DOI:
10.1037/met0000154

Document Version
Peer reviewed version

Citation for published version (APA):
Goldsmith, K. A., MacKinnon, D. P., Chalder, T., White, P. D., Sharpe, M., & Pickles, A. (2018). Tutorial: The Practical Application of Longitudinal Structural Equation Mediation Models in Clinical Trials. *PSYCHOLOGICAL METHODS*, *23*(2), 191-207. https://doi.org/10.1037/met0000154

Supplemental Material

A Brief Description of Mediation and Structural Equation Models for Mediation

When we assess a treatment in the context of a clinical trial, we estimate a total treatment effect, R, of a treatment versus a control on an outcome of interest, Y (Figure S1, upper diagram). We should generally have a theory of how our treatment works – we have designed it to target an intermediate variable, which in turn has an effect on the outcome. This is our hypothesized treatment mechanism. We can further explore treatment mechanisms – how the treatment works – using mediation analysis. This method partitions the total effect of the treatment R on the outcome Y into an indirect effect transmitted through this intermediate or mediator variable M, and a residual direct effect of R on Y (Figure S1, lower diagram). Mediation analysis evaluates a causal chain of events, here how much of the effect of a randomized treatment R on an outcome variable of interest Y is transmitted through a mediating variable M.

When the mediator and outcome are continuous measures, and other assumptions hold that are described in the next paragraph, the relationships implied by the diagram can be expressed as a series of ordinary least squares regression equations that allow estimation of the total, direct, and indirect effects. Where Y = outcome, M = mediator, and R = randomly assigned treatment:

$$Y_i = d_1 + cR_i + \varepsilon_{i1} \tag{1}$$

$$M_i = d_2 + aR_i + \varepsilon_{i2} \tag{2}$$

$$Y_i = d_3 + c'R_i + bM_i + \varepsilon_{i3}\,. \tag{3}$$

The *i* subscript refers to the unit (often the individual) receiving treatment. The d_1, d_2 and d_3 parameters are intercepts, with e_{i1}, e_{i2}, e_{i3} representing the error terms in the regression Equations 1, 2, and 3. The *a*, *b*, *c*, and *c'* parameters are those that are of interest in mediation analysis, also shown in Figure S1. The total effect of R on Y is estimated as the *c* parameter/path in Equation 1, and is shown in the upper diagram in Figure S1. The effect of R on M is obtained by fitting Equation 2 with the mediator as the dependent variable, and estimating the *a* parameter/path (Figure S1, lower). The effect of M on Y is obtained by fitting Equation 3 with the outcome as the dependent variable, and estimating the *b* parameter/path (Figure S1, lower). The indirect effect from R to M to Y is then calculated as *a* multiplied by *b* (Alwin & Hauser, 1975; Baron & Kenny, 1986). This is the product of coefficients (POC) method for obtaining the indirect or mediated effect (Baron & Kenny, 1986; MacKinnon, 2001, 2008). If there is mediation of the treatment effect, the estimates of the *a* path, *b* path, and indirect or mediated *ab* effect will all be statistically significant. Equation 3 estimates the remaining direct effect of treatment on the outcome, as given by the *c'* parameter/path.

In order to use Equations 1, 2, and 3 to estimate a mediated effect, several other assumptions must hold. These include reliably and validly measured variables, linear relationships between variables – including no R x M interaction on the outcome, no unmeasured confounding of the mediator – outcome relationship as described above (U in the lower part of Figure S1) (Emsley, Dunn, & White, 2010; Imai, Keele, & Tingley, 2010; Judd & Kenny, 1981; MacKinnon, 2008; VanderWeele & Vansteelandt, 2009), and no confounding of the M – Y relationship by post-randomization variables (De Stavola, Daniel, Ploubidis, & Micali, 2015; Robins & Greenland, 1992; VanderWeele, 2009; VanderWeele & Vansteelandt, 2009; VanderWeele, Vansteelandt, & Robins, 2014).

It may be unclear at first why there could be confounding in the context of a randomized trial. The R – M and R – Y relationships should be free of confounding if the randomization was done correctly, however, we are not randomizing M, so the M – Y relationship could be subject to unmeasured confounding, as shown by U in Figure S1. We can think of U as inducing a correlation between M and Y. We cannot allow for such a correlation in a simple mediator model as the model would not be identified, so we generally assume there is no unmeasured confounding. Unmeasured confounders could be a source of substantial bias in mediation analysis (Emsley et al., 2010; Imai et al., 2010; VanderWeele & Vansteelandt, 2009). One way to help alleviate bias is to measure all potential confounders of M and Y and include them in the models, however, there may still be some residual unmeasured confounding.

As shown above, even a simple mediation model with one mediator and one outcome requires fitting two equations. Longitudinal mediation models incorporating repeated measurements require fitting multiple equations. One way we can fit more than one equation simultaneously is to use the SEM framework. SEM consist of both a measurement portion relating observed variables to latent underlying constructs, and a structural portion describing relationships among these latent and other observed variables. We use the path tracing rules described by Wright (Kline, 2011; Wright, 1920a, 1920b), which decompose the covariances between variables modeled using SEM, allowing for the estimation of mediated/indirect effects. In practical terms, we take the products of the estimates along each legal path between two variables (for example, paths cannot travel backwards down an arrow), and then sum the products across all of the paths. More information can be found in an introductory SEM book, such as Kline (2011). Applying path tracing rules to the simple mediation model gives the indirect effect from R to Y via M as *a* times *b*, the same result we obtain when using the POC method. The effect can be obtained in one step in the SEM by simultaneously fitting the two equations implied by the model and asking for the indirect effects to be estimated.

Detail on Simulated Dataset for Use with the Tutorial

The data were simulated using the Mplus MONTECARLO command, with parameter estimates obtained from fitting a model to PACE data standardized to baseline values of the mediator and outcome. The parameters from PACE were modified and then used as population and coverage values in the MONTECARLO command. The parameters used are shown in Figure S2. Two datasets of $n = 320$ were simulated using the modified latent change score model (presented in the manuscript, used because it had the best fit to the real trial data), each with a binary treatment group variable. The two datasets were then merged to form a dataset with four treatment groups.

