Journal of the Korean Data & Information Science Society
2017, **28**(3), 669–684

http://dx.doi.org/10.7465/jkdi.2017.28.3.669
한국데이터정보과학회지

Simple power analysis in causal mediation models for a dichotomous outcome based on the mediation proportion

Young Min Kim[1] · John Bennett Cologne[2] · Harry Michael Cullings[3]

[1]Department of Statistics, Kyungpook National University
[23]Department of Statistics, Radiation Effects Research Foundation

Received 16 March 2017, revised 1 May 2017, accepted 8 May 2017

Abstract

Mediation models are widely used in many fields of research and have recently gained attention in epidemiology. The mediation proportion is a standard measure to evaluate what part of the total exposure effect on an outcome may be explained by a particular mediator and to examine how important that pathway is relative to the overall exposure effect. A common question is how large a sample size is needed to achieve high statistical power or, equivalently, what magnitude of effect can be detected. Current power and sample size calculations for mediation analysis are limited and additional research is needed. We therefore propose a computer-intensive power analysis using the mediation proportion. We conduct simulation studies to calculate statistical powers and sample sizes. And then, we illustrate our power analysis using an example from the Adult Health Study of atomic-bomb survivors and demonstrate that the method is relatively straightforward to understand and compute.

Keywords: Bias-corrected bootstrap, dichotomous outcomes, mediation proportion, natural direct and indirect effects, power analysis.

1. Introduction

Statistical mediation analysis is an important tool in epidemiology (Hafeman and Schwartz, 2009; Richiardi *et al.*, 2013; VanderWeele, 2015) and the social and behavioral sciences (MacKinnon, 2008) because it allows exploring complex causal mechanisms where the total effect of an exposure on an outcome includes a pathway via a mediator. For example, Hong *et al.* (2016) investigated the intermediate effect of depression on the relationship between self-esteem and life satisfaction of Korean children under low socio-economic status and

[1] Corresponding author: Assistant Professor, Department of Statistics, Kyungpook National University, Daegu 41566, Korea. E-mail: kymmyself@knu.ac.kr
[2] Senior Researcher, Department of Statistics, Radiation Effects Research Foundation, Hiroshima 732–0815, Japan.
[3] Department of Chief, Department of Statistics, Radiation Effects Research Foundation, Hiroshima 732–0815, Japan.

Jeong (2017) also conducted causality study for economics to examine its quantile causal association from dollar exchange rate to international oil price. Traditionally, a mediated effect could be quantified using the indirect effect of an exposure working through a specific mediator with a path model (Baron and Kenny, 1986). Another effect measure useful for quantifying the mediating effect is the "mediation proportion" (MP), defined by the proportion of the total effect of an exposure on an outcome that is explained by a mediated effect. Freedman *et al.* (1992) proposed a statistical method to validate a casual mechanism with a dichotomous outcome using outcome-mediator-exposure (OME) and outcome-exposure (OE) models, disallowing a product term between a mediator and an exposure. Lynch et al. (1996) compared differences between outcome-mediator-exposure and outcome-exposure models in a study of acute myocardial infarction using the Cox proportional hazard model to assess the impact of risk factor adjustments on the age-adjusted relative hazard using the proportion of excess relative risk. Ditlevsen *et al.* (2005) reviewed estimation of the proportion of an exposure effect on an outcome explained by a mediator in the context of observational epidemiology studies. Tsiatis *et al.* (1995) studied the relationship between survival and longitudinal CD4 counts as a surrogate variable to assess the proportion of a treatment effect explained. Wang and Taylor (2002) investigated an alternative measure of the MP that adapted an idea in Tsiatis *et al.* (1995), requiring fewer assumptions in estimation and allowing more flexibility in modeling. VanderWeele and Vansteelandt (2010) discussed mediation analysis for a dichotomous outcome using odds ratios (ORs) on the risk difference scale and OR scale by combining logistic and linear regressions when a mediator is continuous. Valeri and VanderWeele (2013) developed an analogous mediation analysis for dichotomous mediators. Despite this wealth of background on mediated effects per se in a wide variety of epidemiological settings, there is little guidance for researchers in terms of designing studies or interpreting results in regards to sample size and power specifically aimed at the MP as the effect measure of interest considering all relationships − exposure-mediator-outcome, exposure-mediator, and a mediator-outcome − as well as various forms of exposure distribution.

Statistical power, the probability of rejecting a false null hypothesis (1−probability of type II error), is a critical component of research (Cohen, 1988; Sedlmaier and Gigerenzer, 1989). Power analysis in mediation models was explored by Fritz and MacKinnon (2007), who showed that sample size requirements can be very large, especially if a small mediated effect is to be detected. To increase statistical power, they suggested using asymmetric confidence intervals (CIs) from resampling methods (e.g., the bootstrap) to test the mediated effect. Preacher and Kelley (2011) investigated effect size measures using the MP from linear outcome-mediator-exposure and mediator- exposure models. Preacher and Selig (2012) studied Monte Carlo (MC) simulation methodologies to compute CIs for the mediator and Zhang (2014) proposed MC-based statistical power analysis for various mediation models such as a simple mediation model, a multiple mediator model with a latent mediator, a multiple-group mediation model, and a longitudinal mediation model. Vittinghoff *et al.* (2009) proposed a simple, fast sample size calculation method based on a null hypothesis of no coefficient in an outcome-mediator-exposure model for the linear, logistic, and Poisson models, and their simulation study evidenced less than adequate statistical power of their calculated sample size in small samples, although they acknowledged that being based on Wald statistics the method would be expected to be less reliable in small samples assuming that the exposure has an effect on the mediator. Wang and Xue (2012) indicated that the

method of Vittinghoff *et al.'* (2009) provides relative small sample sizes because they dealt with the indirect pathway ignoring the relation between exposure and mediator. Vittinghoff and Neilands (2015) revised its problem, providing fast R programs to estimate power of the joint test addressing the issue pointed out by Want and Xue (2012). Moreover, Vittinghoff and Neilands (2015) proposed a proper sample size calculation providing a pre-specified power even though the normal distribution for an exposure was assumed.

In this paper, we propose a strategy to compute statistical power and sample size for detecting a continuous mediated effect on a dichotomous outcome using the MP on the log-OR or OR scale based on statistical approaches described by VanderWeele and Vansteelandt (2010) under a null hypothesis of $H_0 : \text{MP} = h_0$ where $h_0 \in [0, 1]$. The method is particularly useful for applications in epidemiology, where logistic regression is frequently used and h_0 is often close to 0, unlike in many economic, social-science, and psychometric applications. Within the procedure, which loops through input values of $h_1 (\neq h_0)$, we use the bootstrap method to compute CIs to test the null hypothesis against an alternative hypothesis of $H_1 : \text{MP} = h_1$. In Section 2, we review the MP measures on the log-OR scale, in which case they are the ratio of the log-OR of a natural indirect effect and the log-OR of a total effect with a continuous mediator. We propose the power analysis procedure at the end of Section 2. A simulation study is presented in Section 3. Relationships among the effect measure (MP), the probabilities of outcome, mediator and exposure, sample size, and statistical power are also investigated in Section 3. We apply the proposed power analysis method to data on heart disease in Section 4. We interpret the findings and summarize our conclusions in Section 5.

