α = 0.05 if its value exceeds 1.96 in absolute terms. The critical values at a different level of significance in a normal distribution.

Method of standardization of coefficients from the logit model

In this method, the indirect effect is found by multiplication of standardized coefficients. Indirect effects are calculated by standardized 'a' and 'b' because both are coming from different regressions one from logistic and other from OLS if M or Y is binary. When mediator (M) is a binary variable 'a' path will be standardized by the formula proposed by (MacKinnon & Dwyer, 1993) that is given by

$$a^* = \frac{\hat{a}\, S_x}{S_M} \quad \text{(a)}$$

But if M is binary its standard error will be equal to $S_M = \sqrt{a^2\, s_x{}^2 + \frac{\pi^2}{3}}$ while b path will also be standardized if Y is also binary.

Method proposed by Li et. al (2007)

Li et al, (2007) by performing Monte Carlo study claimed that the method proposed by them for finding mediation effect is consistent as compared to four other estimates that turned to be inconsistent in presence of a binary mediator. The following model is used for the analysis to capture the effect of X variable through binary mediator M on dependent variable Y where z represents covariate.

$$M = I\,(\,c_M + \alpha_o \mathrm{X} + \gamma_{zM}\mathrm{Z} + \varepsilon_M > 0) \quad \text{(a)}$$

Where *I (·)* is representing binary mediator in terms of the indicator function. Whereas the equation for Y is given by

$$Y = c_y + \tau\,\mathrm{X} + \beta_o \mathrm{M} + \gamma_{zy}\mathrm{Z} + \varepsilon_y \quad \text{(b)}$$

Error terms in both models are assumed to be mean zero random variables. The subscripts of coefficients are providing information about variable with which it is associated and about whether it is from the equation of mediator or dependent variable Y. For example γ_{zy} is the coefficient of variable Z in the equation of Y.

Taking the definition of mediation effect, that it is changed in the mean value of Y which is transferred through M due to change in X variable, it could be summarized as

$$\frac{\partial\,\{E\,(Y \mid X,Z) - E(Y \mid X,M,Z)\,\}}{\partial X} = \delta_o \quad \text{(c)}$$

Where $E\,(Y \mid X,Z)$ denotes the conditional mean of Y given X and **Z** while $E(Y \mid X,M,Z)$ is conditional mean of Y given X, Z and M. Utilizing equation (b) $E\,(Y \mid X,Z)$ can be written as

$$E\,(Y \mid X, Z) = c_y + \tau\,\mathrm{X} + \beta_o E(Y \mid X, M, Z) + \gamma_{zy}\mathrm{Z} \qquad \text{(d)}$$

$$E\,(Y \mid X, M,\ Z) = c_y + \tau\,\mathrm{X} + \beta_o M + \gamma_{zy}\mathrm{Z} \qquad \text{(e)}$$

Plugging equation 4 and 5 in equation 3 will give the following equation

$$\delta_o = \beta_o\,\frac{\partial\,\{E\,(M \mid X, Y)\}}{\partial X}$$

$$= \beta_o\,\frac{\partial\,\{\Pr(\ c_M + \alpha_o\mathrm{X} + \gamma_{zM}\mathrm{Z} + \varepsilon_M > 0) \quad \mid X, Z\,\}}{\partial X} \qquad \text{(f)}$$

It could be seen that equation (f) is representing the mediation effect which is also equivalent to the product of coefficients.

- **Estimation**

For estimation of mediation effect through utilizing the above theoretical derivation, equation 1 could be estimated through logit or probit regression and equation 2 will be estimated through Ordinary Least Squares (OLS). When ε_M follows a logistic distribution, the mediation effect presented by equation (f) can be written as

$$\delta_{AL} = \alpha_o\beta_o\,\frac{e^{\,c_M + \alpha_o\mathrm{X} + \gamma_{zM}\mathrm{Z}}}{(1+e^{\,c_M + \alpha_o\mathrm{X} + \gamma_{zM}\mathrm{Z}})^{\,2}} \qquad \text{(g)}$$

The estimated $\hat{\delta}_{AL}$ can be calculated by plugging the estimated values of the coefficient given in eq (g). Getting from following equations

$$logit\,(\widehat{M}) = \hat{c}_{M.XY} + \hat{\alpha}_{XM.Z}\mathrm{X} + \gamma_{zM.X}\mathrm{Z} \qquad \text{(h)}$$

$$\hat{Y} = \hat{c}_{Y.XMZ} + \hat{\tau}_{XY.MZ}\mathrm{X} + \hat{\beta}_{MY.XZ}\mathrm{X} + \hat{\gamma}_{ZY.XM}\mathrm{Z} \qquad \text{(i)}$$

