STATISTICS IN MEDICINE
Statist. Med. 2004; **23**:2713–2728 (DOI: 10.1002/sim.1847)

Statistical assessment of mediational effects for logistic mediational models

Bin Huang[1,2,*,†], Siva Sivaganesan[3], Paul Succop[4] and Elizabeth Goodman[5]

[1]*Center for Epidemiology and Biostatistics, Children's Hospital Medical Center, Cincinnati, OH, U.S.A.*
[2]*Division of Adolescent Medicine, Children's Hospital Medical Center, Cincinnati, OH, U.S.A.*
[3]*Department of Mathematics, University of Cincinnati, Cincinnati, OH, U.S.A.*
[4]*Department of Environmental Health, University of Cincinnati, Cincinnati, OH, U.S.A.*
[5]*Heller School for Social Policy and Management, Brandeis University, Waltham, MA, U.S.A.*

SUMMARY

The concept of mediation has broad applications in medical health studies. Although the statistical assessment of a mediational effect under the normal assumption has been well established in linear structural equation models (SEM), it has not been extended to the general case where normality is not a usual assumption. In this paper, we propose to extend the definition of mediational effects through causal inference. The new definition is consistent with that in linear SEM and does not rely on the assumption of normality. Here, we focus our attention on the logistic mediation model, where all variables involved are binary. Three approaches to the estimation of mediational effects—Delta method, bootstrap, and Bayesian modelling via Monte Carlo simulation are investigated. Simulation studies are used to examine the behaviour of the three approaches. Measured by 95 per cent confidence interval (CI) coverage rate and root mean square error (RMSE) criteria, it was found that the Bayesian method using a non-informative prior outperformed both bootstrap and the Delta methods, particularly for small sample sizes. Case studies are presented to demonstrate the application of the proposed method to public health research using a nationally representative database. Extending the proposed method to other types of mediational model and to multiple mediators are also discussed. Copyright © 2004 John Wiley & Sons, Ltd.

KEY WORDS: mediational model; GLM; SEM; Bayesian model; Delta method; bootstrap

1. INTRODUCTION

Understanding the associations between risk factors and health outcomes is fundamental to epidemiological health related research. Such associations are often complex, which involve correlated biological, environmental, social, and behavioural factors. The impact of one specific

*Correspondence to: Bin Huang, Center for Epidemiology and Biostatistics, MLC 5041, Cincinnati Children's Hospital Medical Center, 3333 Burnet Avenue, Cincinnati, OH 45229-3039, U.S.A.
†E-mail: bin.huang@cchmc.org

Received November 2003
Accepted February 2004

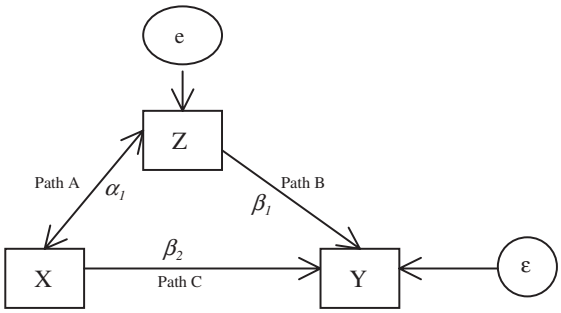

Figure 1. Path diagram for the mediation model, where X is the independent variable, Y is the dependent variable and Z is the mediator.

factor on the health outcome cannot be fully understood unless it is evaluated in the context of other associated factors [1]. There are different roles these other associated factors can play in the relationship between a specific factor, i.e. an independent variable, and an outcome. The role of a mediator is what we focus on in this paper.

According to Baron and Kenny [2], a mediator specifies how, or the mechanism by which, a given effect occurs. A mediational model describes 'the generative mechanism through which the focal independent variable is able to influence the dependent variable of interest' (p. 1173). Simply put, a mediational model shows how the key independent variable affects an intermediate outcome or mediator, which in turn leads to the final outcome. Figure 1 presents a simple mediational model, where X is the independent variable, Y is the dependent variable and Z is the mediator.

The mediational model has broad applications in public health and medical research. It can be used to identify a mediator or an intermediate outcome of an exposure–outcome process, to test a theoretical mediational model, or to design and evaluate an intervention program. For example, the mediational model can help identify an intermediate outcome to cancer, AIDS or cardiovascular disease for early prevention or treatment. The widely used planned health behaviour theory is an example of a theoretical mediational model [3]. According to this theory, intention to acquire a health behaviour mediates the linkage between health attitudes (X) and the actual health behaviour (Y). Validating such a theoretical model using a mediational statistical model could help promote healthy behaviour. Mediational models have been recommended for intervention studies to evaluate the intervention effect [4–6]. In such case, the intervention program (X) is designed to alter an intermediate outcome (Z), which is a precursor leading to the targeted change in the final outcome (Y). For example, a teen smoking intervention (X) may be designed to improve the teen–parents communication (Z) on tobacco issues, in order to prevent the teen from initiation of smoking (Y).

The motivating example of the current study concerns the mediational effect of depression (Z) in the linkage between socioeconomic status (X) and health outcome (Y) among adolescents. Socioeconomic (SES) status, including SES in childhood, has been shown to be a powerful and unique predictor of many health outcomes, but how and why SES is linked to these health outcomes is not clear [7]. A potential mechanism underlying the effect of SES

on health is identified as a depressed mood. For children and adolescents, lower parental SES may lead to higher likelihood of depression in adolescence, which in turn leads to higher prevalence of adverse health outcomes caused by the adolescent depression [7–9]. Here, depression is hypothesized as the mediator that explains the mechanism through which SES is linked with health outcomes. Two major questions are of public health interest. First, is the effect of SES on health mediated by depression in adolescents? Second, if yes, how much of the impact of SES on the health outcome is attributable to depression? Answers to these questions will promote understandings of socioeconomic health differentials, one of the major focuses of Healthy People 2010 [10].

The multiple regression procedure introduced by Baron and Kenny [2] is widely cited as providing the key strategies for evaluating whether a given variable is, in fact, a mediator. In general, under the assumptions that there is no measurement error in the mediator and that the dependent variable does not cause the mediator, this multiple regression procedure consists of three conditions and one comparison. The conditions are:

1. In the regression model of X regressed on Z, X must be significantly associated with Z;
2. In the regression model of X regressed on Y, X must be significantly associated with Y;
3. In the regression model of X and Z regressed on Y, Z must be significantly associated with Y.

