Journal of Clinical Epidemiology 59 (2006) 457–464

Small-scale randomized controlled trials need more powerful methods of mediational analysis than the Baron–Kenny method

Ester Cerin[a,*], Lorian M. Taylor[b], Eva Leslie[a], Neville Owen[a]

[a]Cancer Prevention Research Centre, School of Population Health, The University of Queensland, Herston Road, Herston, Brisbane, QLD 4006, Australia

[b]Centre for Health Promotion Studies, The University of Alberta, Alberta, Canada

Accepted 7 November 2005

Abstract

Objective: To devise more-effective physical activity interventions, the mediating mechanisms yielding behavioral change need to be identified. The Baron–Kenny method is most commonly used, but has low statistical power and may not identify mechanisms of behavioral change in small-to-medium size studies. More powerful statistical tests are available.

Study Design and Setting: Inactive adults ($N = 52$) were randomized to either a print or a print-plus-telephone intervention. Walking and exercise-related social support were assessed at baseline, after the intervention, and 4 weeks later. The Baron–Kenny and three alternative methods of mediational analysis (Freedman–Schatzkin; MacKinnon et al.; bootstrap method) were used to examine the effects of social support on initial behavior change and maintenance.

Results: A significant mediational effect of social support on initial behavior change was indicated by the MacKinnon et al., bootstrap, and, marginally, Freedman–Schatzkin methods, but not by the Baron–Kenny method. No significant mediational effect of social support on maintenance of walking was found.

Conclusions: Methodologically rigorous intervention studies to identify mediators of change in physical activity are costly and labor intensive, and may not be feasible with large samples. The use of statistically powerful tests of mediational effects in small-scale studies can inform the development of more effective interventions.

Keywords: Randomized controlled trials; Mediational effect; Small sample; Exercise-related social support; Walking; Statistical power

1. Introduction

To improve the effectiveness of physical activity interventions, there is the need to identify how (analysis of mediators), for whom (analysis of personal moderators), and under what circumstances (analysis of situational moderators) they can lead to increases in physical activity [1]. As well as improving the understanding of key elements in interventions, such research can also help clarify the general causal mechanisms underlying behavioral change [2].

Research on the mediators of physical activity behavior change has yielded an inconsistent pattern of findings [3,4]. These inconsistencies may in part be due to differences in the statistical power of studies on mediational effects [5]. To find a statistically significant mediational effect (if any exist), intervention studies must (a) produce sizable changes on the hypothesized mediators and outcomes, (b) be based on a sufficiently large sample size, and (c) make use of statistical methods capable of detecting a mediational effect of a specific magnitude for a given sample size.

Exercise-related social support from family and friends is a factor associated with being more likely to be physically active [6,7]. To date, however, only a small number of studies have examined the mediational effect of social support on changes in physical activity behavior, using a randomized controlled trial design. This is the recommended methodological framework for testing hypothetical mediators of behavior change [2]. Of these studies, those with a sample size >250 fully or partially supported the existence of a mediational effect [8–10] and those with <150 participants failed to do so [11,12]. Given that smaller-scale trials typically found small-to-moderate positive associations between the intervention, social support, and physical activity behavior (1%–12% of common variance), it is possible that their failure to identify social support as a significant mediator of changes in physical activity was partly due to a lack of statistical power.

* Corresponding author. Tel.: +852-2817-9846; fax: +852-2818-8042.
E-mail address: ecerin@hku.hk (E. Cerin).

0895-4356/06/$ – see front matter
doi: 10.1016/j.jclinepi.2005.11.008

2. Methods of mediational analysis

2.1. The Baron–Kenny causal steps approach

Currently, in the physical activity behavior domain, the most commonly used method of mediational analysis is the Baron–Kenny [13] causal steps approach, which specifies a series of tests of the links in a causal chain (Fig. 1) [1,3,14]. This approach requires four conclusions for mediation which, if applied for example to a study on the effects of social support in an intervention to influence walking, would read:

1. τ: The intervention is associated with changes in the outcome (walking).
2. α: The intervention is associated with changes in the hypothesized mediator (social support).
3. β: Changes in the mediator (social support) are associated with changes in the outcome (walking) after controlling for the intervention.
4. τ': When controlling for the effect of the intervention on the mediator (social support) and the effect of the mediator (social support) on the outcome (walking), a previously significant relationship between the intervention, τ, and the outcome (walking) is attenuated or no longer significant.

