BMC Medical Research Methodology

Open Access

Incorporating nonlinearity into mediation analyses

CrossMark

George J. Knafl[1*], Kathleen A. Knafl[1], Margaret Grey[2], Jane Dixon[2], Janet A. Deatrick[3] and Agatha M. Gallo[4]

Abstract

Background: Mediation is an important issue considered in the behavioral, medical, and social sciences. It addresses situations where the effect of a predictor variable X on an outcome variable Y is explained to some extent by an intervening, mediator variable M. Methods for addressing mediation have been available for some time. While these methods continue to undergo refinement, the relationships underlying mediation are commonly treated as linear in the outcome Y, the predictor X, and the mediator M. These relationships, however, can be nonlinear. Methods are needed for assessing when mediation relationships can be treated as linear and for estimating them when they are nonlinear.

Methods: Existing adaptive regression methods based on fractional polynomials are extended here to address nonlinearity in mediation relationships, but assuming those relationships are monotonic as would be consistent with theories about directionality of such relationships.

Results: Example monotonic mediation analyses are provided assessing linear and monotonic mediation of the effect of family functioning (X) on a child's adaptation (Y) to a chronic condition by the difficulty (M) for the family in managing the child's condition. Example moderated monotonic mediation and simulation analyses are also presented.

Conclusions: Adaptive methods provide an effective way to incorporate possibly nonlinear monotonicity into mediation relationships.

Keywords: Adaptive regression, Childhood chronic conditions, Fractional polynomials, Mediation, Moderated mediation, Nonlinearity

Background

Mediation is an important issue considered in the behavioral, medical, and social sciences, addressing situations where the effect of a predictor variable X on an outcome (or dependent or response) variable Y is explained to some extent by an intervening, mediator variable M. Methods for addressing mediation have existed for some time [1–3]. Since then, they have undergone a variety of refinements [4–28].

Relationships underlying mediation are commonly treated as linear in Y, X, and M. Mediating relationships, however, can be nonlinear. A few authors have addressed nonlinearity in the mediation context [29–32]. Pearl [30, 31] has developed a general approach for quantifying direct, indirect, and total effects allowing for nonlinearity as well as for categorical variables. Standard polynomials sometimes can be effectively used to address nonlinearity in

predictors, but not in general [33]. To fully address nonlinearity requires more complex methods like nonparametric regression. Methods are needed to assess linearity of mediation relationships and to conduct mediation analyses when the underlying relationships are nonlinear.

The purpose of this paper is 1. To present an approach for assessing linearity of mediation relationships and for conducting nonlinear mediation analyses, based on the adaptive regression methods of Knafl and Ding [34], and 2. To provide example analyses using these methods. Adaptive methods were originally developed for nonlinear modeling in the Poisson regression context with Poisson distributed count outcomes [35], but the methods extend to linear regression with continuous outcomes treated as normally distributed, as often used in mediation analyses, as well as logistic regression with discrete outcomes [36]. SAS® (SAS Institute, Inc., Cary, NC) macros have been developed to support adaptive regression. Knafl [37] presents examples of the use of one of these macros for

* Correspondence: gknafl@unc.edu
[1]School of Nursing, University of North Carolina at Chapel Hill, 5014 Carrington Hall, Campus Box 7460, Chapel Hill, NC 27599-7460, USA
Full list of author information is available at the end of the article

Knafl et al. BMC Medical Research Methodology (2017) 17:45

Page 2 of 21

nonlinear growth curve modeling. These methods address not only nonlinearity in predictors but also nonlinearity in outcomes when those outcomes are continuous and positive valued. They also allow for correlated outcomes and/or non-constant variances.

Behavioral, medical, and social science theories usually do not hypothesize that relationships are nonlinear but implicitly assume they are linear. However, Hayes and Preacher [29] provide a variety of examples of behavioral theories that incorporate nonlinearity. In any case, hypothesized relationships are usually stated in terms of directionality: as X increases, Y increases or decreases. Such statements are not inherently linear but can be represented more generally by possibly nonlinear monotonic relationships. Monotonicity is operationalized in this paper with single power transforms. Moderated monotonic mediation is also addressed in the paper.

Methods

This section starts by formulating standard linear mediation relationships, then nonlinear mediation relationships followed by monotonic mediation relationships. Adaptive regression methods are described next along with how they can be used to assess monotonic mediation and moderated monotonic mediation. The section ends with a description of the two data sets used in example analyses.

Linear mediation relationships

As commonly considered, regression models underlying mediation are formulated as linear in Y, X, and M. As an example, work pressure X could result in increased work stress M leading to increased alcohol consumption Y. The following models are considered relating Y to X (e.g., alcohol consumption to work pressure), M to X (e.g., work stress to work pressure), and Y to M and X (e.g., alcohol consumption to work pressure controlling for work stress), respectively (Fig. 1).

$$Y = i_{Y,1} + c \cdot X + U_{Y,1} \qquad (1)$$

$$M = i_M + a \cdot X + U_M \qquad (2)$$

$$Y = i_{Y,2} + c' \cdot X + b \cdot M + U_{Y,2} \qquad (3)$$

The random variables $U_{Y,1}$, U_M, and $U_{Y,2}$ are assumed to have mean zero. They can be thought of as omitted factors (as in [31]) or as random errors.

The slope c for X in model (1) is the total effect of X on Y (e.g., of work pressure on alcohol consumption) while the slope c' for X in model (3) is the direct effect of X on Y controlling for M (e.g., of work pressure on alcohol consumption controlling for work stress). The indirect effect Δ satisfies
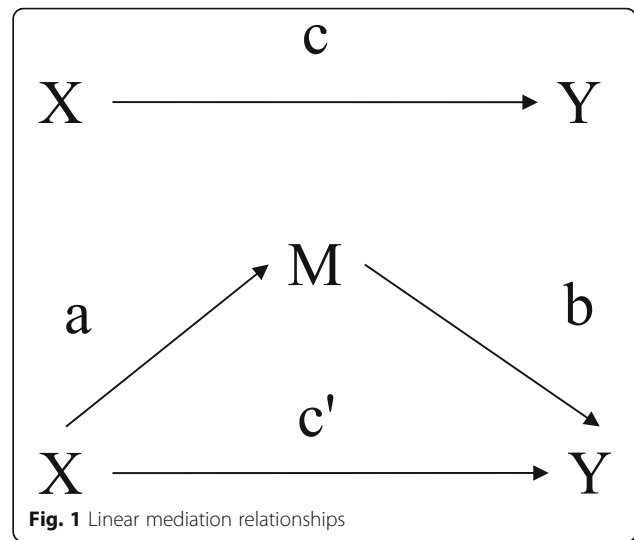
$$\Delta = c - c' = a \cdot b$$

Fig. 1 Linear mediation relationships

as demonstrated by substituting model (2) for the mediator M into model (3), and so the total effect is the sum of the direct and indirect effects. In the Baron and Kenny [1] approach, a significant total effect c was considered necessary for mediation to hold. However, this no longer is considered necessary [11]. There can be a nonzero indirect effect even when the total effect is nonsignificant [38]. However, model (1) is often investigated in practice. In any case, a crucial issue for mediation to hold is a significantly nonzero indirect effect Δ.

A variety of tests for a zero indirect effect are available. Sobel's test [21, 39] is the best known, but MacKinnon et al. [14] considered 14 alternatives including Sobel's test. Difference in coefficients approaches capitalize on the first equality for Δ while product of coefficients approaches including Sobel's test capitalize on the second equality. The assumptions underlying these tests can be questionable, and so the test based on the bootstrapped confidence interval (CI) for Δ has been proposed as a robust alternative [21, 29, 40, 41], especially when bias-corrected [42]. However, bias-corrected bootstrapped CIs can generate inflated Type I errors [43]. The strength of the indirect effect [21] can be measured by the relative indirect effect $\Delta' = \Delta / c$ (often denoted as P_M).

Since mediation addresses causality, temporal precedence is an important issue. This can be addressed by measuring X before measuring M and that before measuring Y [44]. However, mediation sometimes is addressed using cross-sectional data. Equations (1)-(3) are the same for these two cases. Random assignment of the settings of X provides support for causality for the X to Y relationship and the X to M relationship, but the M to Y relationship can still be confounded, and even when X, M, and Y have been measured longitudinally [45]. Nonexperimental or observational studies need theoretical justification for

Knafl *et al. BMC Medical Research Methodology* (2017) 17:45

Page 3 of 21

hypothesized causal relationships [21]. Pearl [31] has provided sufficient conditions for identifying direct, indirect, and total effects for observational studies.

Nonlinear mediation relationships

General approach for addressing nonlinearity

Nonlinear mediation involves the following two regression models (4)-(5) generalizing the linear regression models (2)-(3) while no longer considering model (1).

$$M = i_M + T_{M,X}(X) + U_M \tag{4}$$

$$Y = i_Y + T_{Y,X}(X) + T_{Y,M}(M) + U_Y \tag{5}$$

$T_{M,X}(X)$ and $T_{Y,X}(X)$ are possibly nonlinear transforms of the predictor X while $T_{Y,M}(M)$ is a possibly nonlinear transform of the mediator M. The parameters i_M and i_Y are standard intercepts. When one of the variables X or M is discrete, only its identity transform is considered.

As an example, Hayes and Preacher [29] considered the special case with

$$\begin{aligned} T_{M,X}(X) &= a \cdot \log X, \\ T_{Y,X}(X) &= c' \cdot X, \text{ and} \\ T_{Y,M}(M) &= b_1 \cdot M + b_2 \cdot M^2 \end{aligned} \tag{6}$$

where log denotes the natural log transform. Following Stolzenberg [32], they defined the instantaneous indirect effect, generalizing the product of coefficients approach, as the product of the derivatives $\frac{\partial M}{\partial X}$ and $\frac{\partial Y}{\partial M}$ replacing M by model (4) and assuming U_M and U_Y have no impact on these derivatives. For their example (6), this equals

$$(b_1 + 2 \cdot b_2 \cdot (i_M + a \cdot \log X)) \cdot \frac{a}{X} \tag{7}$$

(as also reported in their eq. (10)).

Pearl's more general approach

Pearl [31] provided the following general model for addressing nonlinearity in the mediation context, also allowing for categorical variables,

$$m = f_M(x, u_M) \text{ and } y = f_Y(x, m, u_y) \tag{8}$$

where x, m, y, u_M, and u_Y are possible values for the random variables X, M, Y, U_M, and U_Y while f_M and f_Y are general transforms. For example, the values m of work stress M could depend nonlinearly on the values x of work pressure X and the values u_M of the omitted factor or error variable U_M while the values y of alcohol consumption Y could depend nonlinearly on the values x of work pressure X, the values m of work stress M, and the values u_y of the omitted factor or error variable U_Y. Pearl also included the relationship $x = f_X(u_X)$, but that is not needed in what follows. He assumed that variables have been standardized, but that is not assumed here. He provided definitions of total, natural direct,

and natural indirect effects under model (8), which generalize those effects for the linear case of models (1)-(3).

The natural direct effect for a change in value of X from x to x' is defined as the expected value $NDE(x', x) = E(nde(x', x))$ where

$$nde(x', x) = f_Y(x', f_M(x, u_M), u_Y) - f_Y(x, f_M(x, u_M), u_Y).$$

In other words, $nde(x', x)$ is the change in y when x changes to x' while m is held fixed at its value for the initial value x. Expectations, here and in what follows, are with respect to U_M and U_Y. The natural indirect effect for a change in value of x to x' is the expected value $NIE(x', x) = E(nie(x', x))$ where

$$nie(x', x) = f_Y(x, f_M(x', u_M), u_Y) - f_Y(x, f_M(x, u_M), u_Y).$$

In other words, $nie(x', x)$ is the change in y when x is held fixed at its initial value while m changes from its value for x to its value for x'. The total effect for a change in value of x to x' is the expected value $TE(x', x) = E(te(x', x))$ where

$$te(x', x) = f_Y(x', f_M(x', u_M), u_Y) - f_Y(x, f_M(x, u_M), u_Y).$$

In other words, $te(x', x)$ is the change in y when x changes to x' and m changes from its value for x to its value for x'.

Adding and subtracting $f_Y(x', f_M(x, u_M), u_Y)$ to $te(x', x)$ gives that

$$TE(x', x) = NDE(x', x) - NIE(x, x')$$

(also eq. (14) of [31]). Thus, the total effect is only the sum of the natural direct and natural indirect effects in the special case that $NIE(x', x) = -NIE(x, x')$, which holds for the linear mediation case of models (1)-(3). This result indicates the shortcoming of trying to generalize the indirect effect using a difference of coefficients approach.