If these conditions are met and if the regression coefficient of X from the last model is less than that of the second model, then one may conclude that Z is a mediator. While this approach is already widely used to demonstrate mediation, the current methodology is limited in scope. First, it does not provide a way to quantify the role of a mediator within the X to Y process. Second, there is no statistical significance test available to test whether Z is indeed a mediator.

The statistical assessment of the role of a mediator has been well developed under the normality assumption. Assuming both the mediator and the outcome are normally distributed, corresponding to the path diagram shown in Figure 1, the linear structural equation model (SEM):

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 z_i + \beta_2 x_i + \varepsilon_i \\ Z_i &= \alpha_0 + \alpha_1 x_i + e_i \end{aligned} \tag{1}$$

where $\varepsilon_i \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma^2)$, $e_i \overset{\text{iid}}{\sim} \mathrm{N}(0, \varpi^2)$ and $\mathrm{Cov}(\varepsilon_i, e_i) = 0$, distinguishes three types of effects [11]. The influence of one variable on another that is unmediated by any other variables is the direct effect (path C), measured by β_2. The indirect effect of a variable is that described by the pathways to and from the mediating variable(s) (paths A and B), measured by $\alpha_1\beta_1$. The sum of the direct and indirect effect is the total effects, measured by $\beta_2 + \alpha_1\beta_1$. Sobel [12] provides an approximate significance test for the indirect effect. Alternative approaches considering empirical distribution of the indirect effect, standardized indirect effect, or asymmetric confidence limits are also available under the normality assumption [13]. The statistical estimate and significance test of the indirect effect quantify the role of the mediator and provide statistical evidence of the hypothesized mechanism. However, in public health research, the outcome or the mediator is often not normally distributed. For example, when the presence or absence of a disease is the outcome of interest, logistic regression modelling is often used

to assess the effect of an exposure on the probability of disease. The methodology derived for path analysis is not directly applicable to the logistic model.

Methods currently available to assess mediational models when the normality is not valid are multi-step regression analysis [2], standardized latent variable multi-step regression for binary or ordinal variables [14], and structural equation models for binary or ordinal variables [15]. However, none of these methods are able to differentiate the direct and indirect effect of X on Y, and thus fail to provide statistical assessment of the mediational effect. When the question of interest specifically concerns the validation of an intermediate endpoint or a surrogate marker for the treatment effect on the presence or survival of a disease, a widely used measure is the proportion of treatment effect (PTE) [16]. The PTE is defined based on the steps 2 and 3 from Baron and Kenny's procedure. Let the regression coefficients of X from the steps 2 and 3 be β and β_a, then $\text{PTE} = 1 - \beta_a/\beta$. Many have suggested that PTE suffer from several serious limitations, including some undesirable statistical properties, and a lack of clear interpretation [17]. Freedman pointed out the need for a full modelling approach, which includes a model for the surrogate marker [18], and overcomes the difficulties in estimation for the logistic model.

In Section 2, we propose a general definition of direct, indirect and total effects through statistical causal inference, under the common notion of a generalized mediational model. For the linear mediational model, we demonstrate that the new definition of mediational effects are consistent with those provided in linear SEM, and are intuitively meaningful. In Section 3, we apply the general definition to the logistic mediational model, where all variables involved are binary, and discuss their epidemiological interpretations. In Section 4, we propose three approaches to the estimation of the mediational effects for the logistic model—Delta method, bootstrap, and a Bayesian method via Monte Carlo simulation from the posterior distribution. Simulation studies are used in Section 5 to examine the statistical properties of proposed methods. A motivating example is used in Section 6 to demonstrate an application of the proposed method to public health research using a nationally representative database. Discussions and suggestions for further studies are presented in Section 7.

2. GENERALIZING MEDIATIONAL EFFECTS THROUGH CASUAL INFERENCE

2.1. Casual interpretation of mediational effects in linear SEM

Although causal inference is the topic of considerable exposition and debate in the statistics community, it has been widely agreed that a cause-and-effect relationship could be understood in terms of the differences between the responses that would arise under different settings of the causal factor. We will re-examine the linear SEM definition of mediational effects through this thinking.

Assuming both Z and Y follow normal distributions, for the linear SEM (1), it is straightforward to derive the expectation of the marginal distribution of $Y \mid x$, $E[Y \mid x] = E_{Z \mid x}[E(Y \mid z,x)] = \beta_0 + \beta_1\alpha_0 + (\alpha_1\beta_1 + \beta_2)x$. As we already know, the regression coefficient $\alpha_1\beta_1 + \beta_2 = E[Y \mid x+1] - E[Y \mid x]$ measures the overall effect of X on Y, which is the difference between the expected responses that is due to one unit change in the exposure X. In order to separate the total effect into direct and indirect effects, one can imagine what the outcome Y would be, had the effect of X on Y not been mediated through Z; in other words,

one unit increase in X does not activate any change in the mediator Z, but only activates the change in the outcome Y by its direct influence. This imaginary response is the potential outcome Y, whose mean is given by,

$$E^P[Y \mid x+1] = \iint y f(y \mid z, x+1) f(z \mid x)\, \mathrm{d}z\, \mathrm{d}y = \beta_0 + \beta_1\alpha_0 + \beta_2 + (\alpha_1\beta_1 + \beta_2)x$$

If Z does not mediate the effect of X on Y at all, then $E^P[Y \mid x+1] = E[Y \mid x+1]$. If the effect of X on Y is completely mediated through Z, then $E^P[Y \mid x+1] = E[Y \mid x]$. Often, only a part of the effect of X on Y is mediated through Z. The difference of $E^P[Y \mid x+1] - E[Y \mid x] = \beta_2$ measures the part that is independent from Z, and the difference $E[Y \mid x+1] - E^P[Y \mid x+1] = \alpha_1\beta_1$ measures the part that is mediated through Z. Clearly, the first difference corresponds to the direct effect and the second to the indirect effect as defined in path analysis. Thus, consistent with linear SEM, we can redefine the total, direct and indirect effects as the differences between expected outcomes that is due to one unit change in X:

$$\begin{aligned} &\textit{Total effect}: && \mathrm{TOT} = E[Y \mid x+1] - E[Y \mid x] \\ &\textit{Direct effect}: && \mathrm{DIR} = E^P[Y \mid x+1] - E[Y \mid x] \\ &\textit{Indirect effect}: && \mathrm{IND} = E[Y \mid x+1] - E^P[Y \mid x+1] \end{aligned}$$

We also consider the scale free parameter IND/TOT proposed by Alwin and Hauser [19], which measures the proportion of the effect of X on Y that is mediated through Z, and name it as relative indirect effect (RIND).