Finally

$$\hat{\delta}_{AL} = \hat{\alpha}_{XM.Z}\hat{\beta}_{MY.XZ}\,\frac{\exp(\,\hat{c}_{Y.XMZ}+\hat{\alpha}_{XM.Z}\,\mathrm{X}+\hat{\gamma}_{ZY.XM}\,\mathrm{Z})}{\{(1+\exp(\,\hat{c}_{Y.XMZ}+\hat{\alpha}_{XM.Z}\,\mathrm{X}+\hat{\gamma}_{ZY.XM}\,\mathrm{Z})\}^{\,2}} \qquad \text{(j)}$$

Estimated mediation will be the average of $\hat{\delta}_{AL}$ for each observation in the sample. Bootstrapping could be used to construct the C-I and test its significance.

Karlson, Holm and Breen (KHB) method

Karlson et al.; (2012, 2013) addressed the problem of comparing coefficients of non-linear regression e.g. logit and probit models with coefficients of linear regression models. They proposed and justified a new method to deal with a binary outcome variable or binary mediator. The method is briefly discussed as follows. The article uses y to denote continuous outcome variable, x for predictor variable and z for mediating variable(s) and y* is used to represent a binary outcome. The following two models are specified

$$H_R : y^* = \beta_{yx}x + e \qquad (a)$$

$$H_F : y^* = \beta_{yx}x + \beta_{yz.x}z + v \qquad (b)$$

Where R stands for reduced model and F for the full model.
Moreover, let x is linearly related to z and could be presented as

$$\text{z} = \theta_{zx} + w \qquad (c)$$

They introduced a new method that utilized the residual term from z on x regression represented by z̃. Equation 'c' given below is the full model with z̃ while equation 'b' is a full model with z

$$H_F^* : \; y^* = \beta_{yx.\tilde{z}}x + \beta_{y\tilde{z}.x}\tilde{z} + k, \qquad sd(k) = \sigma_F \frac{\pi}{\sqrt{3}} \qquad (d)$$

They proved that

$$\beta_{yx} = \beta_{yx.\tilde{z}} \qquad (e)$$

They proposed the following three measures to capture the mediating effect (confounding effect) holding rescaling effect constant

The difference measure

$$b_{yx.\tilde{z}} - b_{yx.z} = \frac{\beta_{yx.\tilde{z}}}{\sigma_F^*} - \frac{\beta_{yx.z}}{\sigma_F} = \frac{\beta_{yx}}{\sigma_F} - \frac{\beta_{yx.z}}{\sigma_F} = \frac{\beta_{yx} - \beta_{yx.z}}{\sigma_F} \qquad (f)$$

Ratio measure

$$\frac{b_{yx.z}}{b_{yx.\tilde{z}}} = \frac{\frac{\beta_{yx.z}}{\sigma_F^*}}{\frac{\beta_{yx.\tilde{z}}}{\sigma_F}} = \frac{\frac{\beta_{yx.z}}{\sigma_F}}{\frac{\beta_{yx}}{\sigma_F}} = \frac{\beta_{yx.z}}{\beta_{yx.\tilde{z}}} = \frac{\beta_{yx.z}}{\beta_{yx}} \qquad (g)$$

Percentage measure

$$\frac{b_{yx.\tilde{z}} - b_{yx.z}}{b_{yx.\tilde{z}}} \times 100\% = \frac{\frac{\beta_{yx.\tilde{z}} - \beta_{yx.z}}{\sigma_F}}{\frac{\beta_{yx.\tilde{z}}}{\sigma_F^*}} \times 100\% = \frac{\beta_{yx.\tilde{z}} - \beta_{yx.z}}{\beta_{x|\tilde{z}}} \times 100$$

$$= \frac{\beta_{yx} - \beta_{yx.z}}{\beta_{yx}} \times 100\% \qquad (h)$$

- Implementation of KHB method to the binary mediator and continuous outcome

Karlson et al., (2013) clearly discussed the case of a binary mediator and continuous outcome variable and summarized it by equations 26a, 26b and 26c in their article

$$y^{*} = \gamma_{yx.z}x + \gamma_{yz^{*}.x}z^{*} + \sigma_{n}\omega \quad \text{(i)}$$

Where z^{*} is representing mediator with a binary outcome.

$$z^{*} = \theta_{z^{*}x}x + m \quad \text{(j)}$$

The total effect β_{yx} for the binary mediator and the continuous outcome is transcribed in the following way

$$\beta_{yx} = \gamma_{yx.z^{*}} + \theta_{z^{*}.x}\gamma_{yz^{*}.x} \quad \text{(k)}$$

Karlson et al. (2012, 2013) not only proposed a new method but also provided a KHB package for STATA. This package has the ability not only to implement the method proposed but also to calculate Average Partial effects (APE).