Despite its widespread use and obvious utility, the Baron–Kenny method has some limitations [5]. For instance, it does not provide a direct estimate of the size of the indirect (mediated) effect of the intervention on the outcome, which is needed for the meta-analysis of the summary statistics of mediational tests. Additionally, the Baron–Kenny approach has low statistical power in studies with a relatively small sample size (e.g., $N \leqslant 50$), even in the presence of large mediational effect sizes (explaining $>25\%$ of the outcome variance, as defined by Cohen [15]). A recent simulation study comparing 14 methods for testing the statistical significance of mediational effects found that, under these conditions (small sample size and large effect size), the Baron–Kenny approach attained as low a power as .47, and some other methods of mediational analysis showed an acceptable power of .80 or higher [5].

One of the reasons why the Baron–Kenny approach has lower statistical power relates to its requirement that a significant overall effect of the intervention on the outcome (τ; Fig. 1) be demonstrated. This requirement is likely to be met if the effect size of the intervention is moderate to large and the samples sizes are moderate to large ($N > 100$) [16]. Public health intervention studies on physical activity are most often conducted in the participants' natural environment. These are relatively uncontrolled settings, in which a considerable number of factors can interact with the intervention. Thus, the effect of the intervention on changes in physical activity behavior is likely to be, at best, small to moderate in size (explaining approximately up to 13% of the outcome variance). Recent reviews and studies indicate that this is often the case [4,17]. Hence, it can be expected that small- to moderate-scale studies adopting the Baron–Kenny approach may fail to detect as statistically significant a mediational effect on physical activity behavior that would be of practical importance.

The need for testing the association between the intervention conditions and the outcome (τ) to determine a mediational effect has been recently questioned by several methodologists [16,18]. Shrout and Bolger [16] suggest that this step in the analysis be avoided when temporally distal (i.e., long-term) rather than temporally proximal (short-term) mediation processes are examined. In fact, with an increase of the interval between an intervention and the assessment of behavior change the size of the effect typically gets smaller. This is due to an increase in likelihood that a behavior change will be (a) mediated by other mechanisms; (b) affected by competing causes; and/or (c) affected by random factors.

2.2. More powerful methods of mediational analysis

Researchers in the field of physical activity and health often seem to be unaware of the existence of an array of methods of mediational analysis that are far more powerful

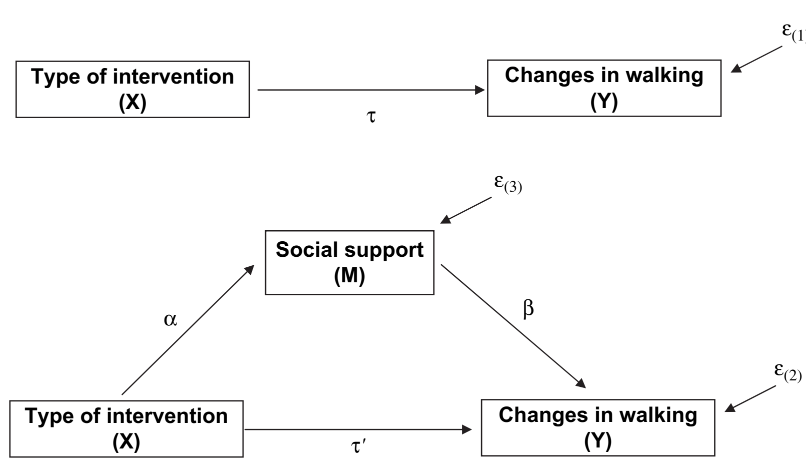

Fig. 1. Path diagram for the analysis of the mediational effect of social support on changes in walking behavior.

than the Baron–Kenny method. For instance, in a simulation study, Freedman and Schatzkin's [19] and Clogg et al.'s [20] difference-in-coefficients tests, and two product-of-coefficients tests by MacKinnon et al. [21] (one based on the distribution of the product of two standard normal variables $z_\alpha z_\beta$, the other based on the empirical distribution of $\alpha\beta/\sigma_{\alpha\beta}$) were found to be the most powerful methods for testing mediating effects [5]. The power of these methods was particularly emphasized in cases when the sample size was small ($N \leqslant 50$) and the population values of α and β (Fig. 1) were low to medium in size. Additionally, these methods yielded accurate type I error rates when both α and β were zero. For the present study, we applied the Freedman–Schatzkin difference-in-coefficients test [19] and the MacKinnon et al. product-of-coefficients test [21] based on the empirical distribution of $\alpha\beta/\sigma_{\alpha\beta}$, because they have been shown to statistically outperform other methods within their own category [5].