2.2. Generalization of mediational effects

In order to extend the linear SEM measurements of mediational effect, we define the generalized mediational model as a system of two generalized linear models (GLM) [20],

$$\begin{aligned} g_1(E[Y_i \mid z_i, x_i]) &= \beta_0 + \beta_1 z_i + \beta_2 x_i \\ g_2(E[Z_i \mid x_i]) &= \alpha_0 + \alpha_1 x_i \end{aligned} \tag{2}$$

where

$$Y_i \mid z_i, x_i, \beta \overset{\text{iid}}{\sim} f_1(\mu_1)$$

$$Z_i \mid x_i, \alpha \overset{\text{iid}}{\sim} f_2(\mu_2)$$

$f_1(.)$ and $f_2(.)$ are known distributions with means of $\mu_1 = E[Y_i \mid Z_i, x_i]$ and $\mu_2 = E[Z_i \mid x_i]$, respectively. For example, when both $f_1(.)$ and $f_2(.)$ are normal and $g_1(.)$ and $g_2(.)$ are the identity links, the mediational model (2) is the same as the linear SEM (1), which we refer to as linear mediational model. When both $f_1(.)$ and $f_2(.)$ are binomial distribution and $g_1(.)$ and $g_2(.)$ are the commonly used logit link, the mediational model is

$$\begin{aligned} \mathrm{logit}[P(Y_i = 1 \mid z_i, x_i)] &= \beta_0 + \beta_1 z_i + \beta_2 x_i \\ \mathrm{logit}[P(Z_i = 1 \mid x_i)] &= \alpha_0 + \alpha_1 x_i \end{aligned} \tag{3}$$

which we referred to as logistic mediational model.

For the generalized mediational model (2), the conditional expectation of the observed outcome Y_i given X_i for the ith individual, $i = 1, \ldots, n$, is

$$E[Y_i \mid x_i] = \int_{z_i} \int_{y_i} y_i f_1(y_i \mid z_i, x_i) f_2(z_i \mid x_i)\, \mathrm{d}y_i\, \mathrm{d}z_i$$

Correspondingly, we define the potential outcome that is due to one unit increase in X_i as

$$E^P[Y_i \mid x_i + 1] = \int_{z_i} \int_{y_i} y_i f_1(y_i \mid z_i, x_i + 1) f_2(z_i \mid x_i)\, \mathrm{d}y_i\, \mathrm{d}z_i$$

The integration should be replaced with summation when Y or Z is discrete, for both $E[Y_i \mid x_i]$ and $E^P[Y_i \mid x_i + 1]$ above. Here, $E^P[Y_i \mid x_i + 1]$ is the (hypothetical) expectation of the outcome Y_i had the effect of X on Y not been mediated through Z. If X and Z are independent, i.e. $f_2(z_i \mid x_i) = f_2(z_i \mid x_i + 1)$, then $E^P[Y_i \mid x_i + 1] = E[Y_i \mid x_i + 1]$. If Z fully mediates the effect of X on Y, i.e. $f_1(y_i \mid z_i, x_i + 1) = f_1(y_i \mid z_i)$, then $E^P[Y_i \mid x_i + 1] = E[Y_i \mid x_i]$. Thus, for a generalized mediational model, we can define the mediational effects for the ith individual through contrasts between expectations of observed and potential outcomes as in the following:

Total effect:	$\mathrm{TOT}(x_i) = g_1(E[Y_i \mid x_i + 1]) - g_1(E[Y_i \mid x_i])$
Direct effect:	$\mathrm{DIR}(x_i) = g_1(E^P[Y_i \mid x_i + 1]) - g_1(E[Y_i \mid x_i])$
Indirect effect:	$\mathrm{IND}(x_i) = g_1(E[Y_i \mid x_i + 1]) - g_1(E^P[Y_i \mid x_i + 1])$
Relative indirect effect:	$\mathrm{RIND}(x_i) = \mathrm{IND}(x_i)/\mathrm{TOT}(x_i)$

The mediational effects as defined above are at the individual level. At the population level, the mediational effects are defined as the corresponding expectations of the individual mediational effects:

$$\text{Population mediational effect} = E[\mathrm{Med}(x)] = \int_x \mathrm{Med}(x) f(x)\, \mathrm{d}x$$

where $\mathrm{Med}(X)$ is the corresponding individual mediational effect. In other words, the population mediational effect is an average of the mediational effect at the individual level weighted by the probability of the independent variable X.

3. DEFINING MEDIATIONAL EFFECTS FOR THE LOGISTIC MEDIATIONAL MODEL

From this point on, we will focus our attention onto the logistic mediational model. Consider a common setting in public health studies, where the outcome (Y) is the presence (1)/absence (0) of a disease, and the mediator (Z) is the presence (1)/absence (0) of a risk factor, and the exposure (X) is the presence (1)/absence (0) of exposure to a hazard. For the logistic

mediational model (3), let

$$I_{\mathrm{H}} = E[Y \mid x=1] = p_{11}q_1 + p_{01}(1-q_1)$$
$$I_{\mathrm{L}} = E[Y \mid x=0] = p_{10}q_0 + p_{00}(1-q_0)$$
$$I_{\mathrm{H}}^{*} = E^{P}[Y \mid x=1] = p_{11}q_0 + p_{01}(1-q_0)$$

where

$$p_{ij} = P(Y=1 \mid z=i, x=j)$$
$$q_i = P(Z=1 \mid x=i)$$

for $i,j=0,1$. Here, I_{H} is the prevalence of disease for the high exposure ($X=1$) population, and I_{L} is the prevalence of disease for the low exposure ($X=0$) population, and I_{H}^{*} is the hypothetical expected prevalence in the high exposure population, had the prevalence of the risk factor Z in the higher exposure group been equal to the prevalence of Z in the low exposure group. Then, following the proposed definitions, we have,

$$\mathrm{TOT} = \mathrm{logit}(I_{\mathrm{H}}) - \mathrm{logit}(I_{\mathrm{L}}), \quad \mathrm{DIR} = \mathrm{logit}(I_{\mathrm{H}}^{*}) - \mathrm{logit}(I_{\mathrm{L}}), \quad \mathrm{IND} = \mathrm{logit}(I_{\mathrm{H}}) - \mathrm{logit}(I_{\mathrm{H}}^{*})$$

and

$$\mathrm{RIND} = \frac{\mathrm{logit}(I_{\mathrm{H}}) - \mathrm{logit}(I_{\mathrm{H}}^{*})}{\mathrm{logit}(I_{\mathrm{H}}) - \mathrm{logit}(I_{\mathrm{L}})}$$