The simulated dataset is called *longitudinal mediation.dat*. The dataset contains the following variables: y0, y1, y2, y3, m0, m1, m2, m3, r1, r2, r3, and r4 in that order. The y variables represent measures of the outcome at baseline (y0), and at three time points post-randomization, with the m variables representing mediator measurements taken at the same time points. The four treatment group variables r1, r2, r3, and r4 are dummy coded variables coded = 1 for membership in that randomized group and = 0 otherwise. These variables are described in more detail in the body of the manuscript.

Mplus input and output files

The names of the input and output files for the models are as follows:

a) *simplex lagged* (simplex model with lagged *b* paths)

b) *simplex contemporaneous* (simplex model with contemporaneous *b* paths)

c) latent growth

d) latent change

e) modified latent change

We also include the code for the simplex model with lagged b paths in this supplemental document and refer to it in the following section, which provides some detail on obtaining the mediated effects using Mplus.

Fitting Longitudinal Mediation Models Using Mplus

Below we have provided information about where to find important effects in the Mplus code for the simplex lagged model, and also discuss the calculation of the indirect and direct effect parameters. The code is set up with similar patterns for all models, which will allow the reader to extrapolate across models.

For example, some of the important paths described in the manuscript can be found as follows in the simplex lagged model code below:

- The R1 treatment group a_{11} path is FM1 on R1 in the Mplus code, line 50.
- The R3 treatment group a_{31} path is FM1 on R3 in the Mplus code, line 52.
- The common b_L path is FY2 on FM1 or FY3 on FM2 in the Mplus code, lines 82 and 83.

The Mplus subcommand MODEL INDIRECT calculates indirect and direct effects, and provides statistical tests and CI for these effects. Mplus applies matrix equations for indirect and direct effects to obtain the effect decomposition for the model (Alwin & Hauser, 1975; MacKinnon, 2008; Sobel, 1986). For the example of effects from the R1 treatment to third time point (y3) in the simplex model shown in Figure 1 in the manuscript, the command is "y3 ind R1;" (see line 117 in Mplus code for the simplex model with lagged *b* paths).

Extra code is needed in the Mplus program to obtain some of the indirect and direct effects defined by Cole and Maxwell (Cole & Maxwell, 2003). SEM path tracing rules and Mplus define direct effects as a single directed path between two variables only, for example, the R1 to FY_1 effect in the simplex model in Figure 1 in the manuscript. The Cole and

Maxwell direct effect example described in the body of the manuscript, R1 -> FY_1 -> FY_2 -> FY_3 -> Y_3, is considered to be an indirect effect in the SEM framework, Mplus and other SEM software. This means that what Mplus outputs as the "total indirect effect" includes some effects that we are defining as direct effects, for example R1 -> FY_1 -> FY_2 -> FY_3 -> Y_3, and also leaves these direct effects out of its calculation of "total direct effects". In fact, in the models fitted here, there is no "direct" path between treatment and outcome for the third post-randomization time point, that is, no effect going directly from treatment to the outcome at the third time point. All of the effects for the third time point are therefore reported as indirect in Mplus, as they would also be in other SEM software. If we want the Cole and Maxwell defined total/overall direct and indirect effects, we have to ask Mplus (and likely other software programs) to calculate and display these effects.

These effects and their associated CI can be obtained in Mplus using the MODEL CONSTRAINT command, as shown starting at line 124 in Mplus code for the simplex model with lagged *b* paths. This command enables the specification of functions of parameters. First each new parameter must be given a name (using the NEW option), and then each constraint is specified. We have calculated the effects separately for each of the R1, R2, and R3 treatment groups. We then need to sum up several product terms, so on each line we calculate the product for a single path, then we go on to specify the time-specific and overall direct and indirect effects as sums of the calculated product terms. The order is

- individual effects,
- total for that type of effect (indirect or direct),
- then the total effect at that time point.

The naming convention for the parameters follows this order

- treatment group,
- type of effect,

- time point,
- effect number (the latter is assigned for convenience).

For the simplex with lagged *b* paths model, the R1 treatment group and third post-randomization time point, the naming convention is as follows (see lines 153 to 167 in the simplex lagged code):

- r1i31: R1 = treatment group, I = indirect, 3 = time point, 1 = effect number, first indirect effect for R1 at third post-randomization time point, line 153;
- r1i32: second indirect effect for R1 at third post-randomization time point, line 155;
- r1i33: third indirect effect for R1 at third post-randomization time point, line 157;
- r1i3t: total indirect effect for R1 at third post-randomization time point, sum of r1i31, r1i32, and r1i33, line 159;
- r1d31: first direct effect for R1 at third post-randomization time point, line 161;
- r1d32: second direct effect for R1 at third post-randomization time point, line 163;
- r1d3t: total direct effect for R1 at third post-randomization time point, sum of r1d31 and r1d32, line 165;
- r13t: total effect for R1 for third post-randomization time point, sum of the total indirect (r1i3t) and total direct (r1d3t), line 167.

The appropriate effects are calculated in a similar manner in the code for each model and repeated for treatments R2 and R3. Note that if this code were being applied to a trial where there were only two treatment groups, the effects would reduce down to just those for R1. Also note that although not necessary, where the factor loading/final path in calculating the effect is equal to one, we have multiplied by this in the code in order to make the code explicit (see the body of the manuscript for explanation of the paths used in the calculations). The estimates and CI for the effects requested in the MODEL CONSTRAINTS command can be found under “New/Additional Parameters” in the Mplus output.

Bootstrapped CIs for the effects are requested by adding the line ‘ANALYSIS: BOOTSTRAP = number of repetitions;’ to the program and requesting ‘CINTERVAL(BOOTSTRAP)’ on the OUTPUT line (for example, see Mplus code for the simplex model with lagged *b* paths, lines 16 and 257).