The definitions of $NDE(x', x)$, $NIE(x', x)$, and $TE(x', x)$ are sufficient for handling categorical predictors X. For the case of continuous predictors X, they can be used to define instantaneous natural direct, natural indirect, and total effects using limits as follows:

$$\frac{dNDE(x)}{dx} = \lim_{x' \to x} \frac{NDE(x', x)}{x' - x},$$

$$\frac{dNIE(x)}{dx} = \lim_{x' \to x} \frac{NIE(x', x)}{x' - x},$$

and

$$\frac{dTE(x)}{dx} = \lim_{x' \to x} \frac{TE(x', x)}{x' - x}.$$

The relative instantaneous natural indirect effect function can be computed as

Knafl *et al. BMC Medical Research Methodology* (2017) 17:45

Page 4 of 21

$$\Delta' = \frac{\dfrac{dNIE(x)}{dx}}{\dfrac{dTE(x)}{dx}}.$$

Examples

For the Hayes and Preacher [29] nonlinear example, $m = f_M(x, u_M)$ and $y = f_Y(x, m, u_Y)$ are determined by eqs. (4)-(5) evaluated at (6). Thus,

$$nde(x', x) = c' \cdot (x' - x) = NDE(x', x)$$

so that $\frac{dNDE(x)}{dx} = c'$ as would be expected for a linear model for Y in X. Also,

$$
\begin{aligned}
nie(x', x) &= b_1 \cdot f_M(x', u_M) + b_2 \cdot f_M^2(x', u_M) \\
&\quad - b_1 \cdot f_M(x, u_M) - b_2 \cdot f_M^2(x, u_M) \\
&= b_1 \cdot a \cdot (\log x' - \log x) + b_2 \cdot ((i_M + a \cdot \log x' + u_M)^2 \\
&\quad - (i_M + a \cdot \log x + u_M)^2) \\
&= b_1 \cdot a \cdot (\log x' - \log x) + b_2 \cdot (2 \cdot (i_M + u_M) \\
&\quad + a \cdot (\log x' + \log x)) \cdot a \cdot (\log x' - \log x)
\end{aligned}
$$

so that

$$\frac{dNIE(x)}{dx} = (b_1 + 2 \cdot b_2 \cdot (i_M + a \cdot \log x)) \cdot \frac{a}{x},$$

which agrees with result (7).

As a second example, consider the following case with $M = X + U_M$ (i.e., the linear model (2) with $i_m = 0$ and $a = 1$) and $Y = i_Y + c' \cdot X + b \cdot M^3 + U_Y$. Since $\frac{\partial M}{\partial X} = 1$ and $\frac{\partial Y}{\partial M} = b \cdot 3 \cdot M^2$, the instantaneous indirect effect for a fixed value x of X using the product of coefficients approach of Hayes and Preacher [29] would be $b \cdot 3 \cdot x^2$ (i.e., substituting x for M based on the relationship $M = X + U_M$ and ignoring the associated value u_M for U_M). On the other hand, the general approach of Pearl [31] gives

$$
\begin{aligned}
nie(x', x) &= b \cdot f_M^3(x', u_M) - b \cdot f_M^3(x, u_M) \\
&= b \cdot \left((x' + u_M)^3 - (x + u_M)^3 \right)
\end{aligned}
$$

so that

$$
\begin{aligned}
\lim_{x' \to x} \frac{nie(x', x)}{x' - x} &= b \cdot 3 \cdot (x + u_M)^2 \\
&= b \cdot 3 \cdot (x^2 + 2 \cdot x \cdot u_M + u_M^2).
\end{aligned}
$$

Since $E(u_M) = 0$, the instantaneous natural indirect effect satisfies

$$\frac{dNIE(x)}{dx} = b \cdot 3 \cdot (x^2 + E(U_M^2)).$$

The expected value $E(U_M^2) > 0$ except in trivial cases. Consequently, this is not the same value as obtained by the Hayes-Preacher product of coefficients approach. The problem with that approach is that U_M cannot

always be ignored in computing indirect effects. Similar problems could occur if f_Y is nonlinear in u_Y. This result indicates the shortcoming of trying to generalize the indirect effect using a product of coefficients approach.

Monotonic mediation

Theorized relationships are reasonably operationalized as monotonic relationships. Monotonic mediation can be formulated with models (9)-(10).

$$M^q = i_M + a \cdot X^{q'} + U_M \tag{9}$$

$$Y^p = i_Y + c' \cdot X^{p'} + b \cdot M^{q \cdot p''} + U_Y \tag{10}$$

For example, transformed work stress M^q could depend nonlinearly on transformed work pressure $X^{q'}$ in model (9) while transformed alcohol consumption Y^p could depend nonlinearly on transformed work pressure $X^{p'}$ controlling for transformed work stress $M^{q \cdot p''}$ in model (10). When $Y > 0$, $M > 0$, and $X > 0$, the power transforms Y^p, $X^{p'}$, $X^{q'}$, M^q, and $M^{q''}$ are well-defined for arbitrary real valued powers p, p', q, q', and $q'' = q \cdot p''$. An approach for extending this to arbitrary valued variables is presented later. The power $p = 0$ represents the natural log transform rather than the constant transform. These are Box-Tidwell transformations [46], although Box and Tidwell considered them only for predictors. Models (9)-(10) provide for transformation of outcomes as well as predictors as opposed to just predictors as in models (4)-(5) and (8). Fractional polynomials of degree 1 in X, M, and Y have been used in (9)-(10) to guarantee that relationships are monotonic.

Models (9)-(10) can be represented in the general form of model (8) replacing m and y by m^q and y^p giving

$$m^q = f_M(x, u_M) = i_M + a \cdot x^{q'} + u_M$$

and

$$y^p = f_Y(x, m, u_Y) = i_Y + c' \cdot x^{p'} + b \cdot m^{q \cdot p''} + u_M.$$

Hence, $nde(x', x) = c' \cdot ((x')^{p'} - x^p) = NDE(x', x)$ so that the instantaneous natural direct effect satisfies

$$\frac{dNDE(x)}{dx} = c' \cdot p' \cdot x^{p'-1}$$

for $p' \neq 0$ while $\frac{dNDE(x)}{dx} = c' \cdot x^{-1}$ for $p' = 0$. Also,

$$
\begin{aligned}
nie(x', x) &= b \cdot f_M^{p''}(x', u_M) - b \cdot f_M^{p''}(x, u_M) \\
&= b \cdot \left(\left(i_M + a \cdot (x')^{q'} + u_M \right)^{p''} \right. \\
&\quad \left. - \left(i_M + a \cdot (x)^{q'} + u_M \right)^{p''} \right).
\end{aligned}
$$

Note that $nie(x, x') = -nie(x', x)$ so that $te(x', x) = nde(x', x) + nie(x', x)$, and so the total effect is the sum of the

Knafl *et al. BMC Medical Research Methodology* (2017) 17:45

Page 5 of 21

natural direct and natural indirect effects as for the linear mediation case of models (1)-(3).

Examples

For the special case with $p'' = 1$,

$$nie(x', x) = b \cdot a \cdot \left((x')^{q'} - x^{q'} \right) = NIE(x', x)$$

so that

$$\frac{dNIE(x)}{dx} = b \cdot a \cdot q' \cdot x^{q'-1}$$

for $q' \neq 0$ while $\frac{dNIE(x)}{dx} = b \cdot a \cdot x^{-1}$ for $q' = 0$. Figure 2 represents this model (assuming $q' \neq 0$ and $p' \neq 0$). The paths are labeled with derivatives of expected values for M^q and Y^p relative to X or to M^q generalizing the slopes used in Figure 1. Note that as in the linear case of Figure 1, the instantaneous natural indirect effect equals the product $b \cdot a \cdot q' \cdot x^{q'-1}$ of the labels for the two upper paths and the instantaneous natural direct effect equals the label $c' \cdot p' \cdot x^{p'-1}$ for the lower path.

For the special case with $p'' = 2$ and $q'' \neq 0$,

$$nie(x', x) = b \cdot \left(\left(i_M + a \cdot (x')^{q'} + u_M \right)^2 - \left(i_M + a \cdot x^{q'} + u_M \right)^2 \right)$$

$$= b \cdot \left(2 \cdot (i_M + u_M) + a \cdot (x')^{q'} + a \cdot x^{q'} \right) \cdot a \cdot \left((x')^{q'} - x^{q'} \right)$$

so that

$$\frac{dNIE(x)}{dx} = b \cdot 2 \cdot \left(i_M + a \cdot x^{q'} \right) \cdot a \cdot q' \cdot x^{q'-1}.$$
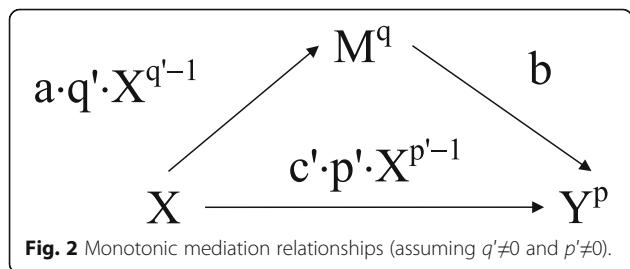
For the special case with $p'' = 3$ and $q \neq 0$,

$$nie(x', x) = b \cdot \left(\left(i_M + a \cdot (x')^{q'} + u_M \right)^3 - \left(i_M + a \cdot x^{q'} + u_M \right)^3 \right)$$

so that

$$\lim_{x' \to x} \frac{nie(x', x)}{x' - x} = b \cdot 3 \cdot \left(i_M + a \cdot x^{q'} + u_M \right)^2 \cdot a \cdot q' \cdot x^{q'-1}$$

$$= b \cdot 3 \cdot \left(\left(i_M + a \cdot x^{q'} \right)^2 + 2 \cdot \left(i_M + a \cdot x^{q'} \right) \cdot u_M + u_M^2 \right) \cdot a \cdot q' \cdot x^{q'-1}.$$

Hence, the instantaneous natural indirect effect satisfies

Fig. 2 Monotonic mediation relationships (assuming $q' \neq 0$ and $p' \neq 0$).

$$\frac{dNIE(x)}{dx} = b \cdot 3 \cdot \left(\left(i_M + a \cdot x^{q'} \right)^2 + \mathrm{E}\left(U_M^2 \right) \right) \cdot a \cdot q' \cdot x^{q'-1}.$$

We recommend using the special case with $p'' = 1$ due to its desirable properties. It provides a natural generalization of the linear case of models (1)-(3) to account for monotonicity. In what follows, $p'' = 1$ is assumed unless otherwise stated.

Adaptive regression

Methods are needed for estimating the relationships of models (4)-(5) and (9)-(10) and for assessing whether those relationships are non-constant in X or M and whether they are reasonably treated as linear in Y, X, and/or M or are distinctly nonlinear in any of those variables. We use adaptive regression modeling [34] for these purposes. Methods are needed as well for assessing whether the instantaneous natural indirect effect $\frac{dNIE(x)}{dx}$ is nonzero. This can be addressed as in linear mediation with bootstrapped CIs, but now computed for a grid of possible values x for X.

Knafl et al. [35] developed adaptive regression methods for nonlinear modeling of Poisson distributed count outcome variables Y. Knafl et al. [36] extended this to generalized linear modeling including adaptive linear regression with outcomes treated as normally distributed, as often used in mediation analyses (but they used these methods to address modeling of electronically monitored medication adherence data rather than mediation). How adaptive regression analyses are conducted is described next.

Adaptive regression methods use likelihood cross-validation (LCV) scores (as defined later) to evaluate and compare models. These scores generalize to contexts where estimation is based on maximizing likelihood-like functions such as extended quasi-likelihood functions [47]. Heuristic (i.e., rule-based) search techniques guided by LCV scores are used to identify effective fractional polynomial models [33] in primary predictor variables including predictors X and mediators M as needed for mediation. Fractional polynomials generalize standard polynomials, which use only nonnegative integer powers, to allow for the possibility of negative and fractional powers. Fractional polynomial models for continuous outcomes are linear in associated parameter vectors (consisting of the intercept and slopes for power transforms of predictors), and so they are linear models estimated using linear regression methods. However, these models are based on relationships that are in general nonlinear (or curvilinear) in the predictor X for models (4)-(5) and (9)-(10) as well as in the mediator M for models (5) and (10).

Power transforms $f(X, p)$ are defined for arbitrary real valued primary predictors X and powers p by setting $f(X, p)$ to X^p when $X > 0$, to 0 when $X = 0$, and to $\cos(\pi \cdot p) \cdot |X|^p$ when $X < 0$ where cos denotes the standard cosine function, π is the usual constant, and $|X|$ denotes the

Knafl *et al. BMC Medical Research Methodology* (2017) 17:45

Page 6 of 21

absolute value of X. Note that for $X < 0$, the sign of $f(X, p)$ oscillates between ± 1 as the power p varies. For simplicity of notation, $f(X, p)$ is denoted as X^p.

Likelihood cross-validation

LCV scores are computed by randomly partitioning observations into k disjoint sets called folds, calculating likelihoods for folds using parameter estimates computed from the remaining data in the complement of the fold, and combining these deleted likelihoods into a geometric mean deleted likelihood score. Formally, let $S = \{s : 1 \le s \le n\}$ denote the index set for the n subjects (or observations or cases) in the data set under analysis, $\boldsymbol{\theta}$ the vector of model parameters, and $L(\cdot; \boldsymbol{\theta})$ a likelihood function or a likelihood-like function (e.g., the extended quasi-likelihood function used with generalized linear models [47]). Randomly partition S into $k > 1$ disjoint nonempty subsets $S(h)$, called folds, for $h \in H = \{h : 1 \le h \le k\}$. The LCV score is defined as

$$\mathrm{LCV} = \prod_{h \in H} L^{1/n}(S(h); \boldsymbol{\theta}(S \backslash S(h)))$$

where $L(S(h); \boldsymbol{\theta}(S \backslash S(h)))$ denotes the joint likelihood for the observations with indexes in $S(h)$ using the maximum likelihood estimate $\boldsymbol{\theta}(S \backslash S(h))$ of the parameter vector $\boldsymbol{\theta}$ computed using the data in the complement $S \backslash S(h)$ of the fold $S(h)$. For a given data set, the same random fold assignment is used for all models so that the LCV scores for those models are comparable. LCV scores for multivariate data are normalized by the number of outcome measurements for all subjects rather than by the number of subjects.