Note that the exponential transformed mediational effects correspond to three odds ratios (OR),

$$\mathrm{OR}_{\mathrm{T}} = \mathrm{e}^{\mathrm{TOT}} = \frac{\mathrm{Odds}(I_{\mathrm{H}})}{\mathrm{Odds}(I_{\mathrm{L}})}, \quad \mathrm{OR}_{\mathrm{D}} = \mathrm{e}^{\mathrm{DIR}} = \frac{\mathrm{Odds}(I_{\mathrm{H}}^{*})}{\mathrm{Odds}(I_{\mathrm{L}})}, \quad \mathrm{OR}_{\mathrm{I}} = \mathrm{e}^{\mathrm{IND}} = \frac{\mathrm{Odds}(I_{\mathrm{H}})}{\mathrm{Odds}(I_{\mathrm{H}}^{*})}$$

The odds ratio of the total effect equals the product of the odds ratios of the direct and indirect effect, $\mathrm{OR}_{\mathrm{T}} = \mathrm{OR}_{\mathrm{D}} \times \mathrm{OR}_{\mathrm{I}}$. The odds ratio of the indirect effect, OR_{I}, measures the increased odds ratio of the disease that is due to the mediational mechanism. If $\mathrm{OR}_{\mathrm{I}} > 1$, then $\mathrm{OR}_{\mathrm{T}} > \mathrm{OR}_{\mathrm{D}}$, that is the increased odds ratio of the disease that is attributable to the mediating effect of Z is measured by $(\mathrm{OR}_{\mathrm{I}} - 1)$. Similarly, if $\mathrm{OR}_{\mathrm{I}} < 1$, then $\mathrm{OR}_{\mathrm{T}} < \mathrm{OR}_{\mathrm{D}}$, i.e. the odds ratio of the disease decreases by $(\mathrm{OR}_{\mathrm{I}} - 1)$ due to the mechanism of mediator. $\mathrm{OR}_{\mathrm{I}} = 1$ indicates null mediation effect, i.e. the effect of X on Y is not at all mediated through Z; while $\mathrm{OR}_{\mathrm{D}} = 1$ indicates a full mediation effect, i.e. the effect of X on Y is completely mediated through Z. If $\mathrm{OR}_{\mathrm{I}} = \mathrm{OR}_{\mathrm{D}}$ then 50 per cent of the TOT is attributable the mediator; If $\mathrm{OR}_{\mathrm{I}} > \mathrm{OR}_{\mathrm{D}}$ then more than 50 per cent of the TOT is attributable the mediator and vise versa.

4. ESTIMATION OF MEDIATIONAL EFFECTS

In this section, we consider three approaches to the estimations of mediational effects for the logistic mediational model (3).

4.1. Delta method

The Delta method [21] is one of the most commonly used methods for approximating standard error. However, considering the extremely complicated functional form, $f(\alpha_0, \alpha_1, \beta_0, \beta_1, \beta_2)$, of the mediational effects here, we have turned to some approximations as given below, (details are given in the Appendix A)

$$\text{logit}(E[Y \mid x]) \approx \beta_0 + \beta_1 P(Z = 1 \mid x) + \beta_2 x$$

$$\text{logit}(E^P[Y \mid x]) \approx \beta_0 + \beta_1 P(Z = 1 \mid x - 1) + \beta_2 x$$

With these approximations, the mediational effects are approximated by

$$\text{DIR} \approx \beta_2, \quad \text{IND} \approx \beta_1(q_1 - q_0) \quad \text{and} \quad \text{TOT} \approx \beta_1(q_1 - q_0) + \beta_2$$

The variance estimates using the Delta method are given below:

$$\text{Var}(\hat{\text{DIR}}) \approx \text{Var}(\hat{\beta}_2)$$

$$\text{Var}(\hat{\text{IND}}) \approx C^2 \,\text{Var}(\hat{\beta}_1) + \beta_1^2[A^2 \,\text{Var}(\hat{\alpha}_0) + 2AB \,\text{Cov}(\hat{\alpha}_0, \hat{\alpha}_1) + B^2 \,\text{Var}(\hat{\alpha}_1)]$$

$$\text{Var}(\hat{\text{TOT}}) \approx \text{Var}(\hat{\text{IND}}) + \text{Var}(\hat{\text{DIR}}) + 2C^2 \,\text{Cov}(\hat{\beta}_1, \hat{\beta}_2)$$

$$\text{Var}(\hat{\text{RIND}}) \approx \frac{\hat{\text{DIR}}^2 \times \text{Var}(\hat{\text{IND}}) + \hat{\text{IND}}^2 \times \text{Var}(\hat{\text{DIR}}) - 2\hat{\text{DIR}} \times \hat{\text{IND}} \times C \times \text{Cov}(\hat{\beta}_1, \hat{\beta}_2)}{\hat{\text{TOT}}^4}$$

where

$$q_0 = \frac{e^{\hat{\alpha}_0}}{(1 + e^{\hat{\alpha}_0})}, \quad q_1 = \frac{e^{\hat{\alpha}_0 + \hat{\alpha}_1}}{(1 + e^{\hat{\alpha}_0 + \hat{\alpha}_1})}, \quad A = q_1 - q_1^2 - q_0 - q_2^2, \qquad B = q_1 - q_1^2$$

and

$$C = q_1 - q_0$$

These asymptotic approximations are reasonable when the prevalence of disease is not too close to 0 or 1 and when the sample size is large.

4.2. Bootstrap estimate

The bootstrap method [22] can be a useful tool for finding standard error of an estimate when the sampling distribution is unknown or when the exact calculation of the estimate is intractable. To obtain the bootstrap estimate, we first randomly draw a bootstrap sample of size N with replacement from the original data of the size N, and then cross-tabulate X, Z by Y and X by Z to obtain bootstrap estimate of p_{ij} and q_i. This allows for the bootstrap estimates of I_{H}, I_{H}^* and I_{L}, and the estimates of the mediational effects. The process is repeated B times ($B = 200$ in this study). The bootstrap estimates of the 95 per cent confidence interval (2.5th and 97.5th percentiles) of the mediational effects were calculated from the 200 draws in our study.

4.3. Bayesian estimate

In Bayesian inference, information about unknown parameters is expressed as a posterior distribution, which combines both *a priori* information and the likelihood of data. Often, when *a priori* information is not available, a non-informative prior is used to obtain Bayesian inference. We will discuss Bayesian estimation using a non-informative prior in this section.