Simplex model with lagged b paths code

```
1  DATA:
2            !PROVIDE NAME OF DATAFILE
3            FILE IS longitudinal mediation.dat ;
4
5  VARIABLE:
6            !PROVIDE VARIABLE NAMES
7            Names are y0 y1 y2 y3 m0 m1 m2 m3 r1 r2 r3 r4 ;
8            !PROVIDE NAMES OF VARIABLES TO BE USED IN MODEL
9            !NOTE ONE TREATMENT DUMMY VARIABLE LEFT OUT TO BE REFERENCE CATEGORY
10           USEVARIABLES ARE y0 y1 y2 y3 m0 m1 m2 m3 r1 r2 r3 ;
11           !MISSING DATA ARE CODED AS 99 IN THE DATASET
12           Missing are all (99);
13
14           !REQUEST BOOTSTRAP ANALYSIS TO GET CONFIDENCE INTERVALS FOR INDIRECT
15 EFFECTS
16 ANALYSIS: BOOTSTRAP = 1000 ;
17
18 MODEL:
19            !MEASUREMENT MODEL FOR MEDIATOR, ALLOW UNDERLYING
20            !TRUE SCORE FACTOR FOR EACH OBSERVED VARIABLE
21            !(1) AT THE END OF EACH LINE INDICATES THAT THE
22            !FACTOR LOADINGS CONSTRAINED TO BE EQUAL
23            !ONE FACTOR LOADING HAS TO BE = 1, SO HERE ALL
24            !FACTOR LOADINGS CONSTRAINED = 1
25                       fm0 by m0 (1) ;
26                       fm1 by m1 (1) ;
27                       fm2 by m2 (1) ;
28                       fm3 by m3 (1) ;
29
30            !(2) CONSTRAINS RESIDUAL VARIANCES TO BE EQUAL ACROSS
31            !MEDIATOR MEASURES
32                       m0 m1 m2 m3 (2);
33
34            !MEASUREMENT MODEL FOR OUTCOME, ALLOW UNDERLYING
35            !TRUE SCORE FACTOR FOR EACH OBSERVED VARIABLE
36            !(1) AT THE END OF EACH LINE INDICATES THAT THE
37            !FACTOR LOADINGS CONSTRAINED TO BE EQUAL
38            !ONE FACTOR LOADING HAS TO BE = 1, SO HERE ALL
39            !FACTOR LOADINGS CONSTRAINED = 1
40                       fy0 by y0 (1) ;
41                       fy1 by y1 (1) ;
42                       fy2 by y2 (1) ;
43                       fy3 by y3 (1) ;
44
45            !(3) CONSTRAINS RESIDUAL VARIANCES TO BE EQUAL ACROSS
46            !OUTCOME MEASURES
47                       y0 y1 y2 y3 (3);
48
49            !A PATH EFFECTS OF TREATMENT GROUP ON MEDIATOR AT FIRST TIME POINT
50                       fm1 on r1*0 (a11) ;
```

```
                        fm1 on r2*0 (a21) ;
                        fm1 on r3*0 (a31) ;

          !A PATH EFFECTS OF TREATMENT GROUP ON MEDIATOR AT SECOND TIME POINT
                        fm2 on r1*0 (a12) ;
                        fm2 on r2*0 (a22) ;
                        fm2 on r3*0 (a32) ;

          !NO EFFECT OF TREATMENT GROUP ON MEDIATOR AT THIRD TIME POINT
                        fm3 on r1@0 r2@0 r3@0 ;

          !SIMPLEX MODEL STABILITY EFFECT BETWEEN TIME POINTS - MEDIATOR
                        fm1 on fm0*0.8 ;
                        fm2 on fm1*0.8 (m2) ;
                        fm3 on fm2*0.8 (m3) ;

          !C' PATH EFFECTS OF TREATMENT GROUP ON OUTCOME AT FIRST TIME POINT
                        fy1 on r1*0 (c11) ;
                        fy1 on r2*0 (c21) ;
                        fy1 on r3*0 (c31) ;

          !C' PATH EFFECTS OF TREATMENT GROUP ON OUTCOME AT SECOND TIME POINT
                        fy2 on r1*0 (c12) ;
                        fy2 on r2*0 (c22) ;
                        fy2 on r3*0 (c32) ;

          !NO EFFECT OF TREATMENT GROUP ON OUTCOME AT THIRD TIME POINT
                        fy3 on r1@0 r2@0 r3@0 ;

          !LAGGED B PATH EFFECTS
                        fy1 on fm0*0 ;
                        fy2 on fm1*0 (b) ;
                        fy3 on fm2*0 (b) ;

          !SIMPLEX MODEL STABILITY EFFECT BETWEEN TIME POINTS - OUTCOME
                        fy1 on fy0*0.8 ;
                        fy2 on fy1*0.8 (y2) ;
                        fy3 on fy2*0.8 (y3) ;

          !ALLOW FOR COVARIANCE BETWEEN BASELINE MEDIATOR
          !AND OUTCOME TRUE SCORES
          !CONSTRAIN COVARIANCE BETWEEN BASELINE MEDIATOR
          !AND OUTCOME TRUE SCORES = 0 AT OTHER TIME POINTS
                        fm0 with fy0*0 ;
                        fm1 with fy1@0 ;
                        fm2 with fy2@0 ;
                        fm3 with fy3@0 ;
                        fy1 with fy2@0 ;
                        fy2 with fy3@0 ;
                        fy1 with fy3@0 ;

           !ALLOW FOR RESIDUAL COVARIANCE BETWEEN MEDIATOR AND OUTCOME AT EACH
           !TIME POINT AND CONSTRAIN TO BE EQUAL (4 AT END OF EACH LINE)
                        m0 with y0 (4) ;
                        y1 with m1 (4) ;
                        y2 with m2 (4) ;
                        y3 with m3 (4) ;

          !USE OF THE BUILT IN MPLUS COMMAND TO REQUEST INDIRECT EFFECTS
          !FOR EACH POST-RANDOMISATION TIME POINT
MODEL INDIRECT:         y1 ind r1 ;
                        y1 ind r2 ;
                        y1 ind r3 ;
                        y2 ind r1 ;
                        y2 ind r2 ;
                        y2 ind r3 ;
                        y3 ind r1 ;
```

```
                y3 ind r2 ;
                y3 ind r3 ;