Latest Approach: Marginal Mediation

Marginal mediation is the latest approach to deal with binary mediators and/ or binary outcomes and to get effect size of each mediator without any interpretability issues. (Barrett, 2018) discussed the existing methods and their limitations to deal with categorical outcomes and mediation. The study discussed five current approaches i.e. estimation of each path by logistic regression which lacks estimation of indirect effect size, polychoric correlation in SEM, the method proposed by (Iacobucci, 2012), the sequence of logistic regression models and the fifth one is ignoring the distribution of outcomes. The pros and cons of each method are discussed in detail and the conclusion is drawn in the following words

> “ In the end, none of these approaches can flexibly handle categorical variables of various kinds (e.g., binary, multinomial) or other non-normal distributions (e.g., costs, counts); none can produce intuitive and meaningful effect sizes and confidence intervals across these variable types; and none can consistently combine two different types of estimates (e.g., binary mediator with continuous outcome, count mediator with binary outcome). Although the current approaches are useful in some situations—particularly the SEM approach—a more complete framework is needed” (Barrett, 2018).

Barrett, (2018) proposed a new technique named “Marginal Mediation Analysis” by the integration of Average Marginal Effects (AMEs) proposed by (Bartus, 2005) in mediation analysis. It is illustrated in the study that the marginal mediation analysis makes the interpretation of indirect effects more meaningful and simple in the models which were previously considered problematic (models with categorical mediators and outcomes). This method takes into account the marginal effect of each variable for each unit in the sample and then this marginal effects for every observation are averaged to get the meaningful number called Average Marginal Effect (AME). It is elaborated that AMEs are robust in case of model misspecification and unobserved heterogeneity and generally applicable to any type and mix of models.

The process of marginal mediation analysis follows the following steps that integrate AME method to mediation analysis. First, the following equations are estimated, same notations are used as in (Barrett, 2018).

$$M_{ij} = a_0 + \sum_{k=1}^{p} a_k x_{ki} + \varepsilon_i \quad \text{for j= 1 , ...m mediators} \tag{a}$$

$$Y_i = \beta_0 + \sum_{j=1}^{m} b_j M_{ij} + \sum_{k=1}^{p} c'_k x_{ki} + \varepsilon_i \tag{b}$$

Where there are k= 1, 2 ....p predictors, 1 to m mediators, and a, b and c′ paths are the same as in the usual mediation analysis. These equations are representing mediation model as it is in practice, and estimation can be done keeping in view the type of outcome in each equation. The novelty of marginal mediation method is in post-estimation steps and integration of AME with above-mentioned equations. For continuous predictor, AME of path a for the kth variable is presented by the equation given below

$$AME_k^a = a_k \frac{1}{n} \sum_{i=1}^{n} f(aX) \tag{c}$$

Where f is denoting the probability density function, linear combination of X variables (predictors) is represented by aX in the above equation. The above mention equation is identical to the equation given below.

$$AME_k^a = \frac{1}{n} \sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} \frac{f(a\boldsymbol{X_1}) - f(a\boldsymbol{X_2})}{2h} \tag{d}$$

Where aX_1 and aX_2 are given by the following matrices

$$aX_1 = \begin{bmatrix} ax_{11} & ax_{12} & \dots & ax_{1k}+h & \dots & ax_{1p} \\ ax_{21} & ax_{22} & \dots & ax_{2k}+h & \dots & ax_{2p} \\ \vdots & \vdots & \dots & \vdots & \dots & \vdots \\ \vdots & \vdots & \dots & \vdots & \dots & \vdots \\ \vdots & \vdots & \dots & \vdots & \dots & \vdots \\ ax_{n1} & ax_{n2} & \dots & ax_{nk}+h & \dots & ax_{np} \end{bmatrix} \tag{e}$$

$$aX_2 = \begin{bmatrix} ax_{11} & ax_{12} & \dots & ax_{1k} - h & \dots & ax_{1p} \\ ax_{21} & ax_{22} & \dots & ax_{2k} - h & \dots & ax_{2p} \\ \vdots & \vdots & \dots & \vdots & \dots & \vdots \\ \vdots & \vdots & \dots & \vdots & \dots & \vdots \\ \vdots & \vdots & \dots & \vdots & \dots & \vdots \\ ax_{n1} & ax_{n2} & \dots & ax_{nk} - h & \dots & ax_{np} \end{bmatrix} \qquad \text{(f)}$$

Matrices given above are the predicted changes in outcome due to very small changes h, where h = 1 $\times$ 10^{-7}.