For the logistic mediational model (3), let z_i be the number of subjects with $Z=1$ among the total of m_i subjects with $X=i$, and y_{ij} be the number of subjects with $Y=1$ among the total of n_{ij} subjects with $Z=i$ and $X=j$. Following the notation from Section 3, we have both Y and Z following binomial distributions,

$$Z_i \mid (q_i, m_i) \sim \text{Binomial}(q_i, m_i)$$

and

$$Y_{ij} \mid (p_{ij}, z_i, m_i) \sim \text{Binomial}(p_{ij}, n_{ij} = jz_i + (1-j)(m_i - z_i))$$

The joint probability distribution of Z and Y is proportional to

$$\propto \prod_{i=0}^{1} \left[q_i^{z_i} (1-q_i)^{m_i - z_i} \prod_{j=0}^{1} p_{ij}^{y_{ij}} (1-p_{ij})^{j \cdot z_i + (1-j)(m_i - z_i) - y_{ij}} \right]$$

Given non-informative priors of $q_i \sim \text{Uniform}(0,1)$ and $p_{ij} \sim \text{Uniform}(0,1)$, it follows that the posteriors of p_{ij} and q_i are $q_i \mid \{m,z\} \sim \text{Beta}(z_i+1, m_i - z_i + 1)$ and $p_{ij} \mid \{m,z,y\} \sim \text{Beta}(y_{ij} + 1, jz_i + (1-j)(m_i - z_i) - y_{ij} + 1)$. This allows us to simulate directly from the posterior distributions. We first draw a large sample (10 000 in this study) of p_{ij} and q_i from their posterior distributions, and then construct the posteriors of mediational effects. The point estimate can be obtained either by mean or median. The 95 per cent confidence estimates can be obtained through the upper and lower 2.5th percentiles.

5. SIMULATION STUDIES

We used simulations to examine the behaviours of the Delta, bootstrap and the non-informative prior based Bayesian estimates of the logistic mediational model using the coverage probabilities of the resulting 95 per cent confidence intervals (CI), and the mean square error of the resulting estimates. The mediation effects considered are the direct, indirect, relative indirect and total effects. Simulation of data, model fitting and comparison were done using Mathematica® [23], a mathematics computation platform. For each simulation design, 1000 independent trials were conducted. (When we increased simulation sample to 2000, we found the results remained stable.) We calculated the coverage probability of a 95 per cent CI as the number of samples in which the resulting 95 per cent CI contains the true value of the corresponding parameter, divided by 1000. Similarly, we calculated the average root mean square error (RMSE) over all 1000 repeated simulation samples. In rare cases, the Delta method could not be obtained due to the failure to converge for the maximum likelihood estimation of regression parameters. These cases were excluded, and the rest of the 1000 simulations were used for the calculation of the 95 per cent CI coverage rate and RMSE.

The data sets were simulated using model 3, and the parameters were designed based on the case study; α_0 was fixed at -2 and β_0 was fixed at -3, so that the $P(Z=1 \mid x=0)$ was

approximately equal to 12 per cent and $P(Y=1\,|\,z=0,\ x=0)$ was approximately equal to 5 per cent. X was simulated from Bernoulli distribution with a probability of 0.5. Three sets of parameter settings:

$$\text{Design A:}\quad \{\alpha_0=-2,\ \alpha_1=1,\ \beta_0=-3,\ \beta_1=1,\ \beta_2=1\}$$

$$\text{Design B:}\quad \{\alpha_0=-2,\ \alpha_1=0.5,\ \beta_0=-3,\ \beta_1=2,\ \beta_2=1\}$$

$$\text{Design C:}\quad \{\alpha_0=-2,\ \alpha_1=2,\ \beta_0=-3,\ \beta_1=0.5,\ \beta_2=1\}$$

and three sample sizes, $N=100$, 500 and 1000, were considered.

The simulation results are summarized in Table I. In terms of the 95 per cent CI coverage rate, both the bootstrap and Bayesian estimates are nominal. Although nominal for the TOT and DIR, the 95 per cent CI coverage rates of the Delta method are less than 90 per cent for the RIND and IND under design A and B, which suggests that the variances are underestimated. This may be caused by the loss of precision from the approximation introduced in the asymptotic estimates. In terms of the RMSE, for small sample sizes ($N=100$), the RMSE is smallest for the Bayesian method, intermediate for the bootstrap method and largest for the Delta method. Both the Delta and bootstrap estimates present unusually large RMSE for the small sample size ($N=100$) as the result of the extremely lower cell count found in several simulation runs. In fact, zero probability was found in 23 simulations under design A and size 100. The Delta method failed to obtain an estimate in these simulations, and thus the coverage rate and RMSE are calculated using the remaining 977 simulation results. When the sample size is moderate or large ($N=500, 1000$), while the RMSE of the Bayesian estimates stay the smallest, it tends to be larger for the bootstrap estimates compared to the Delta estimates. In summary, the Bayesian method outperforms both the Delta and bootstrap methods with regard to the 95 per cent CI coverage rate and the RMSE. The bootstrap method may be a better choice than the Delta method for a sample size smaller than 500.

In terms of computational performance, the Delta method requires the least runtime. Comparing the two computationally intensive approaches, the bootstrap method took nearly 45 times the runtime as the Bayesian method, while the latter took only about twice the runtime as the Delta method. In fact, since p and q are conditionally independent and their posteriors are of known distributions, direct simulation from the posterior distributions of p and q is possible. The estimations of the mediational effects are obtained through straightforward Monte Carlo integration in this study.

6. MOTIVATING EXAMPLE: MEDIATIONAL EFFECT OF DEPRESSION TO SOCIOECONOMIC HEALTH DIFFERENTIALS

Self-rated health (SRH) correlates with physical, mental and social health. It is widely accepted as a valid and reliable indicator that predicts the diagnosis of serious illness and the likelihood of survival [24, 25]. As we noted in the background section, the parental SES gradient in SRH in the adolescent population may be mediated through depression, one of the most notable adverse health outcomes of adolescents in the United States. If the SES gradient in SRH is substantially accounted for through the mediating effect of depression, one may

Table I. Simulation study results.