2. Power analysis using the mediation proportion

Let (X, M, Y) denote an exposure of interest, a potential mediator, and an outcome, respectively. Let C denote baseline covariates, not affected by the exposure. The causal diagram for these variables in Figure 2.1 describes the role of the mediator. To address the mechanism in Figure 2.1 and various questions concerning mediation analysis, the counterfactual framework has been considered (Rubin, 1990; Robins and Greenland, 1992; Pearl, 2001; Hernan, 2004). Define Y_x and M_x, respectively, as the values of an outcome and a mediator that would be realized if $X = x$, possibly contrary to fact. Let Y_{xm} denote the value of an outcome that would have been observed if X and M set to x and m, also possibly contrary to fact, respectively. To make inference about the MP by applying counterfactual approaches, the following assumptions must hold:

Assumptions:

A1) There is no unmeasured confounding for the total effect, (i.e., given covariates $C = c$, the two random variables Y_{xm} and X, are independent for all x).

A2) Given an exposure $X = x$ and covariates $C = c$, there is no unmeasured confounding in the outcome-mediator relation, (i.e., given an exposure $X = x$ and covariates $C = c$, the two random variables M and Y_{xm}, are independent for all x).

A3) Given covariates $C = c$, there is no unmeasured confounding in the mediator-exposure relation, (i.e., given covariates $C = c$, the two random variables M_x and X, are independent for all x).

A4) Given covariates $C = c$, the two random variables Y_{xm} and M_{x^*}, are independent for all x, $x = x^*$, $x \neq x^*$ and m.

Remark A1)-A3) state no unmeasured confounding of the exposure-outcome, mediator-outcome, or exposure-mediator relationships and A4) states the absence of variables influenced by exposure which also confound the mediator-outcome relationship.

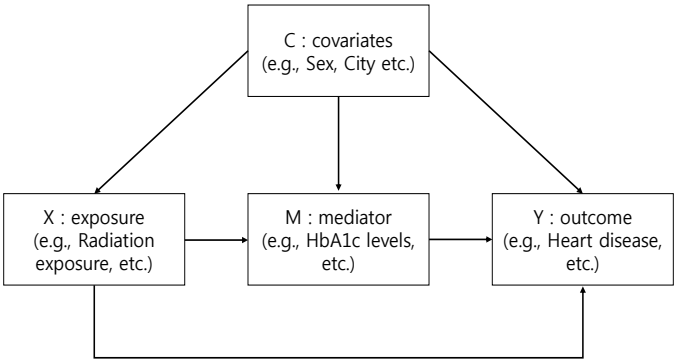

Figure 2.1 (Example of a simple diagram for a mediation model) In this diagram, radiation exposure (X) causes incident heart disease (Y) via HbA1c levels (M) considering sex and city (C : covariates). It means that HbA1c levels is a mediator from people who exposed to radiation to the incident probabilities of heart disease for those people.

2.1. The MP on the log-OR scale for a continuous mediator

Suppose that A1)-A4) hold. We describe a total effect (TE) comparing exposure levels x and x^* on the OR scale given covariates $C = c$ as

$$\mathrm{OR}^{\mathrm{TE}}_{x,x^*|c} = \frac{\mathrm{P}\left(Y_x = 1|c\right) / \left\{1 - \mathrm{P}\left(Y_x = 1|c\right)\right\}}{\mathrm{P}\left(Y_{x^*} = 1|c\right) / \left\{1 - \mathrm{P}\left(Y_{x^*} = 1|c\right)\right\}}, \tag{2.1}$$

which is the ratio of odds for outcome $Y = 1$ given $X = x$ and that given $X = x^*$, given a set of adjustment covariates $C = c$. The natural direct effect (NDE) on the OR scale is the ratio of odds of the outcome, Y, if the exposure had been to x, M took on the value it naturally would take if X were set to be x^* to odds of outcome if the exposure had been to x^* and M took on the value it naturally would take if X were set to be x^* as the same M_{x^*}, conditional on covariate $C = c$. It will usually vary across participants. In other words, given definition of the natural direct and indirect effects, the mediator M is not fixed at any given value and the only X is set in this context. The NDE on the OR scale, $\mathrm{OR}^{\mathrm{NDE}}_{x,x^*|c}$, can be denoted analogously as

$$\mathrm{OR}^{\mathrm{NDE}}_{x,x^*|c} = \frac{\mathrm{P}\left(Y_{xM_{x^*}} = 1|c\right) / \left\{1 - \mathrm{P}\left(Y_{xM_{x^*}} = 1|c\right)\right\}}{\mathrm{P}\left(Y_{x^*M_{x^*}} = 1|c\right) / \left\{1 - \mathrm{P}\left(Y_{x^*M_{x^*}} = 1|c\right)\right\}}. \tag{2.2}$$

Given $C = c$, $\mathrm{OR}^{\mathrm{NDE}}_{x,x^*|c}$ can be interpreted as a ratio of odds for outcomes, $Y_{xM_{x^*}}$ and $Y_{x^*M_{x^*}}$. Likewise, the natural indirect effect (NIE) on the OR scale, $\mathrm{OR}^{\mathrm{NIE}}_{x,x^*|c}$, can be denoted as

$$\mathrm{OR}^{\mathrm{NIE}}_{x,x^*|c} = \frac{\mathrm{P}\left(Y_{xM_x} = 1|c\right) / \left\{1 - \mathrm{P}\left(Y_{xM_x} = 1|c\right)\right\}}{\mathrm{P}\left(Y_{xM_{x^*}} = 1|c\right) / \left\{1 - \mathrm{P}\left(Y_{xM_{x^*}} = 1|c\right)\right\}}. \tag{2.3}$$

Given $C = c$, $\mathrm{OR}^{\mathrm{NIE}}_{x,x^*|c}$ can be interpreted as a ratio of odds for outcomes, Y_{xM} and $Y_{xM_{x^*}}$. The formula, (2.1), decomposes into a product of (2.2) and (2.3) as $\mathrm{OR}^{\mathrm{TE}}_{x,x^*|c} = \mathrm{OR}^{\mathrm{NDE}}_{x,x^*|c} \times \mathrm{OR}^{\mathrm{NIE}}_{x,x^*|c}$. Taking the logarithm of both sides in the previous equation, we have $\log(\mathrm{OR}^{\mathrm{TE}}_{x,x^*|c}) = \log(\mathrm{OR}^{\mathrm{NDE}}_{x,x^*|c}) + \log(\mathrm{OR}^{\mathrm{NIE}}_{x,x^*|c})$ on the log-OR scale.