Larger LCV scores indicate better models, but not necessarily substantially better models. This issue is assessed with LCV ratio tests, analogous to likelihood ratio tests in being based on the χ^2 distribution. LCV ratio tests are expressed in terms of a cutoff for a substantial (or distinct or significant) percent decrease in the LCV score, changing with the sample size. The formula for the cutoff is provided in Section 4.4.2 of Knafl and Ding [34] and in eq. (6) of Knafl et al. [36]. LCV ratio tests are more conservative than standard tests for zero coefficients (examples are provided in [48–50]) in the sense that removal from the model of a predictor with a significant slope does not always generate a substantial percent decrease in the LCV score. Consequently, LCV ratio tests are similar in effect to adjustments for multiple comparisons.

As an example, suppose that a model M_1 generates a score $\mathrm{LCV}(M_1)$ smaller than the score $\mathrm{LCV}(M_2)$ for another model M_2. If the percent decrease in these LCV scores, that is,

$$\frac{\mathrm{LCV}(M_2) - \mathrm{LCV}(M_1)}{\mathrm{LCV}(M_2)} \cdot 100\%,$$

is larger than the cutoff for a substantial percent decrease, then model M_2 substantially improves on model M_1. On the other hand, if the percent decrease is less than or equal to the cutoff, model M_1 is a competitive alternative to model M_2. If model M_1 is also simpler (e.g., based on fewer parameters or containing no versus some interactions) than model M_2, then it would be preferable to model M_2 as a parsimonious, competitive alternative.

Overview of the adaptive modeling process for a fixed outcome

Adaptive fractional polynomial models for a fixed outcome in terms of primary predictors are identified using a heuristic search process beginning with an expansion phase, systematically adding power transforms of those primary predictors into the model, followed by a contraction phase, systematically removing any extraneous power transforms and adjusting the powers of the remaining transforms to improve the LCV score. The search process is controlled by tolerance parameters indicating the change in LCV scores that can be tolerated for each step in the search process. For example, the contraction stopping tolerance parameter setting is based on a LCV ratio test so that the final selected model is parsimonious. The tolerance parameter settings change with the number of measurements in the sample, thereby adjusting the search process by the sample size. The full adaptive modeling process is formulated in Chapter 20 of [34]. That full process is needed for estimating models (4)-(5) with arbitrary nonlinearity. Estimation of the monotonic models of (9)-(10) requires a simpler search process as addressed later.

The adaptive modeling process can generate adaptive models for variances (or more generally dispersions [47]) as well as for means. For example, when M and/or Y are continuous, the omitted factors or errors U_M and U_Y can be treated as normally distributed (or at least approximately so) with non-constant variances that are functions of X, M, and/or some other primary predictors. The log of the variances is modeled as linear in the coefficient parameters for possibly power transformed primary predictors. Coefficient parameters for the means and variances are estimated jointly using maximum likelihood with likelihoods based on the normal distribution. For correlated continuous outcomes, for example, outcomes measured over clusters such as family members or patients of the same provider, likelihoods are based on the multivariate normal distribution.

Adaptive nonlinear moderation can be addressed simply by including interactions as primary predictors. More generally, the adaptive modeling process can automatically generate geometric combinations of two or more primary predictors, that is, products of power transforms of distinct subsets of the primary predictors using possibly different powers to transform those primary predictors. The subset of transforms in the geometric combinations and their powers are generated adaptively.

Searching for power transforms

A base model M_0 is expanded to include a transform of a predictor X as follows. Let $M_0(p)$ denote the model M_0 with the power transform X^p added to it. A grid search is conducted first to maximize $\text{LCV}(M_0(p))$ over an initial set of powers. By default, the initial powers $p = -3, -2.5, \cdots, -0.5, 0.5, \cdots, 2.5, 3$ are used, but this set can be adjusted if desired. The power 0 is purposely not considered. For the case with $X > 0$, the effect of $p = 0$ on M_0 depends on whether or not M_0 includes an intercept. When there is an intercept, the power 0 corresponds to the natural log transform (demonstrated by taking the limit as $p \to 0$); otherwise it corresponds to the constant transform adding in an intercept parameter to M_0. When X has zero or negative values, the effect is more complex. Not considering the power 0 avoids slowing the adaptive modeling process to check M_0 to see what the effect of that power is. In any case, powers p close to 0 approximate the appropriate case without having to check to see which one it is. By default, the smallest powers in absolute value that are considered are ± 0.0001, but this can be adjusted.

Let p_0 denote the power which maximizes $\text{LCV}(M_0(p))$ for this initial set of powers. When $p_0 = -3$, a search is conducted over powers $p = p_0 + i \cdot \delta$ for integer multiples i of $\delta = -1$ until

$$\text{LCV}(M_0(p_0 + (i+1)\cdot\delta)) < \text{LCV}(M_0(p_0 + i\cdot\delta)) \quad (11)$$
$$> \text{LCV}(M_0(p_0 + (i-1)\cdot\delta))$$

When $p_0 = 3$, δ is set to $+1$ instead. These produce powers $p_1 = p_0 + i \cdot \delta$ with $i \geq 0$ for these two cases. For $-3 < p_0 < 3$, $p_1 = p_0$. The choice of an initial power for X is given by p_1.

Next the choice of a power p_2 for changes in powers of $\delta = \pm 0.1$ is identified through a similar search on either side of p_1. Then, this is iterated searching around p_2 over changes $\delta = \pm 0.01$, then $\delta = \pm 0.001$, and so on. By default, changes of at most $\delta = \pm 0.0001$ are considered, but this can be adjusted. At any stage of this process, if the smallest of the three LCV scores analogous to those of inequality (11) compared to the largest of those three scores generates a percent decrease greater than the associated tolerance parameter (indicating these three LCV scores are not close to each other), continue the search with one more decimal digit; otherwise stop the process. When the process stops at the jth stage, the selected model is $M_0(p_j)$.

By default, the expansion stopping tolerance is set generously to 2.5 times the cutoff for a LCV ratio test, and so it is likely a transform of X would be added to the base model. However, it is also possible that the expansion might not add a transform of X to the base model, supporting the conclusion that X does not have an effect on the outcome Y after controlling for the transforms of the base model.

Identification of a transform of a predictor to add to a base model is a small part of the full adaptive modeling process. Multiple predictors need to be considered as part of the expansion; the best transform to remove from a base model needs to be determined as part of the contraction. With each such removal, the powers of the remaining transforms need to be adjusted to improve the LCV score. However, models (9)-(10) with $p'' = 1$ and p and q fixed require identification of only the single power q' for X in (9) and the single power p' for X in (10). An algorithm for identifying choices for these two powers is defined later.

For the general adaptive modeling process, tests for zero coefficients for transforms in selected models are inappropriate to conduct since these tests are usually significant. Due to the contraction heuristics, the removal of each transform of the selected model generates a substantial percent decrease in the LCV score. However, this is not the case for models (9)-(10) based on single transforms of X. While the selected power transform for X in one of these models generates an optimal LCV score, the LCV score for the associated constant base model might not generate a substantial percent decrease in the LCV score, and so the slope for that selected power transform might not always be significantly nonzero. However, a LCV ratio test can also be used to assess the impact of including X in the model versus not including it and is likely to be more conservative than the test for a zero slope for X.

Setting the number of folds

Knafl et al. [35] used 10 folds for estimating nonlinear individual-subject medication adherence curves on the recommendation of Kohavi [51]. However, the data sets they used had limited sample sizes at most 100. The choice of the number k of folds may need to be adjusted for different sample sizes. Knafl and Grey [52] investigated this issue for exploratory factor analysis models. They found that for three different sets of items the same number of factors was selected by maximizing LCV scores as long as the value of k was not too small. Consequently, they recommended using the first local maximum in k over multiples of 5 folds for some benchmark analysis, in their case the selection of the number of factors extracted through maximum likelihood. This choice balances the need for a sufficiently large number of folds while limiting the amount of computation. Section 2.8 of [34] provides a more complete assessment of this issue.

The composite mediation model

Models (9)-(10) with $p'' = 1$ can be combined into a single bivariate outcome model as follows. With superscript T denoting the transpose operator, let $\boldsymbol{Y}' = (M^q \ Y^p)^T$, $\boldsymbol{I}_1 = (1 \ 0)^T$, $\boldsymbol{I}_2 = (0 \ 1)^T$, $\boldsymbol{X}_1 = (X \ 0)^T$, $\boldsymbol{X}_2 = (0 \ X)^T$, $\boldsymbol{M} = (0 \ M)^T$, and $\boldsymbol{U} = (\boldsymbol{U_M} \ \boldsymbol{U_Y})^T$. The model

$$\boldsymbol{Y}' = \beta_1 \cdot \boldsymbol{I}_1 + \beta_2 \cdot \boldsymbol{I}_2 + \beta_3 \cdot \boldsymbol{X}_1^{q'} + \beta_4 \cdot \boldsymbol{X}_2^{p'} + \beta_5 \cdot \boldsymbol{M}^q + \boldsymbol{U}, \quad (12)$$

Knafl *et al. BMC Medical Research Methodology* (2017) 17:45

Page 8 of 21

where a power transform of a vector is the vector of its entries transformed by that power, satisfies $\beta_1 = i_M$, $\beta_2 = i_Y$, $\beta_3 = a$, $\beta_4 = c'$, and $\beta_5 = b$, and so provides a single model for the five coefficient parameters for the means of models (9)-(10). Assume that U is bivariate normally distributed with covariance matrix Σ, thereby allowing for possibly correlated omitted factors or errors. Model (12) is a path model nonlinear in the outcome Y, the mediator M, and the predictor X with Y and M measured with error and X measured without error.

Let θ be the vector of model parameters, including β_j $1 \le j \le 5$ and all the parameters for modeling the covariance matrix Σ. Using subscripts s for $s \in S$ as defined earlier, the likelihood $L(S; \theta)$ satisfies $\log L(S; \theta) = \prod_{s \in S} \ell_s$ where

$$\ell_s = -{}^1/_2 \cdot \boldsymbol{u}_s^T \cdot \boldsymbol{\Sigma}_s^{-1} \cdot \boldsymbol{u}_s - {}^1/_2 \cdot \log(|\boldsymbol{\Sigma}_s|) - {}^1/_2 \cdot 2 \cdot \log(2 \cdot \pi),$$

$$\boldsymbol{\mu}_s = \beta_1 \cdot \boldsymbol{I}_1 + \beta_2 \cdot \boldsymbol{I}_2 + \beta_3 \cdot \boldsymbol{X}_{1,s}^{q'} + \beta_4 \cdot \boldsymbol{X}_{2,s}^{p'} + \beta_5 \cdot \boldsymbol{M}_s^q,$$

Σ_s is the covariance matrix for the sth observation, $|\Sigma_s|$ its determinant, and $\boldsymbol{u}_s = \boldsymbol{Y'}_s - \boldsymbol{\mu}_s$ the associated residual vector. For model (12), there is only one correlation parameter ρ that is the same for all s, but each of the variances for $U_{M,s}$ and for $U_{Y,s}$ might change with s or be the same for all s.

When the omitted factors or errors are independent, that is, when $\rho = 0$,

$$L(S; \theta) = L_M(S; \theta) \cdot L_Y(S; \theta)$$

separates into two terms corresponding to the likelihoods $L_M(S; \theta)$ and $L_Y(S; \theta)$ for models (9) and (10), respectively. The LCV score for model (12) also separates, but into

$$\text{LCV} = \text{LCV}_M^{1/2} \cdot \text{LCV}_Y^{1/2}$$

where LCV_M and LCV_Y are LCV scores for models (9) and (10), respectively. This holds because the LCV score for model (12) is normalized by the number $2 \cdot n$ of outcome measurements while LCV_M and LCV_Y each are normalized by the number n of subjects. If q is fixed (e.g., with $q = 1$ so models (9)-(10) are linear in M), maximum likelihood estimation and adaptive modeling of models (9) and (10) separately generate the same results as for modeling them in combination using model (12), assuming consistent fold assignments for these two cases. However, this only holds when q is fixed and $\rho = 0$. Even when $\rho = 0$, identification of an appropriate value for q requires comparing LCV scores for the composite model (12) since the same value of q is used in its submodels (9)-(10).

Identifying power transforms for X *in the composite model*
For fixed choices for the powers p and q, the adaptive expansion process is constrained to generate a single transform for X_1 and for X_2 as follows. Let M_0 be the base model with means based only on I_1 and I_2 along with a fixed choice for the covariance matrix of U_M and U_Y. Let $M_0(q'_1)$ be this base model expanded to include a power transform of X_1 with selected power q'_1 and $M_0(p'_1)$ expanded to include a power transform of X_2 with selected power p'_1. If

$$\text{LCV}(M_0(p'_1)) > \text{LCV}(M_0(q'_1)),$$

then expand $M_0(p'_1)$ to include a power transform of X_1 with power q'_2, possibly different than q'_1. Otherwise, expand $M_0(q'_1)$ to include a power transform of X_2 with power p'_2, possibly different than p'_1. This is the standard adaptive expansion process constrained to include a single power transform of primary predictors X_1 and X_2 for the means.