Design			A		B		C	
			Coverage (per cent)	RMSE	Coverage (per cent)	RMSE	Coverage (per cent)	RMSE
100	DIR	E1	91.9	6.21	96.0	4.03	90.5	6.38
		E2	93.7	2.86	96.7	1.80	93.0	2.60
		E3	95.6	0.64	97.1	0.54	96.6	0.65
	IND	E1	83.4	0.57	88.1	0.21	94.5	1.36
		E2	93.9	0.35	98.3	0.34	93.5	0.64
		E3	95.1	0.18	95.6	0.20	96.3	0.30
	RIND	E1	77.4	5.03	86.7	32.86	91.8	8.31
		E2	95.5	38.69	97.0	30.24	97.3	60.64
		E3	98.6	0.17	99.3	0.18	99.1	0.26
	TOT	E1	90.7	6.18	95.1	4.06	91.9	6.58
		E2	93.7	2.88	96.4	1.84	93.1	2.53
		E3	95.3	0.61	96.5	0.55	96.1	0.60
500	DIR	E1	95.3	0.36	95.0	0.35	95.4	0.38
		E2	98.5	0.44	95.3	0.41	96.5	0.56
		E3	95.1	0.32	95.2	0.27	95.8	0.36
	IND	E1	82.2	0.07	85.0	0.08	93.9	0.14
		E2	96.8	0.10	96.6	0.12	96.5	0.22
		E3	95.0	0.08	95.2	0.09	95.6	0.15
	RIND	E1	83.5	0.15	84.4	0.08	94.6	0.16
		E2	98.0	1.29	97.7	1.88	98.8	1.45
		E3	96.2	0.09	96.2	0.09	96.9	0.16
	TOT	E1	94.8	0.36	93.9	0.36	96.7	0.36
		E2	98.5	0.54	93.9	0.43	96.8	0.52
		E3	95.4	0.32	95.0	0.28	94.5	1.14
1000	DIR	E1	95.4	0.24	91.6	0.25	96.0	0.27
		E2	94.1	0.34	96.4	0.27	97.0	0.37
		E3	94.4	0.24	95.7	0.20	94.5	0.27
	IND	E1	84.3	0.05	83.0	0.06	94.9	0.09
		E2	95.6	0.08	94.9	0.09	94.7	0.15
		E3	95.1	0.06	95.0	0.06	95.4	0.10
	RIND	E1	86.9	0.05	79.0	0.06	94.4	0.09
		E2	96.2	0.10	97.6	0.09	96.2	0.16
		E3	94.3	0.07	94.8	0.07	94.5	0.11
	TOT	E1	95.2	0.24	92.7	0.25	96.4	0.25
		E2	94.4	0.33	97.3	0.28	96.8	0.33
		E3	94.6	0.23	94.2	0.20	95.0	0.24

The results are summarized based on 1000 replications. Three estimation methods, E1, E2, E3 correspond to delta, bootstrap and Bayesian estimates, respectively. The bootstrap estimates are based on 200 bootstrap samples. The Bayesian estimates are based on 10 000 MCMC samples. The RMSE for Bayesian RIND estimate were calculated from median estimator.

design an intervention program specifically targeting depressed mood among lower SES teens. There has been little work testing this theory and quantifying the strength of the mediational effect.

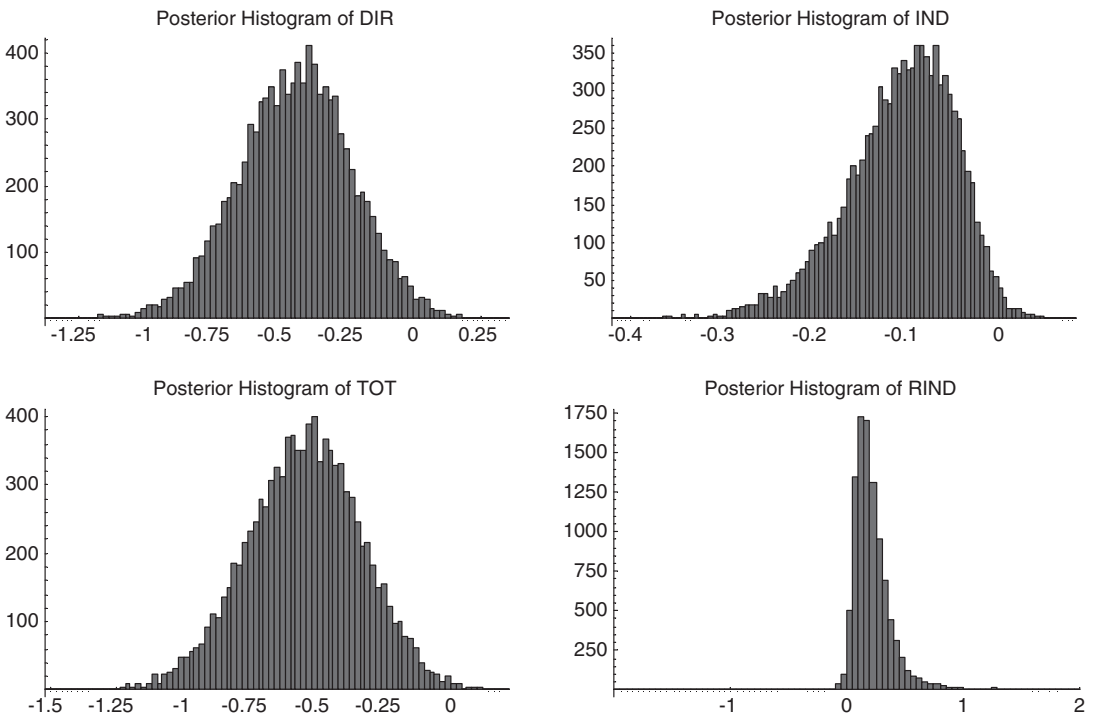

Figure 2. Posterior histograms for the case study.

We used this theory to develop the mediational model of SES (X), depression (Z), and SRH (Y) to test our methodology and applied it to the public use database from the National Longitudinal Study of Adolescent Health (Add Health) [9], which is one of the largest and most recent surveys of adolescents in the U.S. Here, we considered girls only. Depression is measured by the Center for Epidemiological Studies Depression Scale (CES-D), which is dichotomized using the widely applied Robert [26] cutpoint. SRH in Add Health is a five-point Likert scale variable. It is dichotomized at 4 and above, to indicate fair to poor vs good or excellent SRH. Parental education is used here as the indicator for SES. It is dichotomized to indicate whether at least one parent has a college degree or higher education.

Of the total study sample ($N = 2280$), a random subsample ($N = 688$, 30 per cent) was selected and set aside for model validation purposes. The mediational effects were estimated using the remaining data ($N = 1592$). We assessed model fit by comparing predictive posterior distribution with the observed data from the validation sample, using the methods described by Gelfand *et al.* [27]. Results from the analysis suggested good model fit.