The Freedman–Schatzkin difference-in-coefficients test assesses a mediating effect by comparing the relation between the independent variable (the intervention) and the dependent variable (e.g., walking) before and after adjustment for the mediator (e.g., social support). In other words, it tests the null hypothesis that the difference between the unadjusted and adjusted regression coefficients of the independent variable ($\tau - \tau'$) is zero (Fig. 1). In contrast, the MacKinnon et al. product-of-coefficients test assesses the statistical significance of a mediational effect by dividing the product of the coefficients α and β by its standard error. The obtained value is then compared to an empirical (z') distribution. Notably, the MacKinnon et al. test does not suffer from lack of power, as do more traditional product-of-coefficients tests; this is because these tests use a standard normal distribution to assess the statistical significance of the observed mediational effect, although the distribution of the product of regression coefficients α and β is usually asymmetric with high kurtosis [22].

The MacKinnon et al. and Freedman–Schatzkin's tests are conceptually and algebraically equivalent. They test the same null hypothesis, that the moderating effect is not significantly different from zero, and produce the same estimates of mediational effect for ordinary least-square regressions [23]. As we have already noted, however, they use a different statistic for testing the statistical significance of the estimate. In the MacKinnon et al. [5] simulations, with respect to statistical accuracy Freedman and Schatzkin's approach differed from the MacKinnon et al. test in that it had higher type I error rates when the population α was zero and β was nonzero; the opposite was true when α was nonzero and β was zero. The Freedman–Schatzkin test produced acceptable type I error rates when the population α was nonzero and β was zero, but the error rates associated with the MacKinnon et al. test were still relatively high when α was zero and β was nonzero (0.14 for small effect sizes to 0.35 for large effect sizes). Consequently, caution is advised in using the MacKinnon et al. test when only the estimate of β is found to be statistically significant.

2.3. Bootstrap product-of-coefficients test

As noted earlier, most product-of-coefficients tests suffer from a lack of power due to their notably skewed distributions [5]. MacKinnon and colleagues efficiently circumvented this problem by providing an empirical (z') distribution of the ratio of the product of the coefficients α and β and its standard error [5]. An alternative method would be to use bootstrap procedures [16]. Bootstrap methods are appropriate for small sample sizes when data are nonnormally distributed. In contrast to parametric inferential statistics, which rely on the usually unverifiable assumption of normality of a sampling distribution and a readily available formula to calculate the parameters of the distribution, the bootstrap method attempts to estimate sampling distributions for statistical estimators empirically [24]. This is done by using information drawn from the sample of observations used to estimate the statistical model in the first place. Data from the original sample are assumed to be the total population of responses, and the bootstrap program repeatedly randomly redraws a large number of samples (named *bootstrap* samples) from this pseudopopulation, with replacement. At the conclusion of the bootstrap sampling, the average parameter estimates (in this case, the product of the coefficients α and β), the bootstrapped estimate of the standard error (which is the standard deviation of the parameter estimates computed across the bootstrap samples), and 95% percentile confidence intervals of the mean parameter estimates are computed across bootstrap samples. [16]

2.4. A real-data example of mediational analysis in small randomized controlled trials

Telephone-delivered interventions to increase physical activity have been identified as a potentially sustainable, accessible, and cost-effective intervention modality [25] that may, in some instances, be more efficient and effective than are print or face-to-face interventions [4]. Our main objective was to provide, using a telephone-delivered intervention as a case in point, a typical example of a low-power study of mediational effects characterized by a small sample size, an expected small-to-moderate effect size and many missing data. Additionally, although the authors do not maintain that social support is the only mediator of change in walking behavior, for simplicity, the analysis was limited to this single mediator.

To exemplify how to conduct statistically powerful mediational analyses, the Baron–Kenny, Freedman–Schatzkin, MacKinnon et al., and bootstrap methods were used on the data from our small-scale, telephone-delivered intervention trial looking at the mediational effects of social support on initial changes in and maintenance of walking behavior. It was predicted that the Freedman–Schatzkin, MacKinnon et al., and bootstrap methods would be able to detect any small-to-moderate mediating effect of social support on walking, but the Baron–Kenny method would not do so.

3. Methods

3.1. Participants

The study was aimed at recruiting inactive adults aged ⩾45 years. Fifty-two middle-aged to older adults (18 men and 48 women, aged 45–78 years) who identified themselves as physically underactive (i.e., not regularly physically active for >120 minutes a week) participated in the study. Participants were recruited via newspaper advertisements, a radio talk show, and word of mouth at two research sites (Melbourne and Brisbane, Australia). Ethics approval for this study was granted by the ethics committees of Deakin University and the University of Queensland. All participants were screened to ensure they had no limiting health conditions and provided their informed consent to be involved in the study.

3.2. Measures

Self-report measures were used to assess demographics, physical activity behavior, and perceived exercise-related social support from family and friends. Demographic characteristics such as age, gender, ethnicity, marital status, educational attainment, employment status, height, and weight were also assessed.