Since power transforms are added to the model without adjusting the powers of previously added transforms, there might be an improvement if these powers are adjusted. The formal power transform process is defined in Section 20.4.4 of [34]. Informally, each power is adjusted using the power adjustment process described earlier but starting at its current value holding the other powers fixed. Stop if no adjusted powers provide an improvement in the LCV score. Otherwise, continue the process using the adjusted power generating the best improvement in the LCV score. However, since the generated model (12) for fixed p and q has only two transforms, the improvement if any is unlikely to be substantial. Consequently, power adjustment is treated as an optional feature of the adaptive monotonic mediation process.

The model for the variances (technically, the model for the log of the variances) based on I_1 and I_2 results in separate but constant variances for U_M and U_Y. The associated covariance matrix also allowing for the single general correlation ρ is called compound symmetry heterogeneous (CSH). Non-constant variances can be generated by considering X_1, X_2, and/or M as primary predictors for modeling the variances. The full adaptive modeling process can be used for modeling the variances since these need not be monotonic. This is achieved by including I_1 and I_2 in the base model for the variances, not restricting the expansion of the model for the variances, and restricting the contraction to contract only transforms from the model for the variances. By default, the contraction considers removal of the intercepts corresponding to I_1 and I_2 in the model for the variances, but this can be overridden.

Covariates can be included in model (12). For each covariate Z, include $\boldsymbol{Z}_1 = (Z\ 0)^T$ and $\boldsymbol{Z}_2 = (0\ Z)^T$ as primary

predictors for the means and/or variances, respectively, addressing models (9)-(10). The full adaptive modeling process can be used to generate power transforms of Z_1 and Z_2 for modeling the means and/or variances.

Power-adjusted likelihood cross-validation

Assume that $Y > 0$ so that transforms Y^p are well-defined for all p. Model (10) for Y^p for a fixed power p can be estimated using existing adaptive regression methods as described above. Standard LCV scores for these models, not accounting for the power p, as defined earlier are based on the normal density function. The power p can be chosen by maximizing an alternative power-adjusted LCV score that also accounts for power transformation of the outcome Y using the power-adjusted likelihood function as defined next.

If Y^p is normally distributed for $p > 0$, the distribution function $F(y; p, \boldsymbol{\theta})$ for Y satisfies

$$F(y; p, \boldsymbol{\theta}) = P(Y \leq y; p, \boldsymbol{\theta}) = P(Y^p \leq y^p; \boldsymbol{\theta})$$

where $\boldsymbol{\theta}$ is the vector of model parameters. Consequently, the power-adjusted density function $f(y; p, \boldsymbol{\theta})$ for Y satisfies

$$f(y; p, \boldsymbol{\theta}) = \frac{dP(Y \leq y; p, \boldsymbol{\theta})}{dy} = p \cdot y^{p-1} \frac{dP(Y^p \leq y^p; \boldsymbol{\theta})}{dy}$$

where $\frac{dP(Y^p \leq y^p; \boldsymbol{\theta})}{dy}$ is the usual univariate normal density function $\varphi(v, \boldsymbol{\theta})$ evaluated at $v = y^p$ with mean and variance based on the parameter vector $\boldsymbol{\theta}$. When $p < 0$,

$$F(y; p, \boldsymbol{\theta}) = P(Y \leq y; p, \boldsymbol{\theta}) = P(Y^p \geq y^p; \boldsymbol{\theta})$$

and the power-adjusted density function $f(y; p, \boldsymbol{\theta})$ for Y satisfies

$$f(y; p, \boldsymbol{\theta}) = \frac{dP(Y \leq y; p, \boldsymbol{\theta})}{dy} = -p \cdot y^{p-1} \cdot \varphi(y^p; \boldsymbol{\theta}).$$

Thus, for $p \neq 0$,

$$f(y; p, \boldsymbol{\theta}) = |p| \cdot y^{p-1} \cdot \varphi(y^p; \boldsymbol{\theta}).$$

A similar argument for $p = 0$ corresponding to the natural log transform gives

$$f(y; p, \boldsymbol{\theta}) = y^{-1} \cdot \varphi(\log y; \boldsymbol{\theta}).$$

The power-adjusted LCV score is defined as

$$\mathrm{LCV}(p) = \prod_{h \in H} f^{1/n}(S(h); p, \boldsymbol{\theta}(S \backslash S(h); p))$$

where $f(S(h); p, \boldsymbol{\theta}(S \backslash S(h); p))$ denotes the joint power-adjusted likelihood for the subjects with indexes in fold $S(h)$ computed with estimates $\boldsymbol{\theta}(S \backslash S(h); p)$ of the parameter vector $\boldsymbol{\theta}$ generated using the data in the complement $S \backslash S(h)$ of the fold $S(h)$ and with the outcome Y transformed to Y^p. The LCV(p)

score can be maximized in the power p using a grid search to choose an appropriate power transform for the outcome.

The assumption that $Y > 0$ can be relaxed by extending Y^p in the same way as for X^p given earlier, that is, by setting it to 0 when $Y = 0$ and to $\cos(\pi \cdot p) \cdot |Y|^p$ when $Y < 0$. Then, the sign is not always reversed for the case $Y < 0$, affecting the computation of $f(y; p, \boldsymbol{\theta})$. However, the derivative $f(0; p, \boldsymbol{\theta})$ is not always well-defined. For example, when $p < 1$ and $y > 0$,

$$f(y; p, \boldsymbol{\theta}) = p \cdot y^{p-1} \cdot \varphi(y^p; \boldsymbol{\theta}) \uparrow \infty \text{ as } y \downarrow 0.$$

It seems better to add a constant to Y to make it positive valued, which is the approach recommended by Royston and Altman [33] for transforming non-positive predictors.

The univariate outcome transformation process can be applied as well to model (9). The formulation also extends readily to multivariate data. When those data are based on repeatedly measuring the same outcome Y at different times or over different conditions (such as members of the same family or patients of the same provider), it is reasonable to transform each such outcome measurement with the same power p. However, model (12) requires consideration of different powers p and q for Y and M with associated power-adjusted LCV scores LCV(p, q).

Identification of appropriate choices for p and q can be achieved by starting with $p = q = 1$ and using grid searches to adjust p with $q = 1$ fixed giving LCV(p_1, 1) and to adjust q with $p = 1$ fixed giving LCV($1, q_1$). If LCV(p_1, 1) > LCV($1, q_1$), use a grid search in q with $p = p_1$; otherwise use a grid search in p with $q = q_1$. This generates powers p_2 and q_2. In example analyses, the grid searches are first conducted over changes of ±0.5 generating the powers p_2 and q_2, then over changes of ±0.1 starting at the powers p_2 and q_2 generating the powers p_3 and q_3, and then stops identifying powers for Y and M to within one decimal digit.

Monotonic mediation analysis

In what follows, unless otherwise stated, base models for means and covariances are based on the predictors I_1 and I_2, and so with only intercepts i_M and i_Y for the means along with CSH covariance structure. Also assume $p'' = 1$ unless otherwise indicated.

1. Selecting the number k of folds.
   The benchmark analysis is the generation of model (12) constrained so that $p = q = 1$, that is, with untransformed M and Y, by adaptively expanding the model for the means in X_1 and X_2 limiting the number of power transforms for each of these predictors to one. The number of folds to use in all subsequent analyses is the one generating the first local maximum in the standard LCV score over multiples of 5.

Knafl *et al. BMC Medical Research Methodology* (2017) 17:45

Page 10 of 21

2. Selecting the powers p and q.
   Use the search through alternative values for p and q based on power-adjusted LCV scores LCV(p,q) described earlier to generate the full model (12) without changing the CSH covariance structure. Use the selected powers p and q in subsequent analyses unless otherwise indicated.

3. Assessing the need for transforming M and Y.
   Use a LCV ratio test to compare LCV(1,1) (the same as its standard LCV score) generated in Step 1 to LCV(p,q) generated in Step 2.

4. Assessing the need for transforming X.
   For given values of p and q, use a LCV ratio test to compare LCV(p,q) as generated in Step 2, with X_1 and X_2 adaptively transformed, to the LCV(p,q) score for the associated model linear in X_1 and X_2. When the power $q' = 1$, the instantaneous natural indirect effect $\frac{dNIE(x)}{dx} = a \cdot b$ is constant. Whether the instantaneous natural indirect effect is reasonably treated as constant can be assessed with a LCV ratio test comparing a given model with $q' \neq 1$ to that model adjusted to satisfy $q' = 1$.

5. Assessing mediation relationships.
   In the linear mediation case of models (1)-(3), mediation requires significantly nonzero slopes a and b. In the monotonic mediation context of model (12), these issues can be addressed with LCV ratio tests. To assess for a dependence of M^q on X in the component model (9), compare the LCV(p,q) score for the full model (12) to the associated model with X_1 removed (or with $a = 0$) and the transform for X_2 adaptively generated. To assess the dependence of Y^p on M^q in the component model (10) with $p'' = 1$, compare the LCV(p,q) score for the full model (12) to the associated model with M^q removed (or with $b = 0$) and the transforms for X_1 and X_2 adaptively generated. It is also possible to assess the dependence of Y^q on X in the component model (10) by comparing the LCV(p,q) score for the full model (12) to the associated model with X_2 removed (or with $c' = 0$) and the transform for X_1 adaptively generated.

6. Considering non-constant variances.
   With base model the adaptive model (12) with CSH covariance structure generated in Step 2 with initial powers p' and q' for X_1 and X_2, adaptively expand and then contract the model for the variances in X_1, X_2, and M, allowing for possible adjustment of the powers p' and q' of the base model as part of the contraction, but leaving M^q in the base model untransformed (so that $p'' = 1$).

7. Considering covariates.
   Covariates can be considered for inclusion in the component models (9)-(10) of model (12). With base

model the adaptive model of Step 2, adaptively expand and then contract the model for the means and the variances in Z_1 and Z_2 for all covariates Z. As part of the contraction, allow for possible adjustment of the powers p' and q' of the base model, leave M^q in the base model untransformed (so that $p'' = 1$), and only contract transforms of covariates from the model for the means. This can be combined with Step 6 to allow the variances to depend as well on X_1, X_2, and M.

8. Assessing $\rho = 0$.
   For any model generated earlier, compare its LCV score using a LCV ratio test to the associated model constrained to satisfy $\rho = 0$.

9. Assessing $p'' = 1$
   With base model the adaptive model of Step 2, adaptively retransform the model allowing the powers of the single transforms of X_1, X_2, and M^q to be changed to improve the LCV score. The assumption $p'' = 1$ is supported if the associated model generated with $p'' = 1$ in Step 2 is a competitive alternative to this latter model. Similar assessments can be conducted allowing for non-constant variances as in Step 6, for covariates as in Step 7, and/or $\rho = 0$ as in Step 8.

10. Assessment of Model Assumptions
    Yuan and MacKinnon [26] provide a detailed discussion of the impact of violations of the constant variances and normality assumptions of standard regression models. Modeling of variances can address the first problem. Data transformation, applied to outcomes and/or predictors, as considered here can sometimes resolve normality problems, but not always. Consequently, model assumption assessments are important to conduct in general regression contexts including the special case of mediation analyses, whether treated as linear or nonlinear. Should data transformation not resolve such problems, then quantile regression methods as described by Yuan and MacKinnon [26] are more appropriate to use. However, if data transformation resolves model assumption problems, then normality-based methods are optimal [26] when applied to the transformed data and so would likely generate more efficient and powerful estimates. In the case of mediation analyses, the use of bootstrapped CIs on indirect effects circumvents distributional assumption problems for parametric estimates of those effects. However, as demonstrated by the simulations of [26], bootstrap methods cannot fully address distributional assumption problems for the data. Moreover, bootstrapped CIs are likely to be relatively narrower, and so more precise, when data are transformed to be as close as possible to normal, than when untransformed. The example analyses

reported later provide two examples supporting this conclusion. This, of course, assumes that the indirect effect has been consistently estimated so that the true value is in the confidence interval.

For composite model (12), the constant variances assumption can be addressed with a LCV ratio test comparing constant and non-constant variances models to assess whether variances are reasonably close to constant or are distinctly non-constant. If a sufficiently broad set of primary predictors for the variances are considered, the associated non-constant variances model should provide an appropriate depiction of the variances, thereby relaxing the constant variances assumption if necessary. These variances estimates combined with the estimated correlation provide estimates $\Sigma_s(S)$ of the covariance matrices for observed outcome vectors Y'_s for subjects with indexes s in the set S. Associated residual vectors are $u_s(S) = Y'_s - \mu_s(S)$ where $\mu_s(S)$ are estimated mean vectors for subjects s based on X_{1s}, X_{2s}, M_s, and possibly covariates. Associated standardized or scaled residuals are given by $stdu_s(S) = (V_s^T(S))^{-1} \cdot u_s(S)$ where $V_s(S)$ is the square root of $\Sigma_s(S)$ determined by its Cholesky decomposition. The combined standardized residuals over all s can be used to assess the normality assumption by visually checking for linearity in the associated normal (probability) plot and with the Shapiro-Wilk test for normality of the standardized residuals.