To make the problem brief, we assume that there is no product term between the exposure of interest and the mediator in their effects on the outcome. Thus, the logistic OME model takes the following form

$$\log\left(\frac{\mathrm{P}\left(Y_{xM_x} = 1|c\right)}{1 - \mathrm{P}\left(Y_{xM_x} = 1|c\right)}\right) = \beta_0 + \beta_1 x + \beta_2 M_x + \beta_4^T c, \tag{2.4}$$

and the linear ME model is defined by

$$\mathrm{E}\left[M_x|c\right] = \theta_0 + \theta_1 x + \theta_2^T c. \tag{2.5}$$

Then (2.4) and (2.5) and A1)-A4) imply that, for a rare outcome, approximately

$$\mathrm{OR}^{\mathrm{NDE}}_{x,x^*|c} \approx \exp\left\{\beta_1(x - x^*)\right\} \text{ and } \mathrm{OR}^{\mathrm{NIE}}_{x,x^*|c} \approx \exp\left\{\theta_1\beta_2(x - x^*)\right\} \tag{2.6}$$

which do not depend on any baseline covariates (see VanderWeele and Vansteelandt, 2010), when the outcome is rare. The MP on the log-OR scale is the ratio of $\log\left(\mathrm{OR}^{\mathrm{NIE}}_{x,x^*|c}\right)$ and $\log\left(\mathrm{OR}^{\mathrm{TE}}_{x,x^*|c}\right)$ as

$$\mathrm{MP} \equiv \frac{\log\left(\mathrm{OR}^{\mathrm{NIE}}_{x,x^*|c}\right)}{\log\left(\mathrm{OR}^{\mathrm{TE}}_{x,x^*|c}\right)} = \frac{\log\left(\mathrm{OR}^{\mathrm{NIE}}_{x,x^*|c}\right)}{\log\left(\mathrm{OR}^{\mathrm{NIE}}_{x,x^*|c}\right) + \log\left(\mathrm{OR}^{\mathrm{NDE}}_{x,x^*|c}\right)} \tag{2.7}$$

which constitutes a measure of the proportion of the effect of the exposure on an outcome that is explained by the mediator on the log-OR scale. (2.4), (2.5) and (2.6) implies that the MP for a continuous mediator on the log-OR scale is approximately

$$\mathrm{MP} \approx \frac{\theta_1\beta_2}{\theta_1\beta_2 + \beta_1}. \tag{2.8}$$

2.2. MC based power analysis using the resampling method

To conduct power analysis based on the MP on the log-OR scale, we propose a standard resampling method to compute CIs according to sample size using MC simulated samples. The null hypothesis is that the MP is equal to some h_0 (i.e., $H_0 : \mathrm{MP} = h_0$; usually, $h_0 = 0$). To investigate the bootstrap method was appropriate for the power analysis, we conducted a small simulation study to compare performance of percentile and bias-corrected bootstrap methods for the MP with 1,000 MC simulation runs and 1,000 bootstrap replicates. We

calculated empirical coverage probabilities for one- and two-sided 95% confidence intervals for the MP with a continuous mediator considering the true MPs (0.11, 0.31, 0.50), 3 sample sizes (100, 250, 500), and probabilities of event occurrence (approximately 0.1, 0.3, 0.5).

Table 2.1 Empirical coverage probabilities (1CP and 2CP) for one- and two-sided 95% confidence intervals for the MP using percentile and bias-corrected bootstrap methods, considering sample size=100, 250, 500, the true MP=0.11, 0.31, 0.50, and Monte Carlo simulated probabilities of event occurrence (PY)=0.10, 0.30, 0.50 when a mediator is continuous. PerBoot and BCBoot are represented as the percentile and bias-corrected bootstraps, respectively.

True			0.11					0.31					0.50		
MP		PerBoot		BCBoot			PerBoot		BCBoot			PerBoot		BCBoot	
n	PY	2CP	1CP	2CP	1CP	PY	2CP	1CP	2CP	1CP	PY	2CP	1CP	2CP	1CP
100	0.10	0.99	0.99	0.93	0.93	0.10	0.99	0.98	0.93	0.93	0.10	0.97	0.99	0.93	0.95
250	0.10	0.99	0.98	0.95	0.95	0.10	0.98	0.98	0.94	0.94	0.10	0.96	0.97	0.94	0.95
500	0.10	0.98	0.97	0.95	0.95	0.10	0.97	0.96	0.95	0.95	0.10	0.97	0.96	0.95	0.95
100	0.30	1.00	1.00	0.95	0.96	0.31	1.00	1.00	0.95	0.95	0.30	0.98	0.99	0.94	0.94
250	0.30	1.00	0.99	0.94	0.95	0.31	0.98	0.99	0.94	0.95	0.30	0.97	0.98	0.95	0.95
500	0.30	0.99	0.97	0.94	0.94	0.31	0.98	0.97	0.95	0.95	0.30	0.98	0.97	0.96	0.95
100	0.50	1.00	1.00	0.94	0.93	0.50	1.00	1.00	0.94	0.93	0.51	0.98	0.99	0.92	0.93
250	0.50	1.00	0.99	0.95	0.95	0.50	0.99	0.99	0.94	0.95	0.50	0.97	0.99	0.93	0.94
500	0.50	0.99	0.98	0.95	0.95	0.50	0.97	0.98	0.95	0.95	0.51	0.96	0.98	0.95	0.94

Table 2.1 shows that the bias-corrected bootstrap has better performance than the percentile bootstrap based on empirical coverage probabilities for both one- and two-sided 95% confidence intervals for the MP. Thus, the bias-corrected bootstrap was used for this power analysis strategy. Since the observed MP may have a non-symmetric distribution in finite samples, Table 1 shows that the bias-corrected bootstrap method can be useful (Davison and Hinkley, 1997; Carpenter and Bithell, 2000). Sometimes, the bias-corrected and accelerated bootstrap method can also be used (see Davison and Hinkley, 1997; Carpenter and Bithell, 2000).

Power analysis procedure :

0. Assume the true MP, and probability of event occurrence.
1. Generate a sample with size n (i.e., "original" data) including an exposure, a mediator, an outcome and baseline covariates
2. Estimate the MP on the log-OR scale.
3. Test significance of the MP by forming a bootstrap CI.

   Bootstrap CI sub-procedure :

   a. Randomly select a bootstrap sample from the original data (generated in Step 1) with replacement.
   b. Compute a bootstrap MP estimate on the log-OR scale using the bootstrap sample from Step 3a.
   c. Repeat step 3a-3b for a total of B times.
   d. Compute the CI estimate using the B bootstrap MP estimates.

4. Repeat Steps 1-3 for T MC simulation runs.

5. Compute statistical power as $100 \times t/T\%$, where t is the number of tests that reject the null hypothesis in Step 3.

The power analysis procedure can be applied to any other mediation models when the MP is regarded as the important parameter of interest. When you assume the true MP, and the probability of event occurrence, the values may be ascertained from previous studies in the literature or data from pilot studies, or may be hypothesized by the investigator.