                !BUILT IN COMMAND DOES NOT PROVIDE ALL EFFECTS OF INTEREST
                !BELOW SHOWS USE OF MODEL CONSTRAINT COMMAND TO GET ALL
EFFECTS

MODEL CONSTRAINT:
                !EFFECTS FOR R1
                NEW(r1d11 r1d1t r11t);
                NEW(r1i21 r1i2t r1d21 r1d22 r1d2t r12t);
                NEW(r1i31 r1i32 r1i33 r1i3t r1d31 r1d32 r1d3t r13t) ;

                !THERE ARE NO INDIRECT EFFECTS AT TIME POINT 1
                !IN THESE MODELS WITH LAGGED B PATHS
                !FIRST DIRECT EFFECT TIME POINT 1
                r1d11 = c11*1 ;
                !TOTAL DIRECT EFFECT TIME POINT 1
                r1d1t = r1d11 ;
                !TOTAL EFFECT TIME POINT 1 = TOTAL DIRECT
                r11t = r1d1t ;

                !FIRST INDIRECT EFFECT TIME POINT 2
                r1i21 = a11*b*1 ;
                !TOTAL INDIRECT EFFECT TIME POINT 2
                r1i2t = r1i21 ;
                !FIRST DIRECT EFFECT TIME POINT 2
                r1d21 = c11*y2*1 ;
                !SECOND DIRECT EFFECT TIME POINT 2
                r1d22 = c12*1 ;
                !TOTAL DIRECT EFFECT TIME POINT 2
                r1d2t = r1d21 + r1d22 ;
                !TOTAL EFFECT TIME POINT 2 = TOTAL INDIRECT + TOTAL DIRECT
                r12t = r1i2t + r1d2t ;

                !FIRST INDIRECT EFFECT TIME POINT 3
                r1i31 = a11*b*y3*1 ;
                !SECOND INDIRECT EFFECT TIME POINT 3
                r1i32 = a11*m2*b*1 ;
                !THIRD INDIRECT EFFECT TIME POINT 3
                r1i33 = a12*b*1 ;
                !TOTAL INDIRECT EFFECT TIME POINT 3 = OVERALL INDIRECT EFFECT
                r1i3t = r1i31 + r1i32 + r1i33 ;
                !FIRST DIRECT EFFECT TIME POINT 3
                r1d31 = c11*y2*y3*1 ;
                !SECOND DIRECT EFFECT TIME POINT 3
                r1d32 = c12*y3*1 ;
                !TOTAL DIRECT EFFECT TIME POINT 3 = OVERALL DIRECT EFFECT
                r1d3t = r1d31 + r1d32 ;
                !TOTAL EFFECT TIME POINT 3 = TOTAL INDIRECT + TOTAL DIRECT
                r13t = r1i3t + r1d3t ;


                !EFFECTS FOR R2
                NEW(r2d11 r2d1t r21t);
                NEW(r2i21 r2i2t r2d21 r2d22 r2d2t r22t);
                NEW(r2i31 r2i32 r2i33 r2i3t r2d31 r2d32 r2d3t r23t) ;

                !THERE ARE NO INDIRECT EFFECTS AT TIME POINT 1
                !IN THESE MODELS WITH LAGGED B PATHS
                !FIRST DIRECT EFFECT TIME POINT 1
                r2d11 = c21*1 ;
                !TOTAL DIRECT EFFECT TIME POINT 1
                r2d1t = r2d11 ;
                !TOTAL EFFECT TIME POINT 1 = TOTAL DIRECT
                r21t = r2d1t ;

                !FIRST INDIRECT EFFECT TIME POINT 2
```

```
r2i21 = a21*b*1 ;
!TOTAL INDIRECT EFFECT TIME POINT 2
r2i2t = r2i21 ;
!FIRST DIRECT EFFECT TIME POINT 2
r2d21 = c21*y2*1 ;
!SECOND DIRECT EFFECT TIME POINT 2
r2d22 = c22*1 ;
!TOTAL DIRECT EFFECT TIME POINT 2
r2d2t = r2d21 + r2d22 ;
!TOTAL EFFECT TIME POINT 2 = TOTAL INDIRECT + TOTAL DIRECT
r22t = r2i2t + r2d2t ;

!FIRST INDIRECT EFFECT TIME POINT 3
r2i31 = a21*b*y3*1 ;
!SECOND INDIRECT EFFECT TIME POINT 3
r2i32 = a21*m2*b*1 ;
!THIRD INDIRECT EFFECT TIME POINT 3
r2i33 = a22*b*1 ;
!TOTAL INDIRECT EFFECT TIME POINT 3 = OVERALL INDIRECT EFFECT
r2i3t = r2i31 + r2i32 + r2i33 ;
!FIRST DIRECT EFFECT TIME POINT 3
r2d31 = c21*y2*y3*1 ;
!SECOND DIRECT EFFECT TIME POINT 3
r2d32 = c22*y3*1 ;
!TOTAL DIRECT EFFECT TIME POINT 3 = OVERALL DIRECT EFFECT
r2d3t = r2d31 + r2d32 ;
!TOTAL EFFECT TIME POINT 3 = TOTAL INDIRECT + TOTAL DIRECT
r23t = r2i3t + r2d3t ;

!EFFECTS FOR R3
NEW(r3d11 r3d1t r31t);
NEW(r3i21 r3i2t r3d21 r3d22 r3d2t r32t);
NEW(r3i31 r3i32 r3i33 r3i3t r3d31 r3d32 r3d3t r33t) ;

!THERE ARE NO INDIRECT EFFECTS AT TIME POINT 1
!IN THESE MODELS WITH LAGGED B PATHS
!FIRST DIRECT EFFECT TIME POINT 1
r3d11 = c31*1 ;
!TOTAL DIRECT EFFECT TIME POINT 1
r3d1t = r3d11 ;
!TOTAL EFFECT TIME POINT 1 = TOTAL DIRECT
r31t = r3d1t ;

!FIRST INDIRECT EFFECT TIME POINT 2
r3i21 = a31*b*1 ;
!TOTAL INDIRECT EFFECT TIME POINT 2
r3i2t = r3i21 ;
!FIRST DIRECT EFFECT TIME POINT 2
r3d21 = c31*y2*1 ;
!SECOND DIRECT EFFECT TIME POINT 2
r3d22 = c32*1 ;
!TOTAL DIRECT EFFECT TIME POINT 2
r3d2t = r3d21 + r3d22 ;
!TOTAL EFFECT TIME POINT 2 = TOTAL INDIRECT + TOTAL DIRECT
r32t = r3i2t + r3d2t ;

!FIRST INDIRECT EFFECT TIME POINT 3
r3i31 = a31*b*y3*1 ;
!SECOND INDIRECT EFFECT TIME POINT 3
r3i32 = a31*m2*b*1 ;
!THIRD INDIRECT EFFECT TIME POINT 3
r3i33 = a32*b*1 ;
!TOTAL INDIRECT EFFECT TIME POINT 3 = OVERALL INDIRECT EFFECT
r3i3t = r3i31 + r3i32 + r3i33 ;
!FIRST DIRECT EFFECT TIME POINT 3
r3d31 = c31*y2*y3*1 ;
!SECOND DIRECT EFFECT TIME POINT 3
```

```
                      r3d32 = c32*y3*1 ;
                      !TOTAL DIRECT EFFECT TIME POINT 3 = OVERALL DIRECT EFFECT
                      r3d3t = r3d31 + r3d32 ;
                      !TOTAL EFFECT TIME POINT 3 = TOTAL INDIRECT + TOTAL DIRECT
                      r33t = r3i3t + r3d3t ;
OUTPUT: CINTERVAL(BOOTSTRAP) ;
```