11. Assessing natural indirect effects

Once an appropriate choice for composite model (12) with $p'' = 1$ has been identified, possibly including non-constant variances and/or covariates, the assessment of whether the instantaneous natural indirect effect function $\frac{dNIE(x)}{dx}$ for this model is nonzero needs to be assessed. This can be addressed with bootstrapped CIs [21, 29, 40, 41], but computed for a grid of possible values x for the predictor X. For models with a constant instantaneous natural indirect effect (i.e., with $q' = 1$), only one value for X need be considered. The bias-corrected version [42] is recommended by MacKinnon et al. [40], and so is used in example analyses unless otherwise indicated. All reported CIs are based on 1,000 resamples.

The powers p, q, p', and q' for a composite model (12) as well as powers for all covariate predictors and variance predictors are held fixed with associated slope parameters estimated for resamples of the composite data. The generated 95% CI for the instantaneous natural indirect effect function at each nonzero value of x has lower and upper bounds

$$L(x) = b_L \cdot a_L \cdot q' \cdot x^{q'-1} < U(x) = b_U \cdot a_U \cdot q' \cdot x^{q'-1}.$$

Define the normalized width W of these intervals as

$$W = \frac{\max(|L(x)|, |U(x)|) - \min(|L(x)|, |U(x)|)}{\max(|L(x)|, |U(x)|)}$$
$$= 1 - \frac{\min(|b_L \cdot a_L|, |b_U \cdot a_U|)}{\max(|b_L \cdot a_L|, |b_U \cdot a_U|)},$$

which is constant in nonzero x with a value between 0 and 1. Models for the data generating smaller values for W provide more precise predictions of the instantaneous natural indirect effect function.

Moderated monotonic mediation

One of the covariates Z might be considered as a moderator. There are a variety of ways that models (9)-(10) can be adjusted to accommodate moderation. For example, Preacher, Rucker, and Hayes [53] propose five alternatives. Under their fifth alternative, Z moderates the effect of X on M, X on Y, and M on Y. Under this alternative, models (9)-(10) become

$$M^q = \sum_{z \in \mathbb{Z}} \left(i_M(z) + a(z) \cdot X^{q'(z)} \right) \cdot I(Z = z) + U_M \tag{13}$$

$$Y^p = \sum_{z \in \mathbb{Z}} \left(i_Y(z) + c'(z) \cdot X^{p'(z)} + b(z) \cdot M^{q \cdot p''(z)} \right) \cdot I(Z = z) + U_Y \tag{14}$$

possibly with other covariates included where $\mathbb{Z}$ is the set of all possible values z for Z and $I(Z = z)$ the indicator for $Z = z$ (i.e., it equals 1 when $Z = z$ and 0 otherwise). The dependence of M on X and Y on X and M have been defined separately for the values z of Z taking an analysis of variance approach. This formulation allows associated intercepts, slopes, and powers to change with the values z of Z while preserving monotonicity. However, it requires estimation of model parameters for each value z of Z. This requirement is reasonable for moderators Z with discrete numbers of possible values, but problematic for continuous moderators Z with many possible values z but sparse numbers of observations for some values z. In this latter case, the moderator Z could be replaced by a split based on its tertiles, quartiles, etc.

Under (13)-(14), the instantaneous natural direct effect $c'(z) \cdot p'(z) \cdot X^{p'(z)-1}$ (assuming for simplicity that $p'(z) \neq 0$) changes with the values z of Z. When $p''(z) \equiv 1$, the instantaneous natural indirect effect $b(z) \cdot a(z) \cdot q'(z) \cdot X^{(q'(z)-1)}$ (assuming for simplicity that $q'(z) \neq 0$) also changes with the values z of Z. The associated normalized widths $W(z)$ change with z as well.

Model (12) generalizes to

$$Y' = \sum_{z \in \mathbb{Z}} H(z) \cdot I(Z = z) + U \tag{15}$$

where

$$H(z) = \beta_1(z) \cdot I_1 + \beta_2(z) \cdot I_2 + \beta_3(z) \cdot X_1^{q'(z)}$$
$$+ \beta_4(z) \cdot X_2^{p'(z)} + \beta_5(z) \cdot M^q.$$

Knafl *et al. BMC Medical Research Methodology* (2017) 17:45

Page 12 of 21

The adaptive modeling process can be used to identify the powers $q'(z)$ and $p'(z)$ for all z combined.

The individual moderation components of model (15) can be assessed by comparing model (15) to the model with each of those moderation component removed using LCV ratio tests. Specifically, moderation of the effect of X on M in model (13) can be assessed by replacing $\sum_{z \in Z} \beta_3(z) \cdot X_1^{q'(z)} \cdot I(Z = z)$ in (15) with $\beta_3 \cdot X_1^{q'}$. Moderation of the effect of X on Y in model (14) can be assessed by replacing $\sum_{z \in Z} \beta_4(z) \cdot X_2^{p'(z)} \cdot I(Z = z)$ in (15) with $\beta_4 \cdot X_2^{p'}$. Moderation of the effect of M on Y in model (14) can be assessed by replacing $\sum_{z \in Z} \beta_5(z) \cdot M^q \cdot I(Z = z)$ in (15) with $\beta_5 \cdot M^q$.

It is also possible to test effects for specific values z' of Z. Specifically, the effect of X on M in model (13) for the value z' of Z can be assessed by replacing $\sum_{z \in Z} \beta_3(z) \cdot X_1^{q'(z)} \cdot I(Z = z)$ in (15) with $\sum_{z \in Z \setminus \{z'\}} \beta_3(z) \cdot X_1^{q'(z)} \cdot I(Z = z)$. The effect of X on Y in model (14) for the value z' of Z can be assessed by replacing $\sum_{z \in Z} \beta_4(z) \cdot X_2^{p'(z)} \cdot I(Z = z)$ in (15) with $\sum_{z \in Z \setminus \{z'\}} \beta_4(z) \cdot X_2^{p'(z)} \cdot I(Z = z)$. The effect of M on Y in model (14) for the value z' of Z can be assessed by replacing $\sum_{z \in Z} \beta_5(z) \cdot M^q \cdot I(Z = z)$ in (15) with $\sum_{z \in Z \setminus \{z'\}} \beta_5(z) \cdot M^q \cdot I(Z = z)$.

Data on family management of childhood chronic conditions
Example analyses are reported later using a subset of data from a cross-sectional study on family management of childhood chronic conditions [54] reported by 187 partnered mothers. General family functioning is measured using the General Functioning Scale of the McMaster Family Assessment Device [55], coded so that larger values indicate better family functioning with range 1–4. Difficulty managing the child's condition is measured by the family life difficulty scale of the Family Management Measure [54], coded so that larger scores mean more difficulty. This scale measures the extent to which having a child with a chronic condition makes family life more difficult. Child adaptation, in terms of the intensity of the child's conduct-disordered behavior, is measured using the Eyberg Child Behavior Inventory [56], coded so that larger values indicate better child adaptation or less conduct-disordered behavior.

Example analyses use these family management data to demonstrate nonlinear mediation analyses by considering mediation of the impact of family functioning X on child adaptation Y by difficulty M in managing the condition. The cutoff for a substantial percent decrease for these data using models (12) and (15) with $374 = 2 \cdot 187$ measurements is 0.51%. Example analyses assume $p'' = 1$ or $p''(z) = 1$ for all values z of a moderator Z unless otherwise stated.

The proposed mediation relationships for these observational data can be justified on the following basis. General family functioning would be developed by a family prior to the diagnosis of the child's chronic condition, which would affect how difficult that chronic condition makes family life which would then affect the child's adaption to the condition. However, the purpose of these analyses is to provide example mediation analyses not to establish mediation in this context.

Simulated mediation data
Data were simulated for 101 observations with equally spaced values for the predictor X_{sim} between 1 and 10 (i.e., 0.09 units apart) with mediator $M_{sim} = 1 + X_{sim} + U_M$, outcome $Y_{sim} = Y'^{0.4}_{sim}$, and

$$Y'_{sim} = \frac{5 + X_{sim} + M_{sim} + U_Y}{25},$$

where U_M and U_Y are independent standard normal random variables. Y_{sim} was computed by raising Y'_{sim} to the power 2.5; the normalizing value 25 used in computing Y'_{sim} was chosen to be the smallest integer value larger than the maximum generated value for the unnormalized Y'_{sim} values. Normalizing the values of Y'_{sim} avoids generating very large values for Y_{sim}. The true values for the powers are $p = 0.4$, $q = 1$, $p' = 1$, and $q' = 1$ with true constant instantaneous natural indirect effect $a \cdot b = 1 \cdot 1 / 25 = 0.04$. The cutoff for a substantial percent decrease for these data using model (12) with $202 = 2 \cdot 101$ measurements is 0.95%.

Results
Analyses at the beginning of this section consider mediation of the effect of family functioning on child adaptation to a childhood chronic condition by the difficulty in managing that condition. Intercept parameters for the means and variances are constrained in all analyses not to be removed as part of contractions. In computing LCV scores, Y^p and M^q measurements for the same mother are randomly assigned to the same fold. Analyses are also conducted at the end of this section using the simulated mediation data.

Selecting the number of folds
For model (12) applied to the family management data with $p = q = 1$ and CSH covariance structure, the adaptive model in X_1 and X_2 generates a first local maximum at $k = 10$ with 10-fold LCV score 0.015024 and selected powers $p' = -3$ and $q' = 2$. Using $k = 5$, the selected powers are $p' = 3$ and $q' = 1.5$ with 10-fold LCV 0.015018 and insubstantial percent decrease in the LCV score compared to the model selected with $k = 10$ of 0.04% (i.e., less than the cutoff of 0.51% for the data). Using $k = 15$, the selected powers are $p' = -1$ and $q' = 1.5$ with 10-fold LCV 0.015005 and insubstantial percent decrease in the LCV score of 0.13%. Consequently, the generated model is reasonably

Knafl *et al. BMC Medical Research Methodology* (2017) 17:45

Page 13 of 21

robust to the choice of the number of folds. Subsequent analyses all use $k = 10$ folds for computing standard as well as power-adjusted LCV scores.

Using 10 folds, the number of measurements per fold ranges from 26 to 56 for 13 to 28 mothers. Consequently, fold complements contain at least 318 or 85.0% of the 374 available measurements, and so deleted parameter estimates should be reasonably reliable.

Selecting the powers *p* and *q* for *Y* and *M*
For model (12) with CSH covariance structure, the adaptive search for models in X_1 and X_2 with varying choices for p and q, first selects the powers $p = 1$ and $q = 0$ (i.e., the log transform) over changes in these powers of ± 0.5 with $LCV(1, 0) = 0.015683$ and then the powers $p = 1.3$ and $q = 0$ over changes in these powers of ± 0.1 with $LCV(1.3,0) = 0.015776$. For untransformed Y and M with $p = q = 1$, $LCV(1, 1) = 0.015024$ (the same as its standard LCV score reported above), and so the percent decrease is substantial at 4.77%. Consequently, the mediation relationships are distinctly nonlinear in Y and M. Model (12) for the standard linear mediation model with $p = q = p' = q' = 1$ has even smaller $LCV(1,1)$ score 0.014958, and so is substantially improved upon by consideration of monotonicity.

Furthermore, the adaptively generated model with Y transformed and M held untransformed, that is, with $q = 1$, generates the powers $p = 1.4$, $p' = -5$, and $q' = 2$ with $LCV(1.4,1)$ score 0.015143 and substantial percent decrease compared to the model with $p = 1.3$ and $q = 0$ of 4.01%. Consequently, relationships (9)-(10) are distinctly nonlinear in M. Moreover, the adaptively generated model with Y untransformed, that is, with $p = 1$, generates the powers $q = 0$, $p' = -2$, and $q' = 2.4$ with $LCV(1,0)$ score 0.15683 and substantial percent decrease compared to the model with $p = 1.3$ and $q = 0$ of 0.59%. Hence, relationship (10) is distinctly nonlinear in Y.

Assessing $\rho = 0$
Using $p = 1.3$, $q = 0$, $p' = -2.5$, and $q' = 3.5$ as selected with the CSH covariance structure, the estimated correlation is –0.08. Rerunning this model with $\rho = 0$, the $LCV(1.3,0)$ score is 0.015804. Using $p = 1.3$ and $q = 0$ as selected with CSH covariance structure, but with $\rho = 0$, the adaptively generated model (12) has somewhat different powers $p' = -3$ and $q' = 3$ and $LCV(1.3,0)$ score that also rounds to 0.015804. Since this score is larger than the $LCV(1.3,0)$ score under CSH, the omitted variables or errors U_M and U_Y are reasonably treated as independent. Using the values of p and q selected under CSH reduces the computations compared to identifying adaptive values for p and q under $\rho = 0$. However, in this case, the same powers $p = 1.3$ and $q = 0$ are adaptively identified with $\rho = 0$ starting from $p = q = 1$. An alternative approach would be to start the search for p and q with $\rho = 0$ at the values generated for CSH while

searching over grids of ± 0.1 to reduce the computations. This also generates the same solution $p = 1.3$ and $q = 0$.

For the case $p = q = 1$ with selected powers $p' = -3$ and $q' = 2$, a similar result holds. The estimated correlation is -0.05 and the model with $\rho = 0$ has LCV score 0.015043, larger than the score 0.015024 reported earlier for the associated model with CSH covariance structure. Subsequent analyses use $\rho = 0$ since it provides an improvement in these two cases.