3. Numerical Studies

In this section we investigate the performance of our power analysis strategy for dichotomous outcomes. We examined statistical power given sample size for testing the null hypothesis, $H_0 : \text{MP} = 0$ (vs the alternative hypothesis, $H_1 : \text{MP} \neq 0$ or $H_1 : \text{MP} > 0$) on the log-OR scale using two approaches: 1) one MP (MPO) when the denominator of the MP is the total effect Only from an OE model and 2) the other MP (MPS) when the denominator is the Sum of natural direct and indirect effects from OME and ME models. Theoretically, both MPs (MPO and MPS) are equal. The mediation model we considered is

$$M = -0.3 + 0.7X + \varepsilon_M \text{ and } PY = \{1 + \exp(-\beta_0 - 0.35X - \beta_2 M)\}^{-1}, \qquad (3.1)$$

where ε_M is a standard normal variable. The exposure, X, was generated from a lognormal distribution with mean and standard deviation of the distribution on the log-scale with values of 0 and 0.5, respectively. Parameter values of β_0 and β_2 were selected corresponding to the true values of the MP and probabilities of event occurrence via (3.1). We considered the true MP values (0.10, 0.30, 0.50), and probabilities of event occurrence (PY=0.1, 0.3, 0.5). The considered sample sizes are 100, 500, 1000. We do not report the simulation results for sample sizes of 100 and 1000 because they were similar to those for sample size of 500. Each MC simulation had 5000 runs for each combination of the MP and the PY.

Table 3.1 Empirical MP estimates with a total effect only (MPO) and the sum of natural direct and indirect effects (MPS) as a denominator on the log-odds ratio scale considering the true MP=0.1, 0.3, 0.5 and probabilities of even occurrence =0.1, 0.3, 0.5 when sample size is 500 when a mediator is continuous. 1^{st} Qu and 3^{rd} Qu are represented as the 1st and 3rd quantiles, respectively.

True MP	PY	MPS 1^{st} Qu	MPS median	MPS 3^{rd} Qu	MPO 1^{st} Qu	MPO median	MPO 3^{rd} Qu
0.1	0.1	-0.0868	0.0952	0.2936	-0.0874	0.0946	0.2944
	0.3	-0.0102	0.1053	0.2355	-0.0099	0.1055	0.2371
	0.5	-0.0058	0.0999	0.2238	-0.0059	0.0997	0.2246
0.3	0.1	0.1726	0.3061	0.4829	0.1736	0.3101	0.4932
	0.3	0.1974	0.2975	0.4309	0.1994	0.3026	0.4427
	0.5	0.2048	0.3035	0.4298	0.2070	0.3087	0.4402
0.5	0.1	0.3720	0.4993	0.6756	0.3810	0.5190	0.7082
	0.3	0.4023	0.5001	0.6179	0.4216	0.5285	0.6584
	0.5	0.4004	0.4920	0.6108	0.4207	0.5219	0.6530

To examine empirical distributions of the MPS and the MPO in finite samples, median and interquartile are provided in Table 3.1. Table 3.1 shows that in comparison to a smaller

true MP (e.g., MP=0.1), when the true MP is large (e.g., MP=0.5), the empirical MPS can be more unbiased than the empirical MPO in finite samples. On the other hand, when the true MP is not large, the empirical MPS and MPO show similar results, especially for large samples. Moreover, as the PY decreases, variability and magnitude of left-skewedness of the empirical MP increase.

Table 3.2 Empirical MP estimates with a total effect only (MPO) and the sum of natural direct and indirect effects (MPS) as a denominator on the log-odds ratio scale considering the true MP=0.1, 0.3, 0.5 and probabilities of even occurrence =0.1, 0.3, 0.5 when sample size is 500 when a mediator is continuous. Po1 and Po2 are represented as statistical powers for MPS and VM, and MPS, respectively.

		One-sided test					Two-sided test				
True MP	PY	Po1 (%)	MPS (N)	VM (N)	Po2 (%)	MPO (N)	Po1 (%)	MPS (N)	VM (N)	Po2 (%)	MPO (N)
0.3	0.1	81.8	1,000	1157	80.8	3,900	85.4	1,400	1,613	81.7	5,500
	0.3	81.4	500	490	81.8	2,000	84.4	700	672	84.0	3,000
	0.5	86.5	500	483	82.0	1,800	83.0	600	543	82.7	2,600
0.5	0.1	80.7	350	280	81.3	1,400	82.1	550	368	82.5	2,800
	0.3	84.7	160	136	84.6	600	81.9	250	157	80.1	1,100
	0.5	82.4	120	106	83.1	500	83.1	220	136	82.4	1,000

Table 3.3 Empirical statistical powers (power, %) and required sample sizes (N) to achieve statistical power over 80% for 0.05 level one- and two-sided conference interval for the MP for a continuous mediator using the proposed method with the MPS, and the method of Vittinghoff *et al.* (VM, 2009) when an exposure generates from various distributions and ME and OME models are defined as $M = -0.3 + 0.7X + \varepsilon_M$ with standard normal errors, ε_M, of mean of 0 and variance of 1 and the Bernoulli distribution with the probability, $PY = \{1 + \exp(-\beta_0 - 0.35X - \beta_2 M)\}^{-1}$. Po is represented as statistical power.

Exposure	True	PY	One-sided test			Two-sided test		
(X)	MP		Po(%)	MPS(N)	VM(N)	Po(%)	MPS(N)	VM(N)
N(1,0.52)	0.3	0.1	83.5	1,200	1219	85.4	1,800	1614
		0.3	80.5	500	478	84.4	800	637
		0.5	82.7	450	428	82.4	650	535
	0.5	0.1	84.2	400	311	85.1	800	400
		0.3	82.3	150	126	80.1	300	150
		0.5	80.1	120	99	81.7	260	131
N(1,12)	0.3	0.1	83.1	1,100	1203	81.1	1,300	1434
		0.3	80.9	500	483	84.0	700	664
		0.5	84.6	500	454	83.6	600	552
	0.5	0.1	80.4	300	278	85.0	450	399
		0.3	81.1	140	121	85.3	200	172
		0.5	81.5	120	103	80.8	160	128
N(1,22)	0.3	0.1	81.8	1,000	1158	81.3	1,200	1442
		0.3	83.4	600	521	80.6	700	607
		0.5	81.4	500	412	82.0	650	529
	0.5	0.1	84.4	300	313	81.4	350	362
		0.3	81.3	160	122	80.2	200	150
		0.5	84.4	160	113	80.8	180	128

Table 3.4 Empirical statistical powers (Po,%) and required sample sizes (N) to achieve statistical power over 80% for 0.05 level one- and two-sided conference interval for the MP for a continuous mediator using the proposed method with the MPS, and the method of Vittinghoff *et al.* (VM, 2009) when an exposure generates from various distribution and ME and OME models are defined as $M = -1 + 2X + \varepsilon_M$ with standard normal errors, ε_M, of mean of 0 and variance of 1 and the Bernoulli distribution with the probability, $PY = \{1 + \exp(-\beta_0 - X - \beta_2 M)\}^{-1}$. Po is represented as statistical power.