While for dummy variable in "a" path, AME is

$$AME_k^a = \frac{1}{n} \sum_{i=1}^{n} [F(aX|\boldsymbol{x_{ki}} = 1) - F(aX|\boldsymbol{x_{ki}} = 0)] \qquad \text{(g)}$$

Where the $F(aX|\boldsymbol{x_{ki}} = 1)$ is ith obseravtion's predicted value when the dummy is one and $F(aX|\boldsymbol{x_{ki}} = 0)$ is predicted value at value 0. Same equations can be applied to calculate AME for b and c´ paths.

By following the above-mentioned procedure and reporting indirect effect sizes in the form of AMEs makes the results more interpretable. It is explained in the study that the results of individual paths e.g. path 'a' are in metric of the dependent variable in the equation. For example, if a path for a mediator in the binary form will represent risk or probability while if the mediator is continuous then it will be interpreted as the results in ordinary regression are interpreted. The indirect effects are based on both a and b paths, and if mediator and outcome variable are in two different types (i.e binary, continuous or continuous, binary or any other mix) then a and b paths will be in different scale, but the reported indirect effects calculated through marginal mediation approach will be in AMEs and will be easily interpretable. Under marginal mediation approach, the indirect effects are in outcomes original metric and thus do not create the problem of interpretability. Similarly, direct and total effects are measured in the form of AMEs and interpretable in outcomes metric. Because each indirect effect, direct effect and the total effect is in the same metric (i.e. in outcome's metric) therefore they are all comparable. Marginal mediation analysis is implementable due to recently developed and available package "MarginalMediation" in R software, that is easy to use. This package reports both standardized and unstandardized direct effects and specific indirect as well as the confidence intervals through the bootstrapping method.

To provide a clear overview of the mediation methods discussed, Table 1 summarizes their key features, strengths, and weaknesses, particularly in relation to handling binary mediators

Table 1

Summary of Mediation Analysis Techniques and Their Properties

Method	Key features	Handling Binary Mediators	Strengths	Weaknesses
Casual Step Approach	The variable is claimed mediator if significantly reduces the direct effect.	Inadequate for complex models	Easy to understand	It only tests mediation not its strength or coefficient value. Inconsistencies and inconclusive results are possible.
Product of Coefficients Method (Sobel Test)	Indirect effect is calculated as the product of a, b paths	Inadequate for categorical variables	Directly tests Indirect effect. Estimates the strength of mediation effect	It assumes the product of coefficients distribution to be normally distributed which is contradicted by studies
Structural Equation Modelling for Mediation	Simultaneous Estimation of Paths	Necessitates specialized techniques for handling binary mediators	Handles complex models and latent variables	It assumes that all the variables are continuous and have a multivariate normal distribution. Specialized techniques add complexity.

Bootstrapping Technique	Resampling to estimate mediation effects	Flexible	Requires no distributional assumptions	Computationally intensive, though modern software significantly reduces processing time; while applicable to all sample sizes, its benefits are most obvious in small samples.
Iacobucci (2012)	Standardized coefficients from mixed models	Addresses binary mediators	Particularly designed to deal with binary mediators	Relies on Standardization Possibilities of inconsistencies
Li et al. (2007)	Conditional Mean Approach	Addresses binary mediators	Consistent with binary mediators	Assumes specific model specification, relies on assumptions of logistic/probit and OLS; can be complex to implement.
KHB (2013)	Compares coefficients between nested models	Addresses binary mediators	Addresses rescaling issues and making the coefficients comparable in non-linear regressions	Computationally intensive
Marginal Mediation (2018)	Average Marginal effects	Addresses binary mediators	Interpretable effect sizes	----------

Conclusion

This study attempted to discuss the mediation analysis and different complexities involved in estimation of such models. The focus of the paper was dealing with binary mediators. Some recent methods were discussed in detail highlighting the theoretical underpinnings of the techniques. The strengths and weaknesses of these techniques were also deliberated in detail. The paper began with conceptual details of mediation analysis, followed by illustration of traditional estimation methods, and ended by elaboration of contemporary techniques available to deal with binary mediators. Key finding highlight that while traditional methods like product of coefficients have historical significance, they lack rigor in dealing with binary mediators and generate interpretation and comparison issues. Newer methods like (Li et al. 2007; KHB, 2013; Marginal Mediation, 2018) are more efficient in handling binary mediators, with Marginal mediation being more efficient in its capacity to estimate comparable and efficient mediation effects.