WinBUGS [28], a widely used Bayesian computation software, is used for the estimation of posteriors of the direct, indirect, total and relative indirect effects, using a non-informative uniform prior. The posterior distributions are shown in Figure 2 and summarized in Table II. The analyses suggest that the odds of poor self-related general health for the adolescent girls whose parents have less than a college education is 1.7 (OR_T; 95 per cent CI of 1.13–2.55) times to the odds for the adolescent girls from families with at least one college graduated

Table II. Bayesian non-informative posterior estimates obtained from 10 000 MCMC samples for the case study.

	Mean	SD	L95 per cent	Median	U95 per cent
DIR	−0.41	0.207	−0.82	−0.40	−0.02
IND	−0.11	0.060	−0.24	−0.10	−0.01
TOT	−0.52	0.209	−0.94	−0.51	−0.12
RIND	0.21	0.210	0.01	0.20	0.80

parent. This can be explained, in part, by the higher rate of depressive symptoms among adolescent girls whose parents are less well educated: 10 per cent ($OR_I = 1.1$; 95 per cent CI of 1.01–1.28) of increased odds ratio of poor SRH in adolescent girls from less well educated families is attributable to their higher tendency of being depressed. In other words, if these adolescent girls were equally likely to be depressed as those from families with at least one college graduated parent, then the odds ratio of poor self-rated health would be lowered to 1.5 (OR_D; 95 per cent CI of 1.02–2.28).

7. DISCUSSION AND CONCLUSION

The lack of methodological development for the non-linear mediational model has been due to two major barriers: (1) defining and partitioning mediational effects when the association is non-linear, and (2) computation difficulty. In this study, we were able to resolve these barriers by proposing a new definition of mediational effects through causal inference for generalized mediational model; and by taking a Bayesian approach to the estimation of mediational effects. We have demonstrated that the proposed definitions are a general extension of the current concept from linear structural equation model and are intuitively meaningful. Through simulation and case studies, we have shown that the Bayesian approach not only provides good statistical estimates, but also is conceptually and computationally easy to use. The Monte Carlo simulation algorithms used in the Bayesian approach can be easily implemented using either WinBUGS or any computer language such as C, FORTRAN or Mathematica. Since the proposed approximation used in the Delta method is not appropriate when the probability is close to 0 or 1, we would not recommend the use of Delta method in such cases. Also, one should use it with caution when the sample size is small.

We focused our attention here on the logistic mediational model where both the mediator and the outcome are binary variables, for the reason that this is one of the most widely applied models in public health and medical studies. However, it is not necessary to restrict Z as a binary variable. The definition and the Bayesian estimate can be extended to a polytomous type or a continuous type of mediator. When Z is a K-level categorical variable, then I_H, I_L and I_H^* are calculated by summing over all K categories of Z, and $Z \mid X$ are modelled as a multinomial distribution. Given non-informative priors such as Dirichlet distributions (i.e. Jeffery's prior) for p_{ij} and q_i, it follows that both posteriors of p_{ij} and q_i are Dirichlet distributions. In the case of continuous mediator Z, the expected and potential outcomes are calculated from $E[Y \mid x] = E_{Z \mid X}[P(Y = 1 \mid z, x)]$ and $E^P[Y \mid x + 1] = E_{Z \mid X}[P(Y = 1 \mid z, x + 1)]$. With diffused normal distributions as non-informative priors for α and β, the posteriors of

p_{ij} and q_i are not known distributions. One could, however, use Markov Chain Monte Carlo techniques, such as Metropolis–Hastings algorithm, to estimate the mediational effects [29]. This should be easily implemented with WinBUGS. Similarly, one could extend the method into other types of outcome Y. The method could also be extended to consider multiple mediators where $f_2(.)$ is a multivariate distribution. The marginal distribution of $E[Y\,|\,X]$ will be obtained by integrating over the multi-dimensional space of Z, and the definitions of individual and population mediational effects would remain the same. Clearly, further studies are needed to apply the proposed method to other types of mediational model and to extend the method to multiple mediators. More work is also needed to improve the asymptotic estimate.

Throughout this paper, we have implicitly assumed causal association from X to Z or Y, and from Z to Y. Although one may argue that the relationship among an independent variable, mediator, and an outcome may not necessary be 'causal', in many applications, the nature of the mediated relationship is such that the independent variable influences the mediator, which in turn, influences the outcome [30]. This point has been elaborated on by MacKinnon [31]. Thus, the pathway between the independent variable X and the mediator Z can be bi-directional, and the methods proposed in the current study should be useful to such three variables system regardless of the causal assumption.

APPENDIX A

The following notations are used in this appendix:

$$\begin{aligned} I &= P(Y=1\,|\,x) \\ I_a &= \text{Logit}^{-1}(\beta_0+\beta_1 P(Z=1\,|\,x)+\beta_2 x) \\ p_1 &= P(Y=1\,|\,z=1,x) \\ p_0 &= P(Y=1\,|\,z=0,x) \\ q &= P(Z=1\,|\,x) \end{aligned}$$

From the logistic mediational model (3), it is easy to shown that

$$p_1 = \text{Logit}^{-1}(\text{Logit}(p_0)+\beta_1)$$

Thus

$$I = p_1 q + p_0(1-q) = p_0\left(\frac{1}{d_1}\,\mathrm{e}^{\beta_1} q + 1 - q\right)$$

where $d_1 = 1 - p_0 + p_0\mathrm{e}^{\beta_1}$.

Following $\text{Logit}(I_a) = \beta_0 + \beta_1 q + \beta_2 x = \text{Logit}(p_0) + \beta_1 q$, we have

$$I_a = \frac{1}{d_2}\,p_0\mathrm{e}^{\beta_1 q} \quad \text{where } d_2 = 1 - p_0 + p_0\mathrm{e}^{\beta_1 q}$$

Based on the above, we have

$$I - I_a = p_0\left(\frac{\mathrm{e}^{\beta_1 q}}{d_1 + 1 - q} - \frac{\mathrm{e}^{\beta_1 q}}{d_2}\right) \tag{A.1}$$

One can show that

$$I - I_a = O(\beta_1^2)$$

and

$$\sup_{\beta_1} \frac{|I - I_a|}{\beta_1^2} = \frac{1}{48}(3-\sqrt{3})\left(-1+\frac{3-\sqrt{3}}{6}\right)\left(-1+\frac{3-\sqrt{3}}{3}\right) \approx 0.01$$

Thus, approximation (A.1) is bounded and the maximum error can be approximately estimated by $(0.1\beta_1)^2$. The error approaches 0 as one of the conditions meets:

$$(1)\quad q \to 0 \text{ or } 1; \quad (2)\ p_0 \to 0 \text{ or } 1; \quad (3)\ \beta_1 \to 0$$