The Community Healthy Activities Model Program for Seniors (CHAMPS [26]) was used as the source of data on walking behavior. CHAMPS asks the participants to report the past 4-week frequency and duration of specific high-, moderate-, and low-intensity physical activities that older adults usually participate in. This instrument has been shown to be reliable, valid, and sensitive to change over time [26]. Given that walking has been shown to be the most common and feasible form of physical activity in middle-aged and older adults [27], only the four walking-related items of the CHAMPS were analyzed.

Perceived social support from friends and family was measured using a previously validated 13-item social support scale designed by Sallis and colleagues [28]. Participants rated how often over the last 3 months their family or friends had encouraged physical activity behavior, based on a five-point scale ranging from 1 (none) to 5 (very often).

3.3. Procedures

This study was an extension of the Community and Health Advice by Telephone study [4]. It was based on both social cognitive theory [29] and the transtheoretical model of behavior change [30]. In mid-2002, participants attended an initial interview during which they completed all study measures and were randomized to either the print or print-plus-telephone intervention. The participants in the print-plus-telephone intervention group nominated their preferred times of day for receiving telephone calls. Data were collected from participants at baseline, from week 1 to week 12 (intervention period), and from week 12 to week 16 (postintervention follow-up period).

Both groups received a counseling session, instructional newsletters, a pedometer, and weekly activity logs. One of the aims of the face-to-face counseling session was to develop an individualized physical activity plan that encouraged a gradual increase in physical activity frequency, duration, and intensity toward a goal of 30 minutes or more of moderate-intensity physical activity on most days of the week, according to current national physical activity guidelines [31]. Instructional newsletters, containing information on the benefits of physical activity and tips on how to be active in different weather conditions, were distributed to both groups at weeks 3 and 16. All participants completed a weekly activity log on which they recorded type and duration of their physical activity, heart rate while doing physical activity, rate of perceived exertion, and number of steps at the end of each day (based on pedometer readings). Participants in the print-plus-telephone intervention group additionally received six script-based telephone calls by a trained counselor (weeks 1, 2, 4, 6, 8, and 12). The telephone counselor assessed the participants' progress toward their goals and provided individualized encouragement and help in addressing barriers to increasing physical activity. At the conclusion of the 12-week intervention (initial change) and 4 weeks later (week 16; maintenance), the participants were reassessed on the CHAMPS and the social support measure.

3.4. Data analyses

A composite variable for walking was created by summing the scores on the four walking-related items of the CHAMPS and assessed total weekly amount of walking. A total score for social support was computed by summing the scores on all the items of the social support scale to give a possible score ranging from 23 to 115. The Shapiro–Wilk test and normal probability plots were used to verify whether the examined variables were normally distributed. Between-group differences in age and baseline values of social support and walking were analyzed with *t*-tests for independent samples. To assess the mediating effects of social support on walking behavior change, the Baron–Kenny, Freedman–Schatzkin, MacKinnon et al., and bootstrap methods were used.

Two sets of mediational analyses were conducted. The first assessed whether social support was a mediator of initial change in walking (12 weeks); the second, whether social support was a mediator of maintenance of walking behavior 4 weeks after the intervention (follow-up period).

For each mediational analysis, three multiple regression models were constructed.

The first regression model looked at the impact of the intervention condition on the residualized change score for walking after controlling for significant covariates. Here, the residualized change score was the postintervention score (whether after completion of the intervention or after the 4-week follow-up period) adjusted for the preintervention score (i.e., the baseline score), and represented the degree to which an individual's specific behavior increased more than would be expected given his/her initial status. The first regression model provided an estimate for τ (relation between intervention condition and change in walking before adjusting for the mediator), which was needed to conduct the Baron–Kenny and Freedman–Schatzkin's mediational tests (Fig. 1).

The second regression model assessed the impact of the intervention condition on the residual change scores for social support after controlling for significant covariates. This model provided an estimate for α (relation between intervention condition and change in social support), needed in the Baron–Kenny and MacKinnon et al. mediational tests.

The third model analyzed the effect of the intervention condition on residualized change scores for walking after controlling for significant covariates and residualized change scores for social support (estimate of τ', used in the Freedman–Schatzkin's method, representing the independent effect of the intervention condition on changes in walking). At the same time, this model estimated the effect of the change in social support (residualized change scores) on changes in walking (residualized change scores) after controlling for significant covariates and the effect of intervention conditions (estimate of β, used in the Baron–Kenny and MacKinnon et al. methods, representing the independent effect of change in social support on changes in walking). The equations used to compute the standard errors of the mediating effects, according to the Freedman–Schatzkin and MacKinnon et al. methods, are given in the Appendix.