Practical implication

The academic implications of this paper are significant. By thoroughly elaborating and comparing various methods in detail, this paper is a valuable source for researchers across disciplines. It provides a valuable resource to understand the complexities involved in dealing with binary mediators, various techniques available to deal them and to select the most suitable one. Hence helping the researchers to choose the techniques more rigorously and accurately.

Limitations and future research direction

While the paper consists of a comprehensive comparison of existing methods it doesn’t include empirical analysis and simulations. Future research is needed to provide empirical evidence of effectiveness and limitations of contemporary techniques discussed in this paper.

References

Aroian , L. A. (1947). The probability function of the product of two normally distributed variables. *The Annals of Mathematical Statistics*, 18(2), 265-271. https://doi.org/10.1214/aoms/1177730442

Baron, R.M., & Kenny, D.A. (1986). The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, 51(6), 1173-1182. https://doi.org/10.1037/0022-3514.51.6.1173

Barrett, T. S. (2018). *Marginal mediation analysis: A new framework for interpretable mediated effects* (Doctoral dissertation, Utah State University).

Bartus, T. (2005). Estimation of marginal effects using margeff. *The Stata Journal*, 5(3), 309-329. https://doi.org/10.1177/1536867X0500500303

Bollen, K. A., & Stine, R. (1990). Direct and indirect effects: Classical and bootstrap estimates of variability. *Sociological methodology*, 20, 115-140. https://doi.org/10.2307/271084

Breen, R., Karlson, K. B., & Holm, A. (2013). Total, direct, and indirect effects in logit and probit models. *Sociological Methods & Research*, 42(2), 164-191. https://doi.org/10.1177/0049124113494572

Goodman, L. A. (1960). On the exact variance of products. *Journal of the American statistical association*, 55(292), 708-713. https://doi.org/10.1080/01621459.1960.10483369

Holmbeck, G. N. (2002). Post-hoc probing of significant moderational and mediational effects in studies of pediatric populations. *Journal of pediatric psychology*, *27*(1), 87-96. https://doi.org/10.1093/jpepsy/27.1.87

Iacobucci, D. (2012). Mediation analysis and categorical variables: The final frontier. *Journal of Consumer Psychology* 22 (2012) 582–594. https://doi.org/10.1016/j.jcps.2012.03.006

Judd, C. M., & Kenny, D. A. (1981). Process analysis: Estimating mediation in treatment evaluations. *Evaluation review*, 5(5), 602-619. https://doi.org/10.1177/0193841X8100500502

Karlson, K. B., Holm, A., & Breen, R. (2012). Comparing regression coefficients between same-sample nested models using logit and probit: A new method. *Sociological Methodology*, 42(1), 286-313. https://doi.org/10.1177/0081175012444861

Li, Y., Schneider, J. A., & Bennett, D. A. (2007). Estimation of the mediation effect with a binary mediator. *Statistics in Medicine*, 26(18), 3398-3414. https://doi.org/10.1002/sim.2730

Lockwood, C. M., & MacKinnon, D. P. (1998). Bootstrapping the standard error of the mediated effect. *Proceedings of the 23rd annual meeting of the SAS Users Group International* (pp. 997-1002). Cary, NC: SAS Institute.

MacKinnon, D. P., & Dwyer, J. H. (1993). Estimating mediated effects in prevention studies. *Evaluation review*, 17(2), 144-158. https://doi.org/10.1177/0193841X9301700202

MacKinnon, D. P., Fairchild, A. J., & Fritz, M. S. (2007). Mediation analysis. *Annu. Rev. Psychol*., 58(1), 593-614. https://doi.org/10.1146/annurev.psych.58.110405.085542

MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G., & Sheets, V. (2002). A comparison of methods to test mediation and other intervening variable effects. *Psychological methods*, 7(1), 83. https://doi.org/10.1037/1082-989X.7.1.83

Mood, A., Graybill, F., & Duane, C. D. Boes (1974). *Introduction to the Theory of Statistics*. New York: McGraw-Hill.

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior research methods*, *40*(3), 879-891. https://doi.org/10.3758/BRM.40.3.879

Shrout P. E., Bolger N. (2002), Mediation in experimental and non-experimental studies: New procedures and recommendations. *Psychological methods*. 7(4), 422-445. https://doi.org/10.1037/1082-989X.7.4.422

Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. *Sociological methodology*, 13, 290-312. https://doi.org/10.2307/270723

Stone, C. A., & Sobel, M. E. (1990). The robustness of estimates of total indirect effects in covariance structure models estimated by maximum. *Psychometrika*, 55(2), 337-352. https://doi.org/10.1007/BF02295291