References

Alwin, D. F., & Hauser, R. M. (1975). Decomposition of effects in path analysis. *American Sociological Review, 40*(1), 37-47. doi: 10.2307/2094445

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

Cole, D. A., & Maxwell, S. E. (2003). Testing mediational models with longitudinal data: Questions and tips in the use of structural equation modeling. *Journal of Abnormal Psychology, 112*(4), 558-577. doi: 10.1037/0021-843x.112.4.558

De Stavola, B. L., Daniel, R. M., Ploubidis, G. B., & Micali, N. (2015). Mediation analysis with intermediate confounding: Structural equation modeling viewed through the causal inference lens. *American Journal of Epidemiology, 181*(1), 64-80. doi: 10.1093/aje/kwu239

Emsley, R., Dunn, G., & White, I. R. (2010). Mediation and moderation of treatment effects in randomised controlled trials of complex interventions. *Statistical Methods in Medical Research, 19*(3), 237-270. doi: 10.1177/0962280209105014

Imai, K., Keele, L., & Tingley, D. (2010). A general approach to causal mediation analysis. *Psychological Methods, 15*(4), 309-334. doi: 10.1037/A0020761

Judd, C. M., & Kenny, D. A. (1981). Process analysis - Estimating mediation in treatment evaluations. *Evaluation Review, 5*(5), 602-619.

Kline, R. B. (2011). *Principles and practice of structural equation modelling*. New York, NY: The Guildford Press.

MacKinnon, D. P. (2001). Mediating variable. In N. J. Smelser & P. B. Baltes (Eds.), *International encyclopedia of the social and behavioural sciences* (pp. 9503-9507). Oxford, UK: Elsevier Science, Ltd.

MacKinnon, D. P. (2008). *Introduction to statistical mediation analysis*. New York, NY: Taylor & Francis Group LLC.

Robins, J. M., & Greenland, S. (1992). Identifiability and exchangeability for direct and indirect effects. *Epidemiology, 3*(2), 143-155.

Sobel, M. (1986). Some new results on indirect effects and their standard errors in covariance structure analysis. In N. B. Tuma (Ed.), *Sociological methodology 1986* (pp. 159-186). Washington, D.C.: American Sociological Association.

VanderWeele, T. J. (2009). Marginal structural models for the estimation of direct and indirect effects. *Epidemiology, 20*(1), 18-26. doi: 10.1097/EDE.0b013e31818f69ce

VanderWeele, T. J., & Vansteelandt, S. (2009). Conceptual issues concerning mediation, interventions and composition. *Statistics and Its Interface*, *2*(4), 457-468.

VanderWeele, T. J., Vansteelandt, S., & Robins, J. M. (2014). Effect decomposition in the presence of an exposure-induced mediator-outcome confounder. *Epidemiology*, *25*(2), 300-306. doi: 10.1097/Ede.0000000000000034

Wright, S. (1920a). Correlation and causation Part 1. Method of path coefficients. *Journal of Agricultural Research*, *20*, 0557-0585.

Wright, S. (1920b). The relative importance of heredity and environment in determining the piebald pattern of guinea pigs. *Proceedings of the National Academy of Sciences of United States of America, 6*(6), 320-332.