The postintervention social support score adjusted for the baseline social support score was used for the mediational analyses of both initial change and maintenance of walking. This made it possible to determine whether the change in the mediator (social support) preceded a change in the outcome (walking), as suggested by Kraemer et al. [2]. In fact, operationally, for a factor to be considered a mediator in the relationship between an intervention condition and the outcome of interest, temporal precedence of changes in the mediator with respect to changes in the outcome needs to be documented.

Out of 52 participants, 27 (14 from the print and 13 from the telephone group) had incomplete data on at least one of the variables entered in the analysis of the mediational effects on maintenance of walking, and 22 participants (12 from the print and 10 from the telephone group) had missing data on at least one of the variables used for the analysis of the mediational effects on postintervention walking. To account for this problem, multiple regression analyses were conducted using Amos [32], a structural-equation modeling program, which uses full information maximum likelihood (FIML) to generate estimates of the missing values [33,34]. The FIML method has been shown to produce parameter estimates that are both consistent and efficient when data are missing at random [35]. This method has also been shown to outperform other simpler methods even when data have nonignorable missingness (i.e., the probability that a response is missing is related to the values of the outcome variable) [33], the sample size is small ($N = 50$) [35], and there are $\leq$50% of missing data [36]. In the present study, no significant associations between missingness patterns of social support and walking and other variables (e.g., sex, age, intervention condition) were observed. Hence, the data were assumed to be, at least, missing at random.

To perform a bootstrap product-of-coefficients test (product of the coefficients α and β), we created 1,000 bootstrap samples of $N = 52$ from the original dataset. Estimates of regression coefficients α and β, and their product, were computed for each of the bootstrap samples as explained above (regression model 2 and 3), using the FILM method. Finally, the mean, standard deviation, and 95% percentile confidence intervals of the observed products of α and β were computed across bootstrap samples.

4. Results

Table 1 reports descriptive statistics for all the variables examined in this study for each intervention group. A significant difference in mean age was found between the print (53.00 years; standard deviation SD = 5.27) and the telephone group (57.64 years; SD = 6.66; $t(50) = 2.76$; $P = .008$). Hence, age was included as a covariate in subsequent analyses. No significant between-group differences were found in baseline levels of social support and walking.

The three methods yielded different outcomes in the analysis of the mediational effect of social support on initial changes in walking. The Baron–Kenny approach did not reveal a significant mediational effect, but the MacKinnon

Table 1
Descriptive statistics by intervention group

	Print group			Telephone group		
Variable	No.	Mean	SD	No.	Mean	SD
Gender						
Female	21	—	—	17	—	—
Male	7	—	—	7	—	—
Age, years	28	57.6	6.7	24	53	5.3
Social support						
Baseline	24	45.3	9.9	18	45.3	9.7
Post intervention	17	46.5	9.4	16	51.1	14.5
Walking						
Baseline	25	3.8	2.8	22	3.6	2.4
Post intervention	23	5.3	3.3	20	6.1	2.7
Maintenance	21	5.8	3.8	19	5.6	4

et al. and bootstrap methods produced a significant result (Table 2). Similarly, the Freedman–Schatzkin test approached significance ($P = .053$). All four methods of mediational analysis produced similar results with regard to the effect of social support on maintenance: they failed to detect a significant mediational effect of social support on the maintenance of walking (Table 2).

5. Discussion

These findings are in line with the expected differential statistical power of the four methods of mediational analysis. The Baron–Kenny method did not identify a significant mediational effect of exercise-related social support on initial change in walking, but the other two methods did. Although we cannot exclude the possibility that the observed mediational effect was due to chance, it is worth noting that published intervention studies with a relatively large sample size ($N \geqslant 300$) found partial or full confirmation of a significant mediational effect of social support on physical activity change [8–10] but studies with smaller sample sizes did not [11,12,35].

Using the Baron–Kenny method, a sample of ⩾500 participants appears to be needed to attain adequate statistical power (⩾.80) when the size of the mediational effect is moderate. When the effect size is small, a power as low as .11 is achieved with a sample of 1,000 participants [5]. In contrast, if the effect is moderate, the two alternative methods of mediational analysis presented here can achieve acceptable statistical power with samples of as few as 50–100 participants.

Recruiting a sufficient number of participants to achieve adequate power for mediational analysis has practical limitations. Physical activity intervention studies usually last for several weeks or months [14,37], and the level of commitment required from participants is high. This is particularly so for methodologically rigorous investigations in which demands are made of participants in order to control for potential threats to validity. There is the need to minimize the risk of failing to identify significant mechanisms leading to changes in physical activity behavior due to insufficient sample size. One of the ways in which this can be accomplished is by using statistical methods with maximal power to detect mediational effects. The empirical evidence derived from such studies could be further analyzed by conducting meta-analyses of the summary statistics of mediational tests (estimated effect sizes and their standardized errors).