Exposure	True	PY	One-sided test			two-sided test		
(X)	MP		Po(%)	MPS(N)	VM(N)	Po(%)	MPS(N)	VM(N)
Lognormal	0.3	0.1	83.7	2,000	1916	81.8	2,400	2,283
(0,0.5)		0.3	80.8	800	753	80.7	1,000	951
		0.5	83.3	700	681	80.7	800	799
	0.5	0.1	82.1	400	292	81.5	500	362
		0.3	80.5	160	119	84.4	220	168
		0.5	83.4	140	109	84.5	180	142
N(1,0.25)	0.3	0.1	82.2	1,800	1831	82.6	2,200	2,332
		0.3	83.1	800	807	81.6	1,000	975
		0.5	83.5	700	687	80.4	900	794
	0.5	0.1	83.4	350	304	80.3	450	352
		0.3	83.0	160	129	85.1	200	172
		0.5	83.6	140	110	80.9	180	129
N(1,1)	0.3	0.1	82.1	2,000	1826	80.9	2,600	2,231
		0.3	83.1	1,000	807	81.4	1,200	969
		0.5	80.9	800	635	81.0	1,000	805
	0.5	0.1	80.4	400	278	84.2	500	390
		0.3	81.4	200	123	83.4	280	164
		0.5	81.6	180	104	83.8	220	139
N(1,4)	0.3	0.1	80.3	2,800	1732	81.9	3,600	2,291
		0.3	80.2	1,350	740	82.3	1,800	992
		0.5	80.7	1,200	631	84.9	1,500	895
	0.5	0.1	82.1	700	292	84.0	900	388
		0.3	83.1	350	129	83.8	450	165
		0.5	82.6	300	107	84.4	400	141

We also conducted power analyses with the MPS and the MPO, and the method of Vittinghoff *et al.* (VM, 2009), which is a sample-size formula, (3.2), with the same simulation setting. The simulation result is shown in Table 3.2. Surprisingly, when the MPO is used as an effect measure, the required sample size to achieve statistical power over 80% is relatively huge in comparison to that using the MPS and the VM. This result supports the research of Kenny and Judd (2014), who reported that statistical power for the test of the total effect can be much smaller than that for the test of the indirect effect when there is no direct effect, that is to say, when indirect and total effects are the same, the test based on the indirect effect is more powerful. It implies that in finite samples, the bootstrap method is sensitive to estimation of the MPO. In addition, it also implies that mediation analysis based on OME and OE models (the classical Baron and Kenny approach) cannot have better performance than that based on the structural equation approach (OME and ME models) in finite samples using the bootstrap method. Hereafter, we provide the simulation results with the MPS in this manuscript. The explanation for the VM is deferred to Section 3.1 because there are more simulation results. For comparison to the proposed power analysis strategy, we consider the VM approach, which is a power and sample-size formula for use with linear, logistic, Poisson and Cox models, depending on distributions of outcomes for an OME model, assuming a linear ME model to test the null hypothesis, $H_0 : \beta_2 = 0$ (vs

the alternative hypothesis, $H_1 : \beta_2 \neq 0$), not $H_0 : \mathrm{MP} = 0$). Vittinghoff *et al.* (2009) stated that the VM was not always accurate in small sample sizes. The VM formula for sample size calculation with a dichotomous outcome and a continuous mediator is

$$n = \frac{(z_\alpha + z_\gamma)^2}{(\beta_2 \sigma_M)^2 (1 - \rho^2) p(1-p)}, \tag{3.2}$$

where ρ is the correlation of an exposure and a mediator, β_2 is a coefficient of a mediator on the OME model, σ_2 is a standard deviation of ε_M, p is the probability of outcome, and z_α and z_γ are the quantiles of the standard normal distribution corresponding to the specified one-sided type I and II error rates. If the two-sided test is applied, $\alpha/2$ is used instead of α.

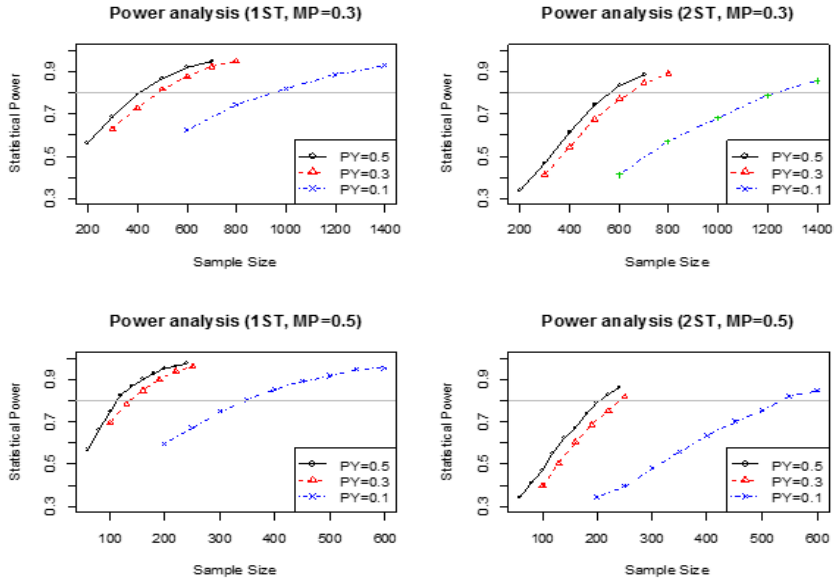

Figure 3.1 Plots of the required sample sizes and statistical powers to detect mediated effects using the MP (the true MP=0.3, 0.5) based on the setup of Table 3.4 when an outcome is dichotomous, a mediator is continuous and an exposure is randomly generated from the lognormal distribution with LogN(0,0.5). 0.05 level one- and two-sided tests (1ST, 2ST) are considered.

We investigated the power analysis strategy with four different exposure distributions including the lognormal distribution used in Table 3.2 and various normal distributions with means and standard deviations of N(1,0.5), N(1,1) and N(1,2) in Tables 3.3 and 3.4. We also considered the true MP of 0.30 and 0.50 and probabilities of target outcomes of 0.1, 0.3, and 0.5. To obtain the numerical results in Tables 3.3 and 3.4, the simulation setup was as follows: the exposure variable, X, was generated by the above various distributions, a continuous mediator, M, was simulated by the linear ME model, $M = -0.3 + 0.7X + \varepsilon_M$ with standard normal errors, ε_M, of mean of 0 and variance of 1 and an outcome variable, Y, was generated by the Bernoulli distribution with the probability, $PY = \{1 + \exp(-\beta_0 - 0.35X - \beta_2 M)\}^{-1}$. The parameter values were chosen as $\beta_2 = 0.50$ and 0.25 to achieve the true MP of 0.3 and 0.5, respectively and β_0 was also variously selected to achieve a proper probability of outcomes using 5,000,000 samples. We also employed other parameter values such as

$M = -1 + 2X + \varepsilon_M$ and $PY = \{1 + \exp(-\beta_0 - X - \beta_2 M)\}^{-1}$ with parameter values of $\beta_2 = 0.50$ and 0.20 for the true MP of 0.3 and 0.5, respectively, and probabilities of outcomes of 0.1, 0.3, and 0.5 in Table 3.4. Statistical powers and required sample sizes to achieve the power equal to or greater than 80% are presented for 0.05 level one- and two-sided tests based on the MPS and the VM in Tables 3.2-4. Figures 3.1-4 illustrate power curves along with sample sizes corresponding to the simulation results with the MPS in Tables 3.2-4. We do not report the figures for power curves for the simulation results in Table 3.4 because all of results have similar patterns to Figures 3.1-3.4.