Table S1

Indirect and direct effects for time points 1 and 2 (95% percentile bootstrap CI) from simplex model with lagged b paths

Paths and effects	Parameters	R1 Estimate	R2 Estimate	R3 Estimate
		Time point 1		
R -> fy_1 -> y_1	$c'_{(r\#)1}$ x 1	**0.60 [0.40, 0.81]**	0.02 [-0.20, 0.25]	**0.48 [0.26, 0.71]**
Direct effect		**0.60 [0.40, 0.81]**	0.02 [-0.20, 0.25]	**0.48 [0.26, 0.71]**
Total effect		**0.60 [0.40, 0.81]**	0.02 [-0.20, 0.25]	**0.48 [0.26, 0.71]**
		Time point 2		
R -> fm_1 -> fy_2 -> y_2	$a_{(r\#)1}$ x b_L x 1	**0.04 [0.001, 0.08]**	-0.003 [-0.01, 0.01]	**0.05 [0.001, 0.10]**
Indirect effect		**0.04 [0.001, 0.08]**	-0.003 [-0.01, 0.01]	**0.05 [0.001, 0.10]**
R -> fy_1 -> fy_2 -> y_2	$c_{(r\#)\ 1}$ x y_2 x 1	**0.58 [0.39, 0.78]**	0.02 [-0.19, 0.24]	**0.47 [0.25, 0.69]**
R -> fy_2 -> y_2	$c'_{(r\#)2}$ x 1	0.13 [-0.09, 0.33]	0.13 [-0.10, 0.32]	0.19 [-0.04, 0.41]
Direct effect		**0.71 [0.46, 0.97]**	0.15 [-0.12, 0.41]	**0.66 [0.39, 0.92]**
Total effect		**0.75 [0.51, 1.00]**	0.14 [-0.12, 0.41]	**0.70 [0.45, 0.96]**

Note. Significant effects shown in bold font, R_1 R_2 and R_3 = dummy variables for randomized treatment group, fm_1 = latent true mediator score at time point 1, fm_2 = latent true mediator score at time point 2, fm_3 = latent true mediator score at time point 3, fy_1 = latent true outcome score at time point 1, fy_2 = latent true outcome score at time point 2, fy_3 = latent true outcome score at time point 3, y_1 = observed outcome score at time point 1, y_2 = observed outcome score at time point 2, y_3 = observed outcome score at time point 3, *(r#)* in the table indicates that the number of the treatment group of interest (R1, R2 or R3) should be substituted.

Table S2

Indirect and direct effects for time points 1 and 2 (95% percentile bootstrap CI) from simplex model with contemporaneous b paths

Paths and effects	Parameters	R1 Estimate	R2 Estimate	R3 Estimate
		Time point 1		
R -> fm_1 -> fy_1 -> y_1	$a_{(r\#)1}$ x b_C x 1	**0.06 [0.03, 0.10]**	-0.005 [-0.02, 0.01]	**0.08 [0.04, 0.12]**
Indirect effect		**0.06 [0.03, 0.10]**	-0.005 [-0.02, 0.01]	**0.08 [0.04, 0.12]**
R -> fy_1 -> y_1	$c'_{(r\#)1}$ x 1	**0.54 [0.34, 0.73]**	0.02 [-0.19, 0.24]	**0.41 [0.19, 0.63]**
Direct effect		**0.54 [0.34, 0.73]**	0.02 [-0.19, 0.24]	**0.41 [0.19, 0.63]**
Total effect		**0.60 [0.40, 0.80]**	0.02 [-0.20, 0.25]	**0.48 [0.26, 0.71]**
		Time point 2		
R -> fm_1 -> fy_1 -> fy_2 -> y_2	$a_{(r\#)1}$ x b_C x y_2 x 1	**0.06 [0.03, 0.09]**	-0.005 [-0.02, 0.01]	**0.07 [0.04, 0.11]**
R -> fm_1 -> fm_2 -> fy_2 -> y_2	$a_{(r\#)1}$ x m_2 x b_C x 1	**0.06 [0.03, 0.10]**	-0.005 [-0.02, 0.01]	**0.07 [0.04, 0.11]**
R -> fm_2 -> fy_2 -> y_2	$a_{(r\#)2}$ x b_C x 1	0.001 [-0.01, 0.02]	-0.003 [-0.02, 0.01]	-0.002 [-0.02, 0.01]
Indirect effect		**0.13 [0.07, 0.19]**	-0.01 [-0.04, 0.01]	**0.14 [0.08, 0.22]**
R -> fy_1 -> fy_2 -> y_2	$c'_{(r\#)1}$ x y_2 x 1	**0.52 [0.32, 0.71]**	0.02 [-0.18, 0.24]	**0.39 [0.18, 0.61]**
R -> fy_2 -> y_2	$c'_{(r\#)2}$ x 1	0.10 [-0.11, 0.29]	0.13 [-0.09, 0.32]	0.16 [-0.05, 0.38]
Direct effect		**0.62 [0.37, 0.88]**	0.15 [-0.11, 0.41]	**0.55 [0.28, 0.80]**
Total effect		**0.75 [0.50, 0.99]**	0.14 [-0.12, 0.41]	**0.70 [0.44, 0.96]**

Note. Significant effects shown in bold font, R_1 R_2 and R_3 = dummy variables for randomized treatment group, fm1 = latent true mediator score at time point 1, fm2 = latent true mediator score at time point 2, fy1 = latent true outcome score at time point 1, fy2 = latent true outcome score at time point 2, y1 = observed outcome score at time point 1, y2 = observed outcome score at time point 2, *(r#)* in the table indicates that the number of the treatment group of interest (R1, R2 or R3) should be substituted.

Table S3

Indirect and direct effects for time points 1 and 2 (95% percentile bootstrap CI) from latent growth model

Paths and effects	**Parameters**	**R1** **Estimate**	**R2** **Estimate**	**R3** **Estimate**
		Time point 1		
Indirect effect				
R -> SM -> SY -> y1	$a_{(r\#)}$ x b x 1	**0.23 [0.10, 0.45]**	-0.02 [-0.09, 0.02]	**0.25 [0.12, 0.50]**
Direct effect				
R -> SY -> y1	$c'_{(r\#)}$ x 1	**0.33 [0.05, 0.53]**	0.11 [-0.07, 0.28]	0.25 [-0.06, 0.47]
Total effect		**0.55 [0.39, 0.73]**	0.08 [-0.10, 0.27]	**0.51 [0.34, 0.69]**
		Time point 2		
Indirect effect				
R -> SM -> SY -> y2	$a_{(r\#)}$ x b x 1.41	**0.32 [0.14, 0.64]**	-0.04 [-0.13, 0.04]	**0.36 [0.16, 0.71]**
Direct effect				
R -> SY -> y2	$c'_{(r\#)}$ x 1.41	**0.46 [0.07, 0.75]**	**0.15 [-0.10, 0.40]**	0.36 [-0.08, 0.67]
Total effect		**0.78 [0.56, 1.03]**	0.12 [-0.14, 0.38]	**0.72 [0.48, 0.97]**

Note. Significant effects shown in bold font, R_1 R_2 and R_3 = dummy variables for randomized treatment group, SM = slope of the mediator, SY = slope of the outcome, y_1 = observed outcome score at time point 1, y_2 = observed outcome score at time point 2, *(r#)* in the table indicates that the number of the treatment group of interest (R1, R2 or R3) should be substituted.