We acknowledge that the MacKinnon et al., Freedman–Schatzkin, and bootstrap methods are not the only methods that have been shown to perform well with small samples. Other statistical tests of mediational effects are also available, some of which have accurate type I error rates in all conditions and have acceptable statistical power when both the mediational effect and the sample are moderate in size [5]. Among these, the joint significance test of α and β is highly recommended for its accuracy and power [5]. In this variant of the causal steps approach, a significant mediational effect is demonstrated when separate tests for each path in the effect are jointly significant. As such, this method provides the most direct test of the simultaneous null hypothesis that path α and β (Fig. 1) are both equal

Table 2
Summary of analyses of mediational effect of social support on initial change and maintenance of walking behavior

Mediational test	Parameter estimates				
Baron–Kenny test [13]					
Parameter	$\hat{\tau}(\widehat{SE})$	t (50)	P	—[a]	—
Initial change	0.96 (0.85)	1.13	.26	—	—
Maintenance	1.43 (1.15)	1.24	.22	—	—
Freedman–Schatzkin [19] test					
Parameter	$\hat{\tau}(\widehat{SE})$	$\hat{\tau}'(\widehat{SE})$	$\hat{\tau}-\hat{\tau}'(\widehat{SE})$	t (50)	P
Initial change	0.96 (0.85)	0.50 (0.90)	0.47 (0.24)	1.98	0.053
Maintenance	1.43 (1.15)	1.37 (1.24)	0.06 (0.33)	0.18	0.858
MacKinnon et al. [21] test					
Parameter	$\hat{\alpha}(\widehat{SE})$	$\hat{\beta}(\widehat{SE})$	$\hat{\alpha}\hat{\beta}(\widehat{SD})$	z'	P
Initial change	6.79 (3.42)	0.07 (0.05)	0.47 (0.41)	1.144	~0.020[b]
Maintenance	6.79 (3.42)	0.01 (0.07)	0.06 (0.49)	0.125	~0.700[b]
Bootstrap test					
Parameter	—	—	Mean $\hat{\alpha}\hat{\beta}(\widehat{SD})$	95% PCI	—
Initial change	—	—	0.50 (0.39)	0.001, 1.461	—
Maintenance	—	—	0.08 (0.50)	−0.960, 1.146	—

Abbreviations and variables: PCI, percentile confidence interval; $\widehat{SD}$, estimate of standard deviation of parameter; $\widehat{SE}$, estimate of standard error of parameter; P, probability level; t, t-ratio; $\hat{\alpha}$, EURC of intervention condition on residualized change score for social support; $\hat{\beta}$, EURC of change in social support on residualized change score for walking; $\hat{\alpha}\hat{\beta}$, the MacKinnon et al. estimate of mediational effect; $\hat{\tau}$, estimate of unstandardized regression coefficient (EURC) of intervention condition on residualized change score for walking; $\hat{\tau}'$, EURC on residualized change score for walking after controlling for social support; $\hat{\tau}-\hat{\tau}'$, Freedman–Schatzkin's estimate of mediational effect.

[a] Inappropriate to assess other causal steps because step 1 was not significant.

[b] Only approximate P-levels are available for z'.

to zero. It is argued that that this method tests all necessary and sufficient conditions to define a mediator [38].

From a conceptual point of view, the main difference between the joint significance test of α and β and the Baron–Kenny causal steps approach pertains to the requirement of demonstrating a significant overall effect of the intervention on the outcome variable. It is argued that the analysis of mediators does not need testing for a significant overall effect of the intervention on the outcome [2,5,16]. In fact, this requirement excludes the possibility of an interaction effect of the intervention and mediator on the outcome [2] and the existence of inconsistent mediational effects (suppression), in which the indirect effect and the direct effect of an intervention may cancel out [39].

Although choosing a method of mediational analysis is likely to depend more on the conceptual framework adopted than on the expected size of mediational effect and number of participants in the study, researchers should keep in mind the existence, statistical performance, and conceptual basis of various procedures. None of the methods presented or mentioned here can, on their own, provide irrefutable evidence of a causal process through which an intervention produces an outcome. They demonstrate only that the causal processes specified by the hypothesized model are consistent with the data. Stronger causal inference can be achieved through carefully planned studies and designs in which both the intervention and the mediator are manipulated in a randomized experiment [5,40].