As $\text{Logit}'(x) = 1/x + 1/(1-x)$ is not bounded at points 0 and 1, the approximation of

$$\text{logit}(E[Y = 1 \,|\, x]) \approx \beta_0 + \beta_1 P(Z = 1 \,|\, x) + \beta_2 x$$

requires that both p_0 and q are bounded away from 0 or 1 with probability 1, i.e. the events of Z and Y have to be existed ones. This assumption is not only a realistic one but also has been assumed by current estimation methods on odds ratio. However, the quality of the approximation becomes worse when the incidence rate approaches 0 or 1. The same proof applies to the approximation of

$$\text{logit}(E^P[Y = 1 \,|\, x]) \approx \beta_0 + \beta_1 P(Z = 1 \,|\, x - 1) + \beta_2 x$$

ACKNOWLEDGEMENTS

The case study uses data from Add Health, a program project designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris, and funded by a grant P01-HD31921 from the National Institute of Child Health and Human Development, with cooperative funding from 17 other agencies. Special acknowledgment is due to Ronald R. Rindfuss and Barbara Entwisle for assistance in the original design. Persons interested in obtaining data files from Add Health should contact Add Health Carolina Population Center, 123 W. Franklin Street, Chapel Hill, NC 27516-2524, U.S.A. (www.cpc.unc.edu/addhealth/contract.html).

REFERENCES

1. Kraemer HC, Stice E, Kazdin A, Offord D, Kupfer D. How do risk factors work together? Mediators, moderators, and independent, overlapping, and proxy risk factors. *American Journal of Psychiatry* 2001; **158**:848–856.
2. Baron RM, Kenny DA. The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology* 1986; **51**:1173–1182.
3. Fishbein M, Ajzen I. *Belief*, *Attitude*, *Intention and Behaviour*. Addison-Wensley: Reading, MA, 1975.
4. MacKinnon DP. Analysis of mediating variables in prevention and intervention research. *NIDA Research Monographs* 1994; **139**:127–153.
5. MacKinnon DP, Taborga MP, Morgan-Lopez AA. Mediation designs for tobacco prevention research. *Drug and Alcohol Dependence* 2002; **68**(Suppl 1):S69–S83.
6. Hansen WB, McNeal RB. The law of maximum expected potential effect: constraints placed on program effectiveness by mediator relationships. *Health Education Research Theory and Practice* 1996; **11**:501–507.
7. Adler NE, Boyce T, Chesney MA *et al.* Socioeconomic status and health. The challenge of the gradient. *American Psychologist* 1994; **49**:15–24.
8. Goodman E. The role of socioeconomic status gradients in explaining differences in US adolescents' health. *American Journal of Public Health* 1999; **89**:1522–1528.

9. Mullan HK, Florey F, Tabor J, Bearman PS, Jones J, Udry JR. The National Longitudinal Study of Adolescent Health: Research Design. URL: http://www.cpc.unc.edu/projects/addhealth/design. 2003.
10. Lee PR, Estes CL. *Healthy People 2010* (6th edn). Jones and Barlett: Sudbury, MA, 2001.
11. Bollen KA. *Structure Equations with Latent Variables*. Wiley: New York, 1989.
12. Sobel ME. Asymptotic confidence intervals for indirect effects in structural equations models. In *Sociological Methodology*, Leinhart S (ed.). Jossey-Bass: San Francisco, 1982; 290–312.
13. MacKinnon DP, Lockwood CM, Hoffman JM, West SG, Sheets V. A comparison of methods to test mediation and other intervening variable effects. *Psychological Methods* 2002; **7**:83–104.
14. MacKinnon DP, Dwyer JH. Estimating mediated effects in prevention studies. *Evaluation Review* 1993; **17**: 144–158.
15. Muthen B. A general structural equation model with dichotomous, ordered categorical and continuous latent variable indicators. *Psychometrika* 1984; **49**:115–132.
16. Freedman LS, Graubard BI, Schatzkin A. Statistical validation of intermediate end-points for chronic diseases. *Statistics in Medicine* 1992; **11**:167–178.
17. De Gruttola VG, Clax P, DeMets DL *et al.* Considerations in the evaluation of surrogate endpoints in clinical trials: summary of a National Institutes of Health Workshop. *Controlled Clinical Trials* 2001; **22**:485–502.
18. Freedman LS. Confidence intervals and statistical power of the 'validation' ratio for surrogate or intermediate endpoints. *Journal of Statistical Planning and Inference* 2002; **100**:443.
19. Alwin DF, Hauser RM. The decomposition of effects in path analysis. *American Sociological Review* 1975; **40**:37–47.
20. McCullagh P, Nelder JA. *Generalized Linear Models* (2nd edn). Chapman & Hall/CRC: New York, 1989.
21. Rao CR. *Linear Statistical Inference and its Applications* (2nd edn). Wiley: New York, 1973.
22. Efron B, Tibshirani RJ. *An Introduction to the Bootstrap*. Chapman & Hall/CRC: Boca Raton, FL, 1998.
23. Wolfram Research. *Mathematica User's Guide*, V 3.0. Wolfram Research, 1996.
24. Kaplan GA, Camacho T. Perceived health and mortality: a nine-year follow-up of the human population laboratory cohort. *American Journal of Epidemiology* 1983; **117**:292–304.
25. Shadbolt B. Some correlates of self-rated health for Australian women. *American Journal of Public Health* 1997; **87**:951–956.
26. Roberts RE, Lewinsohn PM, Seeley JR. Screening for adolescent depression: a comparison of depression scales. *Journal of the American Academy of Child and Adolescent Psychiatry* 1991; **30**:58–66.
27. Gelfand AE, Day DK, Chang H. Model determination using predictive distributions with implementation via sampling-based methods. *Bayesian Statistics* 1992; **4**:147–167.
28. Spiegelhalter D, Thomas A, Best N, Gilks W. *BUGS*: *Bayesian Inference Using Gibbs Sampling*. MRC Biostatistics Unit: Cambridge, 1994.
29. Gilks W, Richardson S, Spiegelhalter D. *Markov Chain Monte Carlo in Practice.* Chapman & Hall/CRC: New York, 2000.
30. Holmbeck GN. Toward terminological, conceptual, and statistical clarity in the study of mediators and moderators: examples from the child-clinical and pediatric psychology literatures. *Journal of Consulting and Clinical Psychology* 1997; **65**:599–610.
31. MacKinnon DP, Krull JL, Lockwood CM. Equivalence of the mediation, confounding and suppression effect. *Prevention Science* 2000; **1**:173–181.