Table S4

Indirect and direct effects for time points 1 and 2 (95% percentile bootstrap CI) from latent change model

Paths and effects	**Parameters**	**R1** **Estimate**	**R2** **Estimate**	**R3** **Estimate**
		Time point 1		
R -> fm_1 -> fy_1 -> y_1	$a_{(r\#)1}$ x b x 1	**0.56 [0.33, 0.88]**	-0.04 [-0.17, 0.09]	**0.65 [0.41, 0.99]**
Indirect effect		**0.56 [0.33, 0.88]**	-0.04 [-0.17, 0.09]	**0.65 [0.41, 0.99]**
R -> fy_1 -> y_1	$c'_{(r\#)1}$ x 1	0.04 [-0.31, 0.33]	0.05 [-0.18, 0.28]	-0.16 [-0.54, 0.13]
Direct effect		0.04 [-0.31, 0.33]	0.05 [-0.18, 0.28]	-0.16 [-0.54, 0.13]
Total effect		**0.60 [0.40, 0.81]**	0.02 [-0.20, 0.24]	**0.49 [0.26, 0.71]**
		Time point 2		
R -> fm_1 -> fy_1 -> y_2	$a_{(r\#)1}$ x b x 1	**0.56 [0.33, 0.88]**	-0.04 [-0.17, 0.09]	**0.65 [0.41, 0.99]**
R -> fm_2 -> fy_2 -> y_2	$a_{(r\#)2}$ x b x 1	-0.02 [-0.14, 0.10]	-0.02 [-0.15, 0.10]	-0.06 [-0.20, 0.06]
Indirect effect		**0.54 [0.34, 0.84]**	-0.06 [-0.20, 0.07]	**0.59 [0.37, 0.90]**
R -> fy_1 -> y_2	$c'_{(r\#)1}$ x 1	0.04 [-0.31, 0.33]	0.05 [-0.18, 0.28]	-0.16 [-0.54, 0.13]
R -> fy_2 -> y_2	$c'_{(r\#)2}$ x 1	0.18 [-0.03, 0.40]	0.14 [-0.08, 0.36]	**0.29 [0.08, 0.51]**
Direct effect		0.23 [-0.13, 0.53]	0.20 [-0.07, 0.45]	0.13 [-0.25, 0.44]
Total effect		**0.77 [0.53, 1.01]**	0.14 [-0.11, 0.41]	**0.72 [0.47, 0.98]**

Note. Significant effects shown in bold font, R_1 R_2 and R_3 = dummy variables for randomized treatment group, fm_1 = latent true change in the mediator between baseline and the first time point, fm_2 = latent true change in the mediator between the first and second time points, fy_1 =latent true change in the outcome between baseline and the first time point, fy_2 = latent true change in the outcome between the first and second time point, y_1 = observed outcome score at time point 1, y_2 = observed outcome score at time point 2, *(r#)* in the table indicates that the number of the treatment group of interest (R1, R2 or R3) should be substituted.

Table S5

Indirect and direct effects for time points 1 and 2 (95% percentile bootstrap CI) from modified latent change model

Paths and effects	**Parameters**	**R1** **Estimate**	**R2** **Estimate**	**R3** **Estimate**
		Time point 1		
R -> fm_1 -> fy_1 -> y_1	$a_{(r\#)1}$ x b x 1	**0.26 [0.14, 0.46]**	-0.02 [-0.08, 0.04]	**0.30 [0.16, 0.53]**
Indirect effect		**0.26 [0.14, 0.46]**	-0.02 [-0.08, 0.04]	**0.30 [0.16, 0.53]**
R -> fy_1 -> y_1	$c'_{(r\#)1}$ x 1	**0.34 [0.07, 0.57]**	0.04 [-0.18, 0.26]	0.18 [-0.10, 0.43]
Direct effect		**0.34 [0.07, 0.57]**	0.04 [-0.18, 0.26]	0.18 [-0.10, 0.43]
Total effect		**0.60 [0.40, 0.80]**	0.02 [-0.20, 0.25]	**0.48 [0.26, 0.71]**
		Time point 2		
R -> fm_1 -> fy_1 -> y_2	$a_{(r\#)1}$ x b x 1.02	**0.27 [0.15, 0.45]**	-0.02 [-0.08, 0.04]	**0.31 [0.17, 0.51]**
R -> fm_2 -> fy_2 -> y_2	$a_{(r\#)2}$ x b x 1	**0.10 [0.03, 0.17]**	-0.02 [-0.09, 0.03]	**0.11 [0.04, 0.19]**
Indirect effect		**0.37 [0.21, 0.58]**	-0.04 [-0.13, 0.04]	**0.42 [0.26, 0.64]**
R -> fy_1 -> y_2	$c'_{(r\#)1}$ x 1.02	**0.35 [0.07, 0.60]**	0.04 [-0.18, 0.25]	0.18 [-0.10, 0.46]
R -> fy_2 -> y_2	$c'_{(r\#)2}$ x 1	0.05 [-0.14, 0.25]	0.14 [-0.07, 0.34]	0.12 [-0.10, 0.35]
Direct effect		**0.40 [0.10, 0.68]**	0.18 [-0.08, 0.43]	0.30 [-0.01, 0.58]
Total effect		**0.76 [0.53, 1.02]**	0.14 [-0.12, 0.41]	**0.73 [0.47, 0.98]**

Note. Significant effects shown in bold font, R_1 R_2 and R_3 = dummy variables for randomized treatment group, fm_1 = modified latent true change in the mediator between baseline and the first time point, fm_2 = modified latent true change in the mediator between the first and second time points, fy_1 = modified latent true change in the outcome between baseline and the first time point, fy_2 = modified latent true change in the outcome between the first and second time point, y_1 = observed outcome score at time point 1, y_2 = observed outcome score at time point 2, *(r#)* in the table indicates that the number of the treatment group of interest (R1, R2 or R3) should be substituted.