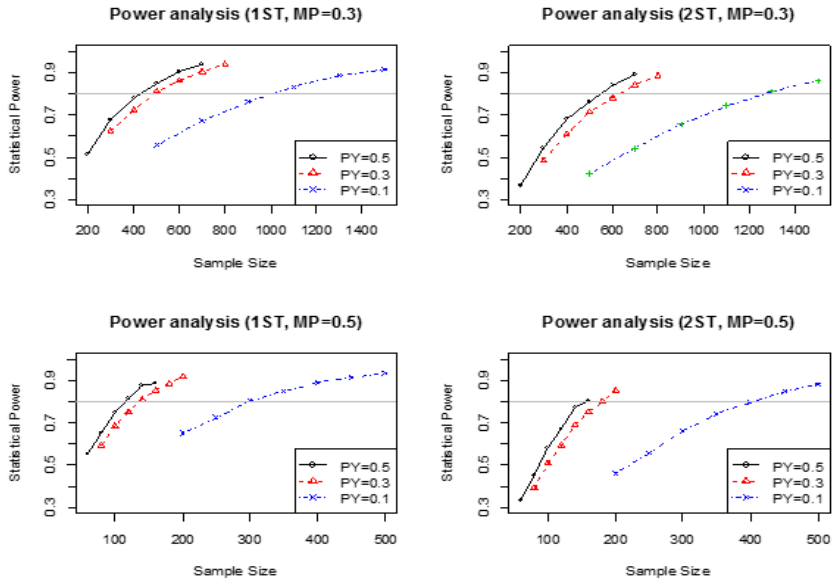

Figure 3.2 Plots of the required sample sizes and statistical powers to detect mediated effects using the MP (the true MP=0.3, 0.5) based on the setup of Table 3.4 when an outcome is dichotomous, a mediator is continuous and an exposure is randomly generated from the normal distribution with N(1,0.25). 0.05 level one- and two-sided tests (1ST, 2ST) are considered.

Tables 3.2-4 show that the smaller the probability of the outcome, the larger the needed sample size. Especially, as the probability of outcomes goes down to around 0.1, the required sample size substantially increases. This implies that if the outcome is rare, we may need extremely large sample size to obtain enough statistical power. Tables 3.2-4, and Figures 3.1-4 show that a 0.05 level one-sided test is much more suitable with this MP than the two-sided test because the MP ranges from 0 to 1 and is asymmetric. In addition, distributions of outcomes and exposures may jointly influence performance of power and sample size calculation. In comparison to the VM, the proposed method provides slightly larger required sample sizes, whose results follow the statement of Vittinghoff *et al.* (2009). Moreover, the proposed method reflects the effects of exposure distributions, which are less considered by the VM.

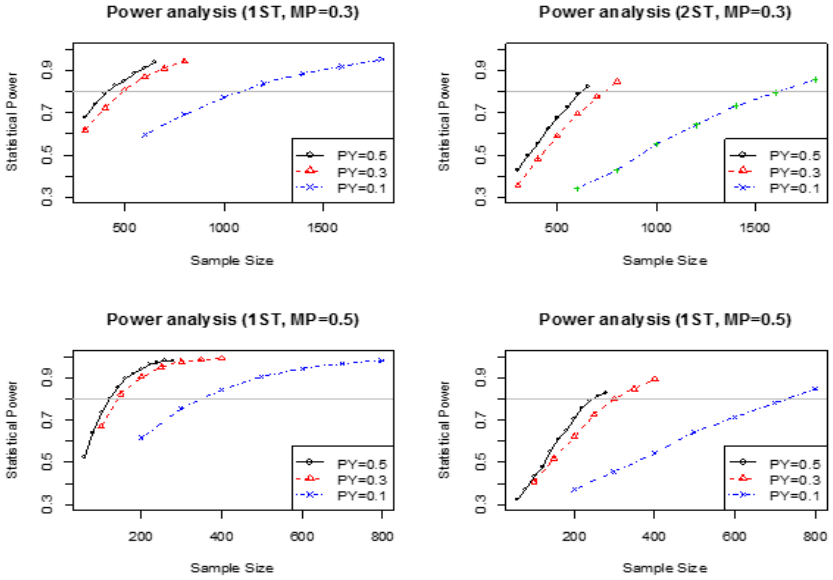

Figure 3.3 . Plots of the required sample sizes and statistical powers to detect mediated effects using the MP (the true MP=0.3, 0.5) based on the setup of Table 3.4 when an outcome is dichotomous, a mediator is continuous and an exposure is randomly generated from the normal distribution with N(1,1). 0.05 level one- and two-sided tests (1ST, 2ST) are considered.

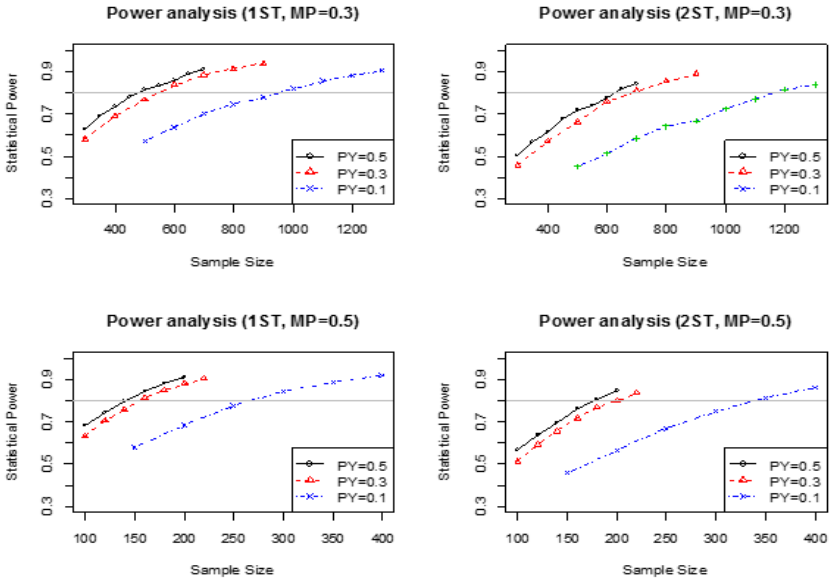

Figure 3.4 Plots of the required sample sizes and statistical powers to detect mediated effects using the MP (the true MP=0.3, 0.5) based on the setup of Table 3.4 when an outcome is dichotomous, a mediator is continuous and an exposure is randomly generated from the normal distribution with N(1,4). 0.05 level one- and two-sided tests (1ST, 2ST) are considered.