Assessing $a = 0$
The adaptively generated model with $p = 1.3$, $q = 0$, $\rho = 0$, and constrained not to include a transform of X_1 in the model for the means has model for the means depending on X_2 transformed by the power $p' = -3$ with $LCV(1.3,0)$ score 0.015077 and substantial percent decrease 4.60% compared to the model also including a transform of X_1. Consequently, transformed difficulty $\log M$ in model (9) is reasonably assumed to depend substantially on family functioning X.

Assessing $b = 0$
The adaptively generated model with $p = 1.3$, $q = 0$, $\rho = 0$, and constrained not to include the transform M^q in the model for the means has model for the means depending on X_1 and X_2 transformed by the powers $p' = -0.4$ and $q' = 3$, respectively, with $LCV(1.3,0)$ and substantial percent decrease 1.07% compared to the model also including the transform of M^q. Consequently, transformed adaptation $Y^{1.3}$ in model (10) is reasonably assumed to depend substantially on transformed difficulty $\log M$ corresponding to $q = 0$.

Assessing $c' = 0$
The adaptively generated model with $p = 1.3$, $q = 0$, $\rho = 0$, and constrained not to include a transform of X_2 in the model for the means has model for the means depending on X_1 transformed by the power $q' = 3$ with $LCV(1.3,0)$ and substantial percent decrease 0.57% compared to the model also including a transform of X_2. Consequently, in model (10), transformed adaptation $Y^{1.3}$ is reasonably assumed to depend substantially on family functioning X, and so the instantaneous natural direct effect is substantial.

Assessing $p'' = 1$
Using the model (12) selected with $p = 1.3$, $q = 0$, and $\rho = 0$ (i.e., with $p' = -3$ and $q' = 3$), an adaptive transformation of this model allowing for $p'' \neq 1$ (and also possible adjustments to p' and q') is the same model generated with $p'' = 1$. Consequently, the simplifying assumption $p'' = 1$ is reasonable in this case.

Assessing non-constant variances

As described earlier, adaptive models can be generated allowing for variances of the omitted factors or errors U_M and U_Y to depend on transforms of X_1, X_2, and M. Starting from the adaptively generated model for $p = 1.3$ and $q = 0$ with $\rho = 0$ (i.e., with $p' = -3$ and $q' = 3$), the expansion adds in transforms of these predictors to the model for the variances, but the contraction removes all of them, leaving the base constant variances model. This result indicates that the variances are reasonably treated as constant in X_1, X_2, and M.

Also considering a covariate

The family management study enrolled parents of children with a variety of chronic conditions. One childhood chronic condition type was Crohn's disease or a bowel disorder with 54 (28.9%) children having this condition. The indicator Z for having this condition can be considered as a possible covariate for models (9)-(10). Adaptive modeling starts with the adaptively generated model for $p = 1.3$, $q = 0$, and $\rho = 0$ (i.e., with $p' = -3$ and $q' = 3$) along with constant variances (and so based on I_1 and I_2). The model for the means is expanded considering the indicators Z_1 and Z_2 (as defined similarly to X_1 and X_2) while the model for the variances is expanded considering arbitrary transforms of X_1, X_2, and M along with Z_1 and Z_2. The contraction is constrained so that M^q is not retransformed in the model for the means (so $p'' = 1$) while X_1 and X_2 are not removed from the model for the means, but associated powers are allowed to be changed.

The generated model has the same powers $p' = -3$ and $q' = 3$ as without consideration of Z_1 and Z_2 along with the covariate Z_2 added to both models for the means and the variances and no transforms of X_1, X_2, and M in the model for the variances. The LCV(1.3,0) score is 0.016062, which is a substantial improvement on the score 0.015804 for the associated model not considering covariates with percent decrease 1.61%. Consequently, the indicator for having Crohn's disease or a bowel disorder substantially influences the means

and variances for model (10), but not for model (9) (since only Z_2 is included in the generated model and not Z_1).

Model assumptions

A standard linear mediation analysis, that is, with model (12) based on $p = q = p' = q' = 1$ and $\rho = 0$, generates the standardized residuals plotted in Fig. 3. While this plot is reasonably close to linear for most of the data, there are exceptions at the low and high ends of the plot. There are also three outliers (i.e., with values outside of ± 3) with standardized residual values of -3.44, 3.03, and 3.07. The Shapiro-Wilk test for normality of the standardized residuals is significant at $p = 0.037$. Consequently, the normality assumption is questionable for the linear mediation model. The corresponding adaptive non-constant variances model with no changes to the model for the means but possible inclusion of transforms of X_1, X_2, and M in the model for the variances contains the single transform X_1^{19} in the model for the variances. The LCV(1,1) score for this adjusted model is 0.015046 with substantial improvement over the score 0.014958 for the associated constant variances model (as reported earlier) with percent decrease 0.58%. Consequently, the constant variance assumption is questionable for the component model (2) of the standard linear moderation model.
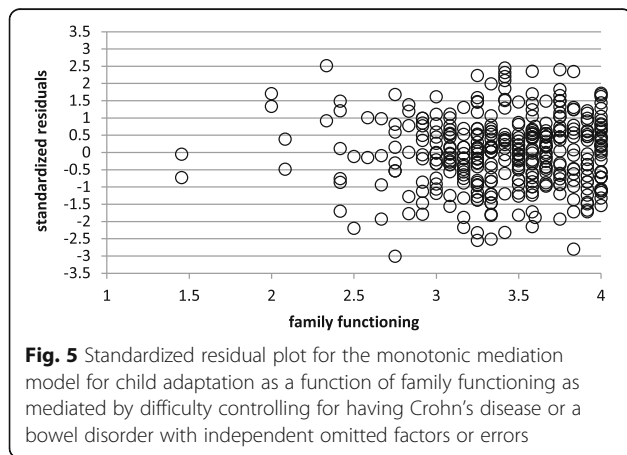
Figure 4 contains the normal plot generated by the monotonic mediation model (with $p = 1.3$, $q = 0$, $p' = -3$, $q' = 3$, and $\rho = 0$) adjusted for the covariate having Crohn's disease or a bowel disorder with the best power-adjusted LCV score generated so far. This plot is reasonably close to linear, the standardized residuals range for -3.01 to 2.51 with only one observation having a value -3.01 outside of ± 3, but very close to the boundary value of -3. The Shapiro-Wilk test for normality is now non-significant ($p = 0.475$), and so the normality assumption seems reasonable for this case. Figure 5 contains the plot of the standardized residuals in terms of family functioning. The
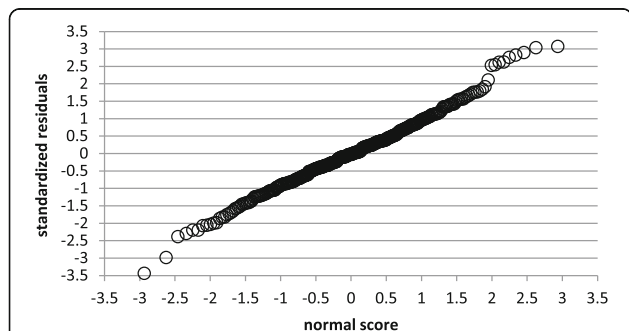
Fig. 3 Normal plot for the linear mediation model for child adaptation as a function of family functioning as mediated by difficulty with independent omitted factors or errors
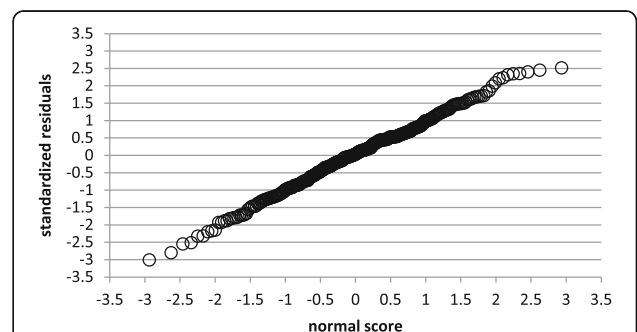
Fig. 4 Normal plot for the monotonic mediation model for child adaptation as a function of family functioning as mediated by difficulty controlling for having Crohn's disease or a bowel disorder with independent omitted factors or errors

Fig. 5 Standardized residual plot for the monotonic mediation model for child adaptation as a function of family functioning as mediated by difficulty controlling for having Crohn's disease or a bowel disorder with independent omitted factors or errors

Table 2 Bias-corrected bootstrapped 95% confidence intervals for the monotonic mediation model for child adaptation as a function of family functioning as mediated by difficulty controlling for having Crohn's disease or a bowel disorder with independent omitted factors or errors

Family functioning x	Lower bound $L(x)$ on natural indirect effect[a]	Upper bound $U(x)$ on natural indirect effect[a]	Normalized width W of the confidence interval[b]
1	2.28	8.19	0.72
2	9.14	32.76	0.72
3	20.56	73.72	0.72
4	36.55	130.05	0.72

[a]Using 1,000 resamples
[b]$W = (U(x) - L(x))/U(x)$, which is constant in x

assumption of constant variances is reasonable at least for all but a few exceptional, low family functioning values, suggesting that the standardized residuals are reasonable close to having constant variances. These results indicate that monotonic transformation can resolve distributional assumption problems for mediation models with continuous positive valued outcomes and mediators.

Results for the selected model

Using this monotonic mediation model with the best LCV score so far, estimated instantaneous total, natural direct, natural indirect, and relative natural indirect effects are presented in Table 1 for the grid of family functioning values 1, 2, 3, and 4 (1 is the smallest possible value while 4 is the largest possible value for the scale). The instantaneous natural indirect effect of family functioning on child adaptation increases with increasing or improving family functioning values while the instantaneous natural direct effect of family functioning on child adaptation decreases in such a way that the relative instantaneous natural indirect effect of family functioning on child adaptation increases. Family functioning needs to be relatively high in order for the relative instantaneous natural indirect effect to be relatively strong; in other words, mediation in this case is quite weak for values of family functioning between 1 and 2 In contrast, the standard linear mediation model (i.e., with $p = q = p' = q' = 1$ and $\rho = 0$) generates a total

effect of 18.5, a natural direct effect of 12.3, a natural indirect effect of 6.2, and a relative natural indirect effect of 0.34, quite different results.

Bootstrapped CIs

The standard linear mediation model (i.e., with $p = q = p' = q' = 1$ and $\rho = 0$) generates the same bootstrapped CI of $2.00 - 11.96$ for the instantaneous natural indirect effect at each value of family functioning. The normalized width W for this CI is 0.83.

Table 2 contains bias-corrected bootstrapped 95% CIs for the natural indirect effect at a range of values of family functioning under the monotonic mediation model with the best LCV score so far, that is, model (12) with powers $p = 1.3$, $q = 0$, $p' = -3$, and $q' = 3$, controlling for the covariate having Crohn's disease or a bowel disorder with independent omitted factors or errors (i.e., $\rho = 0$). The lower and upper bounds on the instantaneous natural indirect effects of family functioning on child adaptation increase with increasing values of family functioning. The normalized width W for these CIs is 0.72. Since this is smaller than the value 0.83 (about 13% smaller) for the standard linear mediation model, the monotonic mediation model generates more precise estimates of the instantaneous natural indirect effects.

The bootstrapped CI of $2.00 - 11.96$ for the linear moderation model (as reported earlier) overlaps with the

Table 1 Estimated instantaneous total, natural direct, natural indirect, and relative natural indirect effects for the monotonic mediation model for child adaptation as a function of family functioning as mediated by difficulty controlling for having Crohn's disease or a bowel disorder with independent omitted factors or errors

Family functioning	Instantaneous total effect	Instantaneous natural direct effect	Instantaneous natural indirect effect	Relative instantaneous natural indirect effect
1	3092.9	3087.9	4.9	0.002
2	212.6	193.0	19.6	0.092
3	82.3	38.1	44.1	0.536
4	90.5	12.1	78.5	0.867

Knafl *et al. BMC Medical Research Methodology* (2017) 17:45

Page 16 of 21

bootstrapped CIs for $x = 1$ and $x = 2$, but is entirely below the bootstrapped CIs for $x = 3$ and $x = 4$, suggesting that the linear moderation model in this case generates biased estimates of natural indirect effects for larger values of family functioning.

Moderated monotonic mediation

Another childhood chronic condition type for the family management study was diabetes with 51 (27.3%) children having this condition. The indicator Z for having this condition can be considered as a possible moderator for models (13)-(15). As before, the indicator for the child having Crohn's disease or a bowel disorder is considered as a covariate to include in both the models for the means and the variances as well as the model for the variances to depend on transforms of X_1, X_2, and M.

The generated model for the case of children with a chronic condition other than diabetes ($Z = 0$) has powers $p'(0) = -1.7$ and $q'(0) = 6$. The generated model for the case of children with diabetes ($Z = 1$) has the power $q'(1) = 0.5$ with no transform of family functioning (so a missing $p'(1)$ power). As before the covariate Crohn's disease or a bowel disorder is included in the model for the means and variances, but only in the component of model (15) addressing the outcome variable, that is, submodel (14). The LCV(1.3,0) score is 0.016215, which is a substantial improvement on the score 0.016062 for the associated model not considering moderation with percent decrease 0.94%. Consequently there is distinct moderated mediation.