6. Conclusions

Randomized controlled trials are the optimal context in which mediational influences can be assessed, but it is costly and challenging to conduct a sufficiently large and methodologically rigorous experimental study capable of detecting the presence of mediational effects. Small-sample intervention studies using statistically powerful methods can contribute to the identification of important mediators of behavior change. Although the findings of a single well-controlled, small-sample study cannot be conclusive, a meta-analysis of the empirical evidence accumulated across a series of such studies can lead to confidence in the existence of a mediational effect on physical activity. This accumulated evidence can be used to guide the development of potentially more-effective interventions.

Acknowledgments

The support and advice of Professor Abby King (Stanford University), Professor Annette Dobson (The University of Queensland), and Dr. Stewart Trost (Kansas State University) are gratefully acknowledged.

Appendix

Freedman and Schatzkin's [17] difference-in-coefficients test assesses the null hypothesis that the difference between the unadjusted and adjusted regression coefficients of the independent variable ($\tau - \tau'$) is zero and is tested using the estimate of the standard error of $\hat{\tau} - \hat{\tau}'$(Fig. 1):

$$\hat{\sigma} = \sqrt{\hat{\sigma}_{\tau}^2 + \hat{\sigma}_{\tau'}^2 - 2\hat{\sigma}_{\tau}\hat{\sigma}_{\tau'}\sqrt{1 - \hat{\rho}_{XM}^2}},$$

where $\left|\widehat{\rho_{XM}}\right|$is the absolute value of the correlation between the independent variable and the mediator. The significance test of the mediating effect is computed by dividing $\hat{\tau} - \hat{\tau}'$by the standard error and comparing the obtained value to a t distribution with $N - 2$ degrees of freedom.

One variant of the MacKinnon et al. [18] product-of-coefficients test assesses the significance of the mediating effect by dividing its estimate (the product of the coefficients α and β), by its standard error, which is equal to:

$$\hat{\sigma}_{\alpha\beta} = \sqrt{\hat{\alpha}^2\hat{\sigma}_{\beta}^2 + \hat{\beta}^2\hat{\sigma}_{\alpha}^2},$$

where $\hat{\sigma}_{\alpha}$and $\hat{\sigma}_{\beta}$are the standard errors of α and β, respectively (Fig. 1). The ratio of the mediating effect to its standard error is then compared to the critical value of the MacKinnon et al. [18] empirical z' distribution (http://www.public.asu.edu/~davidpm/ripl/freqdist.pdf).

References

[1] Bauman AE, Sallis JF, Dzewaltowski DA, Owen N. Toward a better understanding of the influences on physical activity: the role of determinants, correlates, causal variables, mediators, moderators, and confounders. Am J Prev Med 2002;23(2 Suppl):5–14.

[2] Kraemer H, Wilson G, Fairburn C, Agras W. Mediators and moderators of treatment effects in randomized clinical trials. Arch Gen Psychiatry 2002;59:877–83.

[3] Baranowski T, Anderson C, Carmack C. Mediating variable framework in physical activity interventions: how are we doing? How might we do better? Am J Prev Med 1998;15:266–97 [Erratum in: Am J Prev Med 1999;17:98].

[4] Castro C, King A. Telephone-assisted counselling for physical activity. Exerc Sports Sci Rev 2002;30:64–8.

[5] MacKinnon DP, Lockwood CM, Hoffman JM, West SG, Sheets V. A comparison of methods to test mediation and other intervening variable effects. Psychol Methods 2002;7:83–104.

[6] Trost S, Owen N, Bauman A, Sallis J, Brown W. Correlates of adults' participation in physical activity: review and update. Med Sci Sports Exerc 2002;34:1996–2001.

[7] Dzator JA, Hendrie D, Burke V, Gianguilio N, Gillam HF, Beilin LJ, Houghton S. A randomized trial of interactive group sessions achieved greater improvements in nutrition and physical activity at a tiny increase in cost. J Clin Epidemiol 2004;57:610–9.

[8] Calfas KJ, Sallis JF, Nichols JF, Sarkin JA, Johnson MF, Caparosa S, Thompson S, Gehrman CA, Alcaraz JE. Project GRAD: two-year outcomes of a randomized controlled physical activity intervention among young adults. Am J Prev Med 2000;18:28–37.

[9] Miller Y, Trost S, Brown W. Mediators of physical activity behavior change among women with young children. Am J Prev Med 2002;23(2 Suppl):98–103.