4. Data Example

In this section, we apply our proposed method for a continuous mediator to real data on effects of ionizing radiation on glucose and lipid metabolism in the Adult Health Study (AHS) cohort of atomic bomb survivors. This cohort has been followed by biennial health examinations since 1958 at the Radiation Effects Research Foundation (RERF) in Hiroshima and Nagasaki, Japan (Yamada *et al.*, 2004). Usually, researchers would like to know how large a sample is required to achieve statistical power over 80% to investigate an association (between heart disease and radiation dose possibly mediated through hemoglobin A1c (HbA1c) levels in this example). The outcome is prevalence of heart disease and the exposure is radiation dose with mean and standard deviation of 0.23 and 0.45 gray. Radiation dose to the colon (weighted combination of γ-ray dose plus ten times neutron dose, adjusted for dose uncertainty) was used as a surrogate for dose to the heart and major blood vessels, as there are currently no direct estimates of heart and circulatory-system organ doses. The mediator is HbA1c level. For the MC studies to compute statistical power, we used previous information to assume the mediation model (Akahoshi *et al.*, 2003; Ito *et al.*, 2000; Nakanishi *et al.*, 2005; Ozasa *et al.*, 2012; Saito *et al.*, 2011). The bootstrap with 1,000 replicates was used to compute CIs for the MP. 5000 MC simulations were run for the power analysis. Effects of radiation dose on health outcomes, possibly mediated through HbA1c level, were estimated based on the MP assuming linear dependence of radiation dose on HbA1c level and a logistic regression to relate heart disease to radiation dose and HbA1c level. The MP in this study means the proportion of the total effect of radiation dose on the incidence of heart disease that can be explained by HbA1c level.

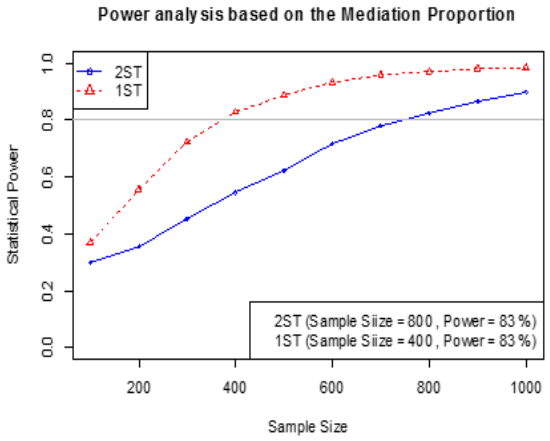

Figure 4.1 Plot of required sample sizes vs statistical powers to test the null hypothesis, based on 0.05 level one- (red line, 1ST) and two-sided (blue, 2ST) tests keeping the prevalence of heart disease. Both radiation dose and generated HbA1c level are continuous variables.

The mediation model for power analysis was assumed to be HBA1c=HBA1c(0)+θ_1Dose and HD$\sim$Ber(P_{HD})0 with $P_{HD} = \{1 + \exp(-\beta_0 - \beta_1 \text{Dose} - \beta_2 \text{HbA1c})\}^{-1}$, where HbA1c(0)

is the background HbA1c level following a normal distribution with mean and standard deviation of 5.6 and 1.1 (mmol/mol), HD is heart disease (i.e., HD=1 for incidence of heart disease and HD=0 otherwise), P_{HD} is the probability of HD, $\beta_0 = \text{logit}(\text{P}(HD(0)))$ with a baseline incidence of heart disease HD(0) (i.e., HD(0)=1 for incident heart disease and HD(0)=0 otherwise) and θ_1, β_1 and β_2 were assumed to be 0.4, 0.4 and 0.5, respectively. We also assumed that there was no product term between radiation dose and HbA1c levels on the incidence of heart disease (i.e. no interaction between exposure and mediator). The MP is 0.334 and the prevalence of heart disease is 0.22.

Figure 4.1 shows that at statistical power of 83%, the required sample size is 800 for a 0.05 level two-sided test (blue line, 2ST). We can also achieve statistical power of 83% with a sample size 400 for a 0.05 level one-sided test (red line, 1ST). Because the prevalence of heart disease is 0.22, if a case-control study is considered, the ratio of case and control group sizes would be 0.22/0.78.

5. Discussion

Our objective in this paper was to propose an alternative power analysis strategy based on the mediation proportion: on the log-odds ratio or odds ratio scale for a dichotomous outcome, using a flexible Monte Carlo (MC) approach with bootstrap replicates in the causal mediation model. The bootstrap procedure was used to test the null hypothesis, $H_0 : \text{MP} = 0$ vs the alternative hypothesis, $H_1 : \text{MP} > 0$ or $H_1 : \text{MP} \neq 0$. We need substantially larger samples to achieve acceptable statistical power when the probability of event occurrence is small.

From the results of Table 2.1, the bias-corrected bootstrap has better performance than the basic bootstrap confidence interval estimation in finite samples. Table 3.2 shows that the required sample sizes based on the mediation proportion using only the total effect as a denominator (MPO), which is the mediation proportion using the total effect only as a denominator, to achieve statistical power over 80% are substantially larger than those based on the mediation proportion using the sum of the natural direct and indirect effects (MPS), which is the mediation proportion using the sum of natural direct and indirect effects as a denominator, in finite samples using the bootstrap methods. This implies that the bootstrap causes more variability for the MPO than for the MPS. Therefore, we conclude that using the MPS and the bias-corrected bootstrap method help improve the problem of volatility of the MP estimator in finite samples and we therefore recommend the bias-corrected bootstrap method.

Our proposed method is computer-intensive because it uses both MC simulation and bootstrap replicates. This is partly related to inefficiency in running two nested simulation loops in R. If lower-level program languages such as Fortran, C++ and JAVA can be used, the computational cost may be reduced. However, large computing time does not translate into complexity of the problem; despite the method being computer-intensive, we believe it is straightforward and intuitive.

To the best of our knowledge, this is the most systematic investigation and illustration of a relatively easy and accessible power analysis for the MP on the log-OR scale for a dichotomous outcome, and a continuous mediator. The findings from simulation studies provide guidelines for researchers who want to consider causal mediation analysis before their research starts, and imply a relationship among factors such as the MP, probability of event

occurrence, probability of discrete mediators and exposures, or distribution of exposures, etc. In addition, we report an association between sample size and statistical power in the casual mediation model with a dichotomous outcome on the log-OR scale. However, when natural direct and indirect effects have opposite direction, the power analysis using the MP is not applicable. In addition, the direction of the product term between a mediator and an exposure on an outcome is opposite to natural direct and indirect effects and is strong, this method is not applicable. Thus, before applying the proposed power analysis, researchers should endeavor to check whether the direction between the two effects and the product term between exposure and mediator in the outcome model are opposite or not.

Acknowledement : The Radiation Effects Research Foundation (RERF) in Hiroshima and Nagasaki, Japan is a public interest foundation funded by the Japanese Ministry of Health, Labour and Welfare (MHLW) and the US Department of Energy (DOE). The research was also funded in part through DOE award DE-HS0000031 to the National Academy of Sciences. This publication was supported by RERF Research Protocol A5-10. The views of the authors do not necessarily reflect those of the two governments. This research was also supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (NRF-2016R1D1A1B03932212).