The linear moderated mediation model (i.e., with $p = q = 1$, $p'(0) = q(0) = 1$, $p'(1) = q(1) = 1$, and $\rho = 0$) allowing for effects of the covariate having Crohn's disease or a bowel disorder on the means and variances as well as possibly nonlinear effects of X_1, X_2, and M on the variances has means and variances depending on the covariate having Crohn's disease of a bowel disorder for only the submodel (14) as for the moderated monotonic mediation model. Its LCV(1,1) score is 0.015460 with substantial percent decrease 4.66%. Consequently, the moderated mediation is distinctly nonlinear.

Model (15) adjusted to remove moderation of the effect of transformed X on transformed M in submodel (13) has LCV(1.3,0) score 0.016112 with substantial percent decrease 0.64% compared to the full model (15). Moreover, model (15) adjusted to remove moderation of the effect of transformed X on transformed Y in submodel (14) has LCV(1.3,0) score 0.016112 (same score as above but a different model) with substantial percent decrease 0.64% compared to the full model (15). Consequently, there is distinct moderation of both effects of the predictor X in model (15).

On the other hand, model (15) adjusted to remove moderation of the effect of transformed M on transformed Y in submodel (14) has LCV(1.3,0) score 0.016155 with insubstantial percent decrease 0.37% compared to the full model (15). Consequently, this is a parsimonious, competitive alternative model, and the effect of transformed M on transformed Y is reasonably considered not to be moderated by having diabetes.

Model (15) with the term $\beta_3(0) \cdot X_1^{q'(0)} \cdot I(Z = 0)$ removed has powers $q'(1) = 0.5$ and $p'(0) = -1$ with LCV(1.3,0) score 0.015846 and substantial percent decrease 2.28%. Thus, there is a distinct effect of transformed X on transformed M for children with a chronic condition other than diabetes in submodel (13). With the term $\beta_5(0) \cdot M^q \cdot I(Z = 0)$ removed, the model has powers $q'(0) = 6$, $q'(1) = 0.5$, and $p'(0) = 0.5$ with LCV(1.3,0) score 0.015949 and substantial percent decrease 1.64%. Thus, there is a distinct effect of transformed M on transformed Y for children with a chronic condition other than diabetes in submodel (14).

Model (15) with the term $\beta_3(1) \cdot X_1^{q'(1)} \cdot I(Z = 1)$ removed has powers $q'(0) = 6$ and $p'(0) = -1$ with LCV(1.3,0) score 0.015674 and substantial percent decrease 3.34%. Thus, there is a distinct effect of transformed X on transformed M for children with diabetes in submodel (13). On the other hand, with the term $\beta_5(1) \cdot M^q \cdot I(Z = 1)$ removed, the model has powers $q'(0) = 6$, $q'(1) = 0.5$, and $p'(0) = -1$ with LCV(1.3,0) score 0.016137 and insubstantial percent decrease 0.48%. Thus, the effect of transformed M on transformed Y for children with diabetes in submodel (14) is not distinct. This latter result indicates that the effect of transformed X on transformed Y is not distinctly mediated by transformed M for children with diabetes.

Table 3 Estimated instantaneous total, natural direct, natural indirect, and relative natural indirect effects for the monotonic mediation model for child adaptation as a function of family functioning as mediated by difficulty controlling for having Crohn's disease or a bowel disorder and as moderated by having diabetes with independent omitted factors or errors

Diabetes[a]	Family functioning	Instantaneous total effect	Instantaneous natural direct effect	Instantaneous natural indirect effect	Relative instantaneous natural indirect effect
no	1	931.5	931.4	0.1	<0.001
no	2	235.8	232.8	3.0	0.013
no	3	126.3	103.5	22.8	0.181
no	4	154.4	58.2	96.1	0.623

[a]There was no mediation for the diabetes = yes case

With the term $\beta_4(0) \cdot X_2^{p'(0)} \cdot I(Z=0)$ removed, the model has powers $q'(0) = 6$ and $q'(1) = 0.5$ with LCV(1.3,0) score 0.016058 and substantial percent decrease 0.97%. This result indicates that there is a distinct effect of transformed X on transformed Y for children with a chronic condition other than diabetes in submodel (14). On the other hand, the fact that $\beta_4(1) \cdot X_2^{p'(1)} \cdot I(Z=0)$ is not in the generated model indicates that the effect of transformed X on transformed Y for children with diabetes in submodel (14) is not distinct.

Using the parsimonious, competitive model with the term $\beta_5(1) \cdot M^q \cdot I(Z=1)$ removed, the standardized residuals range from -2.89 to 2.48 with nonsignificant ($p = 0.430$) Shapiro-Wilk normality test and normal plot reasonably close to linear (not displayed). Table 3 contains the associated estimated instantaneous total, natural direct, natural indirect, and relative natural indirect effects for the grid of family functioning values 1, 2, 3, and 4, but just for families having children with chronic conditions other than diabetes. The instantaneous natural indirect effect of family functioning on child adaptation increases with increasing or improving family functioning values while the instantaneous natural direct effect of family functioning on child adaptation decreases in such a way that the relative instantaneous natural indirect effect of family functioning on child adaptation increases. Compared to Table 1, instantaneous natural indirect effects and instantaneous total effects are smaller for low values of family functioning (1–3) and larger than for the highest value of family functioning (4). Relative instantaneous natural indirect effects are all smaller.

Table 4 contains bias-corrected bootstrapped 95% CIs for associated estimated instantaneous natural indirect effects of family functioning on child adaptation for the grid of family functioning values 1, 2, 3, and 4, also just for families having children with chronic conditions other than diabetes. Values for the lower and upper bounds increase with family functioning. Compared to Table 2, widths of the CIs are wider in absolute value for the highest value for family functioning (4) and narrower in absolute value otherwise (1–3). However, the relative width W is larger, 0.84 versus 0.72. In contrast, the associated linear model (i.e., with powers $p = q = 1$ and $p'(0) = q'(0) = 1$, $\rho = 0$, and having Crohn's disease or a bowel disorder covariate effects on the means and variances for submodel (14)) has estimated constant instantaneous natural indirect effect 5.5 with bias-corrected bootstrapped 95% CI 1.4 − 11.7 and normalized width W 0.88. Hence, the moderated monotonic mediation model generates more precise bootstrapped CIs than the standard moderated linear mediation model.

In summary, moderated mediation for the family management data is distinctly nonlinearly monotonic. However, mediation only occurs for mothers of children with a chronic condition other than diabetes and not for mothers of children with diabetes.

Example analyses of the simulated mediation data

Using the CSH covariance structure with $p = q = 1$, the first local maximum in the LCV score occurs at $k = 15$ with LCV(1,1) = 0.41542, and so $k = 15$ folds are used to compute subsequent LCV scores for the simulated data. The generated CSH model allowing for arbitrary p and q has $p = 0$ and $q = 0.3$ with distinctly improved LCV(0,0.3) = 0.93577 (i.e., the percent decrease for the $p = q = 1$ model is 55.6%, much larger than the cutoff of 0.95% for the data). Also, the estimated correlation for this model is the substantial value 0.92, suggesting highly dependent omitted factors or errors U_M and U_Y. However, the generated model allowing for arbitrary p and q with uncorrelated omitted factors or errors (i.e.., with $\rho = 0$), has $p = 0.4$ and $q = 1$ (i.e., the true values for these powers) with distinctly improved LCV(0.4,1) = 1.08555 (i.e., the percent decrease for the model with $p = 0$ and $q = 0.3$ is 13.8%), indicating that the omitted factors or errors are reasonably treated as independent as simulated. Under this model, $p' = 1.1$ and $q' = 1.2$, close to their simulated values of $p' = q' = 1$.

Table 4 Bias-corrected bootstrapped 95% confidence intervals for the monotonic mediation model for child adaptation as a function of family functioning as mediated by difficulty controlling for having Crohn's disease or a bowel disorder and as moderated by having diabetes with independent omitted factors or errors

Diabetes[a]	Family functioning x	Lower bound $L(x)$ on natural indirect effect[b]	Upper bound $U(x)$ on natural indirect effect[b]	Normalized width W of the confidence interval[c]
no	1	0.029	0.18	0.84
no	2	0.93	5.72	0.84
no	3	7.09	43.43	0.84
no	4	29.88	183.00	0.84

[a]There was no mediation for the diabetes = yes case
[b]Using 1,000 resamples
[c]$W = (U(x) − L(x))/U(x)$, which is constant in x

Knafl *et al. BMC Medical Research Methodology* (2017) 17:45

Page 18 of 21

The constant instantaneous indirect effect model associated with this latter model (i.e., with $q' = 1$) has $LCV(0.4,1) = 1.08074$ and insubstantial percent decrease 0.44%. Consequently, the instantaneous natural indirect effect is reasonably considered to be constant as simulated. Also, the true model as simulated (i.e., also setting $p' = 1$), has $LCV(0.4,1) = 1.07928$ and insubstantial percent decrease 0.58%, indicating that the true model is a competitive alternative for the simulated data.

Using the constant instantaneous natural indirect effect model (i.e., with $p = 0.4$, $q = 1$, $p' = 1.1$, and $q' = 1$), the 95% bootstrapped CI with bias correction for that effect is $0.0358 - 0.0563$, and so contains the true constant instantaneous natural indirect effect 0.04. The associated 95% bootstrapped CI without bias correction is $0.0363 - 0.0567$, and so is quite similar, suggesting that bias correction has not inflated the Type I error in this case.

Using the standard linear mediation model (i.e., with $p = 1$, $q = 1$, $p' = 1$, $q' = 1$, and $\rho = 0$), the 95% bootstrapped CI with bias correction for the constant instantaneous indirect effect is $0.0480 - 0.0974$, and so does not contain the true constant instantaneous natural indirect effect 0.04. The associated 95% bootstrapped CI without bias correction is $0.0481 - 0.0986$, and so is quite similar and also does not contain the true value.

Discussion

We formulated and demonstrated an approach for conducting possibly moderated monotonic mediation analyses based on adaptively selected fractional polynomial models. This formulation considers transformation of outcomes, predictors, and mediators, not just predictors and mediators as previously considered [29–32]. Results of the example analyses of the family management data indicated that transformation of positive valued continuous outcomes can provide distinct improvements over leaving those outcomes untransformed and can resolve problems with model assumptions.

Other nonlinear regression methods could have been used instead to estimate relationships. An advantage of fractional polynomial models is that they are based on linear regression models and so have no more limitations, assumptions, and requirements than models used in linear mediation analyses. Associated derivatives are also readily computed as needed for estimating monotonic instantaneous natural indirect, natural direct, and total effects.

The example analyses used likelihood cross-validation (LCV) for evaluating models, both unadjusted and adjusted for transformation of the outcome. By assessing how a model performs on randomly selected subsets, model evaluation using LCV scores is robust to effects of chance variation compared to using likelihoods based on the complete data. LCV ratio tests, generalizing likelihood ratio tests, can be used to assess whether mediation relationships are constant, linear, or nonlinear as well as a variety of other issues.

The example analyses of the family management data demonstrated the need to address nonlinearity in the context of mediation. Relationships considered in mediation analyses can be distinctly nonlinear. Even when mediation relationships are reasonably treated as linear, consideration of nonlinear alternatives is needed to determine that this assumption holds. It is conventional to treat mediation relationships as linear without checking this assumption. However, like any assumption, it should be checked.

While the example analyses of the family management data demonstrated that distinct nonlinear monotonic mediation can be identified using composite model (12), linear mediation also held. However, the normality assumption was questionable for the linear mediation case and also the constant variances assumption. There are likely to be data sets where mediation can be identified only by consideration of monotonicity. However, there are also likely to be data sets where consideration of monotonicity does not resolve problems related to the normality and constant variances assumptions. Quantile regression methods [26] can be used in such cases.

The example data analyses of the family management data also demonstrated that standard linear mediation analyses can provide a misleading impression that the natural indirect effect is constant in the predictor X when the indirect effect can actually vary quite a bit from this constant value. Moreover, the single 95% bootstrapped CI generated by the linear moderation analysis can be less precise than the CIs generated by a monotonic mediation analysis and can even not overlap for some values of the predictor X, suggesting that a linear approach can generate biased indirect effects. Furthermore, the example analyses of the family management data demonstrated that moderated mediation analyses are important to consider because mediation may be weaker for some subpopulations than others and even not hold for some subpopulations (e.g., mothers of children with diabetes compared to mothers of children with other chronic conditions).

The example analyses of the simulated mediation data support the effectiveness of adaptive mediation modeling since the true model was a competitive alternative to the adaptively selected model and since bootstrapped 95% CIs for the constant instantaneous indirect effect contained the true value. However, these analyses also demonstrate that allowing for correlated omitted factors or errors when in fact they are independent can generate models quite different from the true model, indicating the importance of conducting analyses of both cases. These analyses also demonstrate that conducting a standard linear mediation analysis when in fact the

relationship (9) is nonlinear in the outcome (i.e., $p{\neq}1$) but when the true instantaneous natural indirect effect is constant (i.e., $q'=1$) can result in a biased 95% CI for that constant effect. This is also likely to hold when relationships (8)-(9) are nonlinear in the mediator (i.e., $q{\neq}1$).