[10] Pinto B, Lynn H, Marcus B, DePue J, Goldstein M. Physician-based activity counseling: intervention effects on mediators of motivational readiness for physical activity. Ann Behav Med 2001;23:2–10.
[11] Brassington GS, Atienza AA, Perczek RE, DiLorenzo TM, King AC. Intervention-related cognitive versus social mediators of exercise adherence in the elderly. Am J Prev Med 2002;23(2 Suppl):80–6.
[12] Castro C, Sallis J, Hickmann S, Lee R, Chen A. A prospective study of psychological correlates of physical activity for ethnic minority women. Psychol Health 1999;14:277–93.
[13] Baron R, Kenny D. The moderator–mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. J Pers Soc Psychol 1986;51:1173–82.
[14] Lewis B, Marcus B, Pate R, Dunn A. Psychosocial mediators of physical activity behavior among adults and children. Am J Prev Med 2002;23(2 Suppl):26–35.
[15] Cohen J. Statistical power for the behavioral sciences. 2nd ed. Hillsdale, NJ: Erlbaum; 1988.
[16] Shrout PE, Bolger N. Mediation in experimental and nonexperimental studies: new procedures and recommendations. Psychol Methods 2002;7:422–45.
[17] Green BB, McAfee T, Hindmarsh M, Madsen L, Caplow M, Buist D. Effectiveness of telephone support in increasing physical activity levels in primary care patients. Am J Prev Med 2002;22:177–83.
[18] MacKinnon DP, Krull JL, Lockwood CM. Equivalence of mediation, confounding and suppression effects. Prevention Sci 2000;1:173–81.
[19] Freedman LS, Schatzkin A. Sample size for studying intermediate endpoints within intervention trials of observational studies. Am J Epidemiol 1992;136:1148–59.
[20] Clogg CC, Petkova E, Shihadeh ES. Statistical methods for analyzing collapsibility in regression models. J Educ Stat 1992;17:51–74.
[21] MacKinnon DP, Lockwood C, Hoffman J. A new method to test for mediation. Paper presented at the annual meeting of the Society for Prevention Research, Park City, UT, June 5, 1998.
[22] Sobel ME. Asymptotic confidence intervals for indirect effects in structural equation models. In: Leinhardt S, editor. Sociological methodology 1982. Washington, DC: American Sociological Association; 1982: 290–312.
[23] MacKinnon DP, Warsi G, Dwyer JH. A simulation study of mediated effect measures. Multivariate Behav Res 1995;30:41–62.
[24] Bollen KA, Stine R. Direct and indirect effects: classical and bootstrap estimates of variability. Sociol Method 1990;20:115–40.
[25] Napolitano M, Marcus B. Targeting and tailoring physical activity information using print and information technology. Exerc Sports Rev 2002;30:122–8.
[26] Stewart AL, Mills KM, King AC, Haskell WL, Gillis D, Ritter PL. CHAMPS physical activity questionnaire for older adults: outcomes for interventions. Med Sci Sports Exerc 2001;33:1126–41.
[27] Australian Bureau of Statistics (ABS). Participation in sport and physical activity in Australia (No. 4177.0). Canberra: Australian Bureau of Statistics; 2000.
[28] Sallis J, Grossman R, Pinski R, Patterson T, Nader P. The development of scales to measure social support for diet and exercise behaviors. Prev Med 1987;16:825–36.
[29] Bandura A. Social foundations of thought and action: a social cognitive theory. Englewood Cliffs, NJ: Prentice Hall; 1986.
[30] Prochaska JO, Marcus BH. The transtheoretical model: applications to exercise. In: Dishman RK, editor. Advances in exercise adherence. Champaign, IL: Human Kinetics; 1994:161–80.
[31] Commonwealth Department of Health and Aged Care. National physical activity guidelines for Australians. Canberra, Australia: CDHAC; 1999.
[32] Arbuckle JL. Amos user's guide. Chicago: SmallWaters; 1995.
[33] Enders CK. A primer on maximum likelihood algorithms available for use with missing data. Struct Equation Model 2001;8:128–41.
[34] Bennett DA. How can I deal with missing data in my study? Aust NZ J Pub Health 2001;25:464–9.
[35] Schafer JL, Graham JW. Missing data: our view of the state of the art. Psychol Methods 2002;7:147–77.
[36] Collins LM, Schafer JL, Chi-Ming K. A comparison of inclusive and restrictive strategies in modern missing data procedures. Psychol Methods 2001;6:330–51.
[37] Burke V, Giangiulio N, Gillam H, Beilin LJ, Houghton S. Physical activity and nutrition programs for couples: a randomized controlled trial. J Clin Epidemiol 2003;56:421–32.
[38] Cohen J, Cohen P. Applied multiple regression: correlational analysis for the behavioral sciences. Hillsdale, NJ: Erlbaum; 1983.
[39] MacKinnon DP, Krull JL, Lockwood CM. Equivalence of the mediation, confounding, and suppression effect. Prev Sci 2000;1: 173–81.
[40] Sheets VL, Braver SL. Organizational status and perceived sexual harassment: detecting the mediators of a null effect. Pers Soc Psychol B 1999;25:1159–71.