References

Akahoshi, M., Amasaki, Y., Soda, M., Hida, A., Imaizumi, M., Nashima, E., Maeda, R., Seto, S. and Yano, K. (2003). Effects of radiation on fatty liver and metabolic coronary risk factors among atomic bomb survivors in Nagasaki. *Hypertension Research*, **6**, 965-970.

Baron, R. M. and 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**, 1173-1182.

Carpenter, J. and Bithell, J. (2000). Bootstrap confidence intervals: When, which, what? A practical guide for medical statisticians. *Statistics in Medicine*, **19**, 1141-1164.

Cohen, J. (1988). *Statistical power analysis for the behavior sciences 2nd edition*. Lawrence Eribaum Associates, New Jersey.

Davison, A. C., and Hinkley, D. V. (1997).*Bootstrap methods and their application*. Cambridge University Press.

Ditlevsen, S., Christensen, U., Lynch, J., Damsgaard, M. T. and Keiding, N. (2005). The mediation proportion : A structural equation approach for estimating the proportion of exposure effect on outcome explained by an intermediate variable. *Epidemiology*, **16**, 114-120.

Freedman, L. S., Graubard, B. I. and Schatzkin, A. (1992). Statistical validation of intermediate endpoints for chronic diseases. *Statistics in Medicine*, **11**, 167-178.

Fritz, M. S. and MacKinnon, D. P. (2007). Required sample size to detect the mediated effect. *Psychological science*, **18**, 233-239.

Hafeman, D. M. and Schwartz, S. (2009). Opening the black box: a motivation for the assessment of mediation. *American Journal of Epidemiology*, **38**, 838-845.

Hernan, M. A. (2004). A definition of causal effect for epidemiological research. *Journal of Epidemiology and Community Health*, **58**, 265-271.

Hong Y., Jang, G. and Choi, C. (2016). Life satisfaction and self-esteem of children from low-income class: Testing mediation model of depression. *Journal of the Korean Data & Information Science Society*, **27**, 179-189.

Ito, C., Maeda, R., Ishida, S., Sasaki, H. and Harada, H. (2000). Correlation among fasting plasma glucose, two-hour plasma glucose levels in OGTT and HbA1c. *Diabetes Research and Clinical Practice*, **50**, 225-230.

Jeong, K. (2017). Quantile causality from dollar exchange rate to international oil price. *Journal of the Korean Data & Information Science Society*, **28**, 367-369.

Kenny, D. A. and Judd, C. M. (2014). Power anomalies in testing mediation. *Psychological science*, **25**, 334-339.

Lynch, J. W., Kaplan, G. A., Cohen, R. D., Tuomilehto, J. and Salonen, J. T. (1996). Do cardiovascular risk factors explain the relation between socioeconomic status, risk of all-cause mortality, cardiovascular mortality and acute myocardial infarction? *American Journal of Epidemiology*, **144**, 934-942.

MacKinnon, D. P. (2008). *An introduction to statistical mediation analysis*. Lawrence Erlbaum Associates, New York.

Nakanishi, S., Yamada, M., Hattori, N. and Suzuki, G. (2005). Relationship between HbA(1)c and mortality in a Japanese population. *Diabetologia*, **48**, 230-234.

Ozasa, K., Shimizu, Y., Suyama, A., Kasagi, F., Soda, M., Grant, E. J., Sakata, R., Sugiyama, H. and Kodama, K. (2012). Studies of the mortality of atomic bomb survivors, Report 14, 1950-2003: An overview of cancer and noncancer diseases. *Radiation Research*, **177**, 229-243.

Pearl, J. (2001). Direct and indirect effects. In: Proceedings of the seventeenth conference on uncertainty and artificial intelligence. San Francisco, CA: Morgan Kaufmann 411-420.

Preacher, K. J. and Kelley, K. (2011). Effect size measures for mediation models: quantitative strategies for communicating indirect effects. *Psychological Methods*, **16**, 93-115.

Preacher, K. J. and Selig, J. P. (2012). Advantages of Monte Carlo confidence intervals for indirect effects. *Communication Methods and Measures*, **6**, 77-98.

Richiardi, L., Bellocco, R. and Zugna, D. (2013). Mediation analysis in epidemiology: methods, interpretation and bias. *International Journal of Epidemiology*, **42**, 1511-1519.

Robin, J. M. and Greenland, S. (1992). Identifiability and exchangeability for direct and indirect effects. *Epidemiology*, **3**, 143-155.

Rubin, D. B. (1990). Formal mode of statistical inference for causal effects. *Journal of Statistical Planning and Inference*, **25**, 279-292.

Saito, I., Kokubo, Y., Yamagishi, K., Iso, H. and Inoue, M. (2011). Diabetes and the risk of coronary heart disease in the general Japanese population: The Japan public health center-based prospective (JPHC) study. *Atherosclerosis*, **216**, 187-191.

Sedlmaier, P. and Gigerenzer, G. (1989). Do studies of statistical power have an effect on the power of studies? *Psychological Bulletin*, **105**, 309-316.

Tsiatis, A. A., DeGruttola, V. and Wulfsohn, M. S. (1995). Modeling the relationship of survival to longitudinal data measured with error: Applications to survival and CD4 counts in patients with AIDs. *Journal of the American Statistical Association*, **90**, 27-37.

Valeri, L. and VanderWeele, T. J. (2013). Mediation analysis allowing for exposure-mediator interactions and causal interpretation: Theoretical assumptions and implementation with SAS and SPSS macros. *Psychological methods*, **18**, 137-150.

VanderWeele, T. J. (2015). *Explanation in causal inference: Methods for mediation and interaction*. Oxford University Press, New York, U.S.A.

VanderWeele, T. J. and Vansteeland, S. (2010). Odds ratios for mediation analysis for a dichotomous outcome. *American Journal of Epidemiology*, **172**, 1339-1348.

Vittinghoff, E. and Neilands, T. B. (2015). Sample size for joint testing of indirect effects. *Prevention Science*, **16**, 1128-1135.

Vittinghoff, E. Sen, S. and McCulloch, C. E. (2009). Sample size calculations for evaluating mediation. *Statistics in Medicine*, **28**, 541-557.

Wang, C. and Xue, X. (2012). Power and sample size calculations for evaluating mediation effects in longitudinal studies. *Statistical methods in Medical Research*, **25**, 686-705.

Wang, Y. and Taylor, J. M. G. (2002). A measure of the proportion of treatment effect explained by a surrogate marker. *Biometrics*, **58**, 803-812.

Yamada, M., Wong, F. L., Fujiwara, S., Akahoshi, M. and Suzuki, G. (2004). Noncancer disease incidence in atomic bomb survivors, 1958-1998. *Radiation Research*, **161**, 622-632.

Zhang, Z. (2014). Monte Carlo based statistical power analysis for mediation models: Methods and software. *Behavior Research Methods*, **46**, 1184-1198.