Limitations

The example mediation analyses of the family management data were limited since they were based on cross-sectional data, and so the timing of measurements for predictors, mediators, and outcomes could not be controlled to reflect precedence as needed to support causality [11, 44]. These analyses were also limited by the absence of control over the predictor variable given the non-experimental design of the study [21]. However, the primary purpose of those analyses was to demonstrate nonlinear mediation and the need for such analyses. This purpose was effectively achieved by the example analyses. Further work, though, is needed to address monotonic mediation in situations where the timing of measurement of variables has been controlled and where the predictor is experimentally controlled. An advantage of an experimentally controlled predictor is that it would be categorical and hence not require transformation. However, the outcome Y and the mediator M might benefit from transformation. There is also a need to replicate these analyses using a wider variety of data sets.

Nonlinearity was addressed here for mediation involving univariate outcomes. However, mediation is also conducted using repeated measurements, either over clusters or longitudinally over time, analyzed with multilevel modeling, linear mixed modeling, or structural equation modeling [6, 13, 57–61]. However, linear relationships are usually assumed in these analyses. Further research is needed to extend monotonic mediation to these cases. The example analyses used a continuous outcome and mediator, but outcomes and mediators can sometimes be categorical [13]. Further research is needed to extend monotonic mediation to address categorical mediators and/or outcomes. The analyses also only addressed the case with a single mediator. An extension to monotonic mediation is needed that accounts for multiple mediators.

Conclusions

Mediation relationships are commonly assumed to be linear without assessing the validity of this assumption. Reported example analyses demonstrate that mediation relationships can be nonlinear. Moreover, standard linear mediation analyses can generate models that violate model assumptions and generate biased estimates of indirect effects, but this can in some cases be resolved through more general monotonic mediation analyses. Adaptive methods as extended here to the monotonic mediation and moderated monotonic mediation contexts can effectively account for nonlinearity in mediation relationships. The advantage of restricting to monotonic relationships is that no adjustments are needed to underlying theory about directionality between changes in pairs of variables.

Authors' contributions
The family management data used in the example analyses were collected as part of an instrument development project [54]. All authors were involved in the development and conduct of this project including data collection and data analysis. For the current project, GJK developed the methods, generated the simulated data and the results, and wrote the initial description of those methods, data, and results. All authors reviewed drafts of the manuscript, rewrote sections, offered critical comments, and approved the final version.

Competing interests
The authors declare that they have no competing interests.

Consent for publication
Not applicable.

Ethics approval and consent to participate
The Institutional Review Board of the University of North Carolina at Chapel Hill approved secondary analyses of the family management data used in example analyses.

Author details
[1]School of Nursing, University of North Carolina at Chapel Hill, 5014 Carrington Hall, Campus Box 7460, Chapel Hill, NC 27599-7460, USA. [2]School of Nursing, Yale University, New Haven, USA. [3]School of Nursing, University of Pennsylvania, Philadelphia, USA. [4]College of Nursing, University of Illinois at Chicago, Chicago, USA.

References
1. Baron RM, Kenny DA. The moderator-mediator variable distinction in social psychology research: conceptual, strategic, and statistical considerations. J Pers Soc Psychol. 1986;51:1173–82.

Knafl *et al. BMC Medical Research Methodology* (2017) 17:45

Page 20 of 21

2. James LR, Brett JM. Mediators, moderators, and tests for mediation. J Appl Psychol. 1984;69:307–21.
3. Judd CM, Kenny DA. Process analysis: estimating mediation in treatment evaluation. Eval Rev. 1981;5:602–19.
4. Biesanz JC, Falk CF, Salavei V. Assessing mediational models: testing and interval estimation for indirect effects. Multivariate Behav Res. 2010;45:661–701.
5. Cheung GW, Lau RS. Testing mediation and suppression effects of latent variables: bootstrapping with structural equation models. Organ Res Meth. 2008;11:296–325.
6. Cole DA, Maxwell SE. Testing mediational models with longitudinal data: questions and tips on the use of structural equation modeling. J Abnorm Psychol. 2003;112:558–77.
7. Edwards JR, Lambert LS. Methods for integrating moderation and mediation: a general analytical framework using moderated path analysis. Psychol Methods. 2007;12:1–22.
8. Frazier PA, Tix AP, Barron KE. Testing moderator and mediator effects in counseling psychology. J Couns Psychol. 2004;51:113–34.
9. Gu F, Preacher KJ, Ferrer E. A state space modeling approach to mediation analysis. J Educ Behav Stat. 2014;39:117–43.
10. Kraemer HC. Toward non-parametric and clinically meaningful moderators and mediators. Stat Med. 2008;27:1679–92.
11. Kraemer HC, Kiernan M, Essex M, Kupfer DJ. How and why criteria defining moderators and mediators differ between the Baron & Kenny and MacArthur approaches. Health Psychol. 2008;27:S101–8.
12. Lepage B, Dedieu D, Savy N, Lang T. Estimating controlled direct effects of intermediate confounding of the mediator-outcome relationship: comparison of five different methods. Stat Methods in Med Res. 2012;21:1–18.
13. MacKinnon DP, Fairchild AJ, Fritz MS. Mediation analysis. Annu Rev Psychol. 2007;58:593–614.
14. 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.
15. Matthieu JE, Taylor SR. Clarifying conditions and decision points for mediational type inferences in organizational behavior. J Organ Behav. 2006;27:1031–56.
16. Muller D, Judd CM, Yzerbyt VY. When moderation is mediated and mediation moderated. J Pers Soc Psychol. 2005;89:852–63.
17. Muller D, Yzerbyt VY, Judd CM. Adjusting for a mediator with two crossed treatment variables. Organ Res Meth. 2008;11:224–40.
18. Nuitgen MB, Wetzels R, Matzke D, Dolan CV, Wagenmakers E-J. A default Bayesian hypothesis test for mediation. Behavioral Res Meth. 2015;47:85–97.
19. Pratschke J, Haase T, Comber H, Sharp L, de Camrgo CM, Johnson H. BMC Med Res Methodol. 2016;16:27. doi:10.1186/s12874-016-0130-6.
20. Rochon J, du Bois A, Lange T. Mediation analysis of the relationship between institutional research activity and patient survival. BMC Med Res Methodol. 2014;14:9. doi:10.1186/1471-2288-14-9.
21. Shrout PE, Bolger N. Mediation in experimental and nonexperimental studies: new procedures and recommendations. Psychol Methods. 2002;7:422–45.
22. Taylor AB, MacKinnon DP, Tein J-Y. Tests of the three-path mediated effect. Organ Res Meth. 2008;11:241–69.
23. Valeri L, VanderWeele T. Mediation analysis allowing for exposure–mediator interactions and causal interpretation: theoretical assumptions and implementation with SAS and SPSS macros. Psychol Methods. 2013;18:137–50.
24. Vanderweele TJ. Mediation analysis with multiple versions of the mediator. Epidemiol. 2012;23:454–63.
25. Yuan Y, MacKinnon DP. Bayesian mediation analysis. Psychol Methods. 2009;14:301–22.
26. Yuan Y, MacKinnon DP. Robust mediation analysis based on median regression. Psychol Methods. 2014;19:1–20.
27. Zhao X, Lynch Jr JG, Chen Q. Reconsidering Baron and Kenny: myths and truths about mediation analysis. J Cons Res. 2011;37:197–206.
28. Zu J, Yuan K-H. Local influence and robust procedures for mediation analysis. Multivariate Behav Res. 2010;45:1–44.
29. Hayes AF, Preacher KJ. Quantifying and testing indirect effects in simple mediation models when the constituent paths are nonlinear. Multivariate Behav Res. 2010;45:627–60.
30. Pearl J. The foundations of causal inference. Sociol Methodol. 2010;40:75–149.
31. Pearl J. A general approach to causal mediation analysis. Psychol Methods. 2014;15:309–34.
32. Stolzenberg RM. The measurement and decomposition of causal effects in nonlinear and nonadditive models. Sociol Methodol. 1980;11:459–88.
33. Royston P, Altman DG. Regression using fractional polynomials of continuous covariates: parsimonious parametric modelling. Appl Stat. 1994;43:429–67.
34. Knafl GJ, Ding K. Adaptive regression for modeling nonlinear relationships. Switzerland: Springer International Publishing; 2016.
35. Knafl GJ, Fennie KP, Bova C, Dieckhaus K, Williams AB. Electronic monitoring device event modelling on an individual-subject basis using adaptive Poisson regression. Stat Med. 2004;23:783–801.
36. Knafl GJ, Delucchi KL, Bova CA, Fennie KP, Williams AB. A systematic approach for analyzing electronically monitored adherence data. In: Ekwall B, Cronquist M, editors. Micro electro mechanical systems (MEMS) technology, fabrication processes and applications, (Chapter 1, pp. 1–66). Hauppauge: Nova Science Publishers; 2010. https://www.novapublishers.com/catalog/product_info.php?products_id=19133. Accessed 7 Dec 2016.
37. Knafl GJ. A SAS macro for adaptive regression modeling. In: Proceedings SAS global forum 2009; 2009. http://support.sas.com/resources/papers/proceedings09/110-2009.pdf. Accessed 7 Dec 2016.
38. MacKinnon DP, Krull JL, Lockwood CM. Equivalence of mediation, confounding, and suppression. Prev Sci. 2000;1:173–81.
39. Sobel ME. Asymptotic confidence intervals for indirect effects in structural equation models. In: Leinhart S, editor. Sociological methodology, San Francisco: Jossey-Bass. 1982. p. 290–312.
40. MacKinnon DP, Lockwood CM, Williams J. Confidence limits for the indirect effect: distribution of the product and resampling methods. Multivariate Behav Res. 2004;39:99–128.
41. Preacher KJ, Hayes AF. SPSS and SAS procedures for estimating indirect effects in simple mediation models. Behav Res Methods Instrum Comput. 2004;36:717–31.
42. Efron B. The jackknife, the bootstrap, and other resampling plans, CBMS 38, SIAM-NSF. Philadelphia: Society for Industrial and Applied Mathematics; 1982.
43. Fritz MS, Taylor AB, MacKinnon DP. Explanation of two results in statistical mediation analysis. Multivariate Behav Res. 2012;47:61–87.
44. MacKinnon DP, Luecken LJ. How and for whom? mediation and moderation in psychology. Health Psychol. 2008;27:S99–100.
45. VanderWeele TJ, Vansteelandt S. Conceptual issues concerning mediation, interventions and composition. Stat Interface. 2009;2:457–68.
46. Box GEP, Tidwell PW. Transformation of the independent variables. Technometrics. 1962;4:531–50.
47. McCullagh P, Nelder JA. Generalized linear models. 2nd ed. Chapman & Hall/CRC: Boca Raton, FL; 1999.
48. Knafl GJ, Riegel B. What puts heart failure patients at risk for poor medication adherence? Patient Prefer Adher. 2014;8:1007–18.
49. Meghani SH, Knafl GJ. Patterns of analgesic adherence predict health care utilization among outpatients with cancer pain. Patient Prefer Adher. 2016;10:81–98.
50. Riegel B, Knafl GJ. Electronically monitored medication adherence predicts hospitalization in heart failure patients. Patient Prefer Adher. 2014;8:1–13.
51. Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: Mellish CS, editor. Proceedings of the 14th international joint conference on artificial intelligence. San Francisco: Morgan Kaufman; 1995. p. 1137–43.
52. Knafl GJ, Grey M. Factor analysis model evaluation through likelihood cross-validation. Stat Methods in Med Res. 2007;16:77–102.
53. Preacher KJ, Rucker DD, Hayes AF. Addressing moderated mediation hypotheses: theory, methods, and prescriptions. Multivariate Behav Res. 2007;42:185–227.
54. Knafl K, Deatrick JA, Gallo A, Dixon JK, Grey M, Knafl GJ, O'Malley JP. Development and testing of the Family Management Measure. J Pediatr Psychol. 2011;36:494–505.
55. Epstein N, Baldwin L, Bishop D. The McMaster Family Assessment Device. J Marital Fam Ther. 1983;9:171–80.
56. Eyberg S, Robinson E. Conduct problem behavior: standardization of a behavior rating scale with adolescents. J ClinI Child Psychol. 1983;12:347–54.
57. Judd CM, Kenny DA, McCelland GH. Estimating and testing mediation in within-subject designs. Psychol Methods. 2001;6:115–34.
58. Kenny DA, Korchmaros JD, Bolger N. Lower level mediation in multilevel models. Psychol Methods. 2003;8:115–28.
59. MacKinnon DP. Introduction to statistical mediation analysis. New York: Lawrence Erlbaum; 2008.

Knafl *et al. BMC Medical Research Methodology* (2017) 17:45

Page 21 of 21

60. Kenny DA, Kashy DA, Bolger N. Data analysis in social psychology. In: Gilbert S, Fiske T, Lindsay D, editors. Handbook of social psychology. 4th ed. New York: McGraw Hill; 1998. p. 115–28.
61. Preacher KJ, Hayes AF. Contemporary approaches to assessing mediation in communication research. In: Hayes A, Slater MD, Snyder LB, editors. The SAGE sourcebook of advanced data analysis methods for communication research. Thousand Oaks, CA: Sage; 2008. p. 13–54.
62. Knafl G. Analyzing mediation data. 2016. http://www.unc.edu/~gknafl/mediation.html. Accessed 7 Dec 2016.