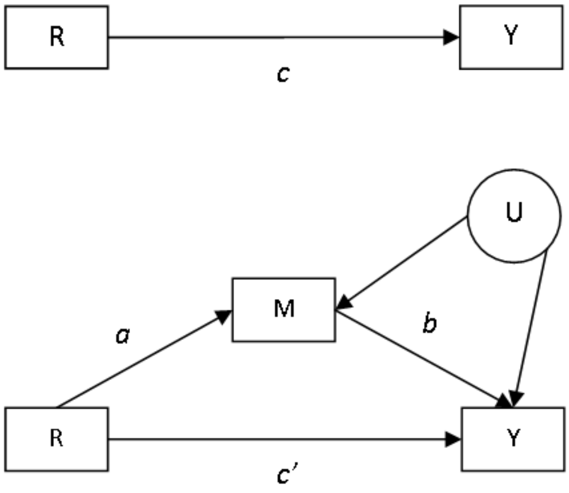


Figure S1. Treatment effect and mediation path diagrams. R = randomized treatment, Y = outcome, M = mediator, U = unmeasured confounders.

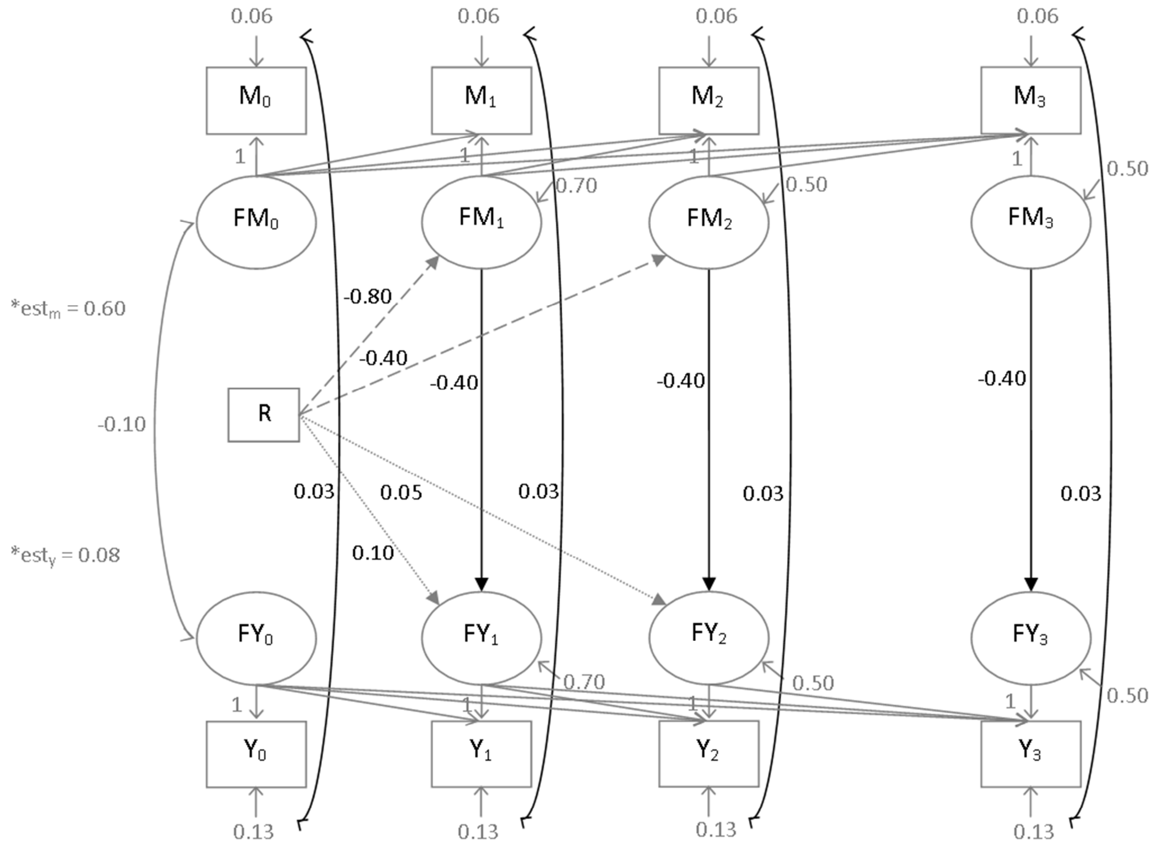


Figure S2. Parameters used to create simulated dataset. *est_m = estimates for all mediator measure factor loadings except at the same time point, which is set = 1 to provide the latent variable scale, this estimate is set equal and so is the same for all factor loadings, est_y = as for the mediator, but in reference to the outcome measure. M_0, M_1, M_2, M_3 = mediator measurements taken at baseline, 1st follow-up time point, 2nd follow-up time point and 3rd follow-up time point post-randomization, Y_0, Y_1, Y_2, Y_3 = outcome measurements taken at the same time points, FM_0 = true latent mediator score at baseline, FM_1, FM_2, FM_3 = modified true latent mediator change between each time point and the previous time point at 1st follow-up time point, 2nd follow-up time point and 3rd follow-up time point post-randomization, FY_0 = true latent outcome score at baseline FY_1, FY_2, FY_3 = modified true latent outcome change between each time point and the previous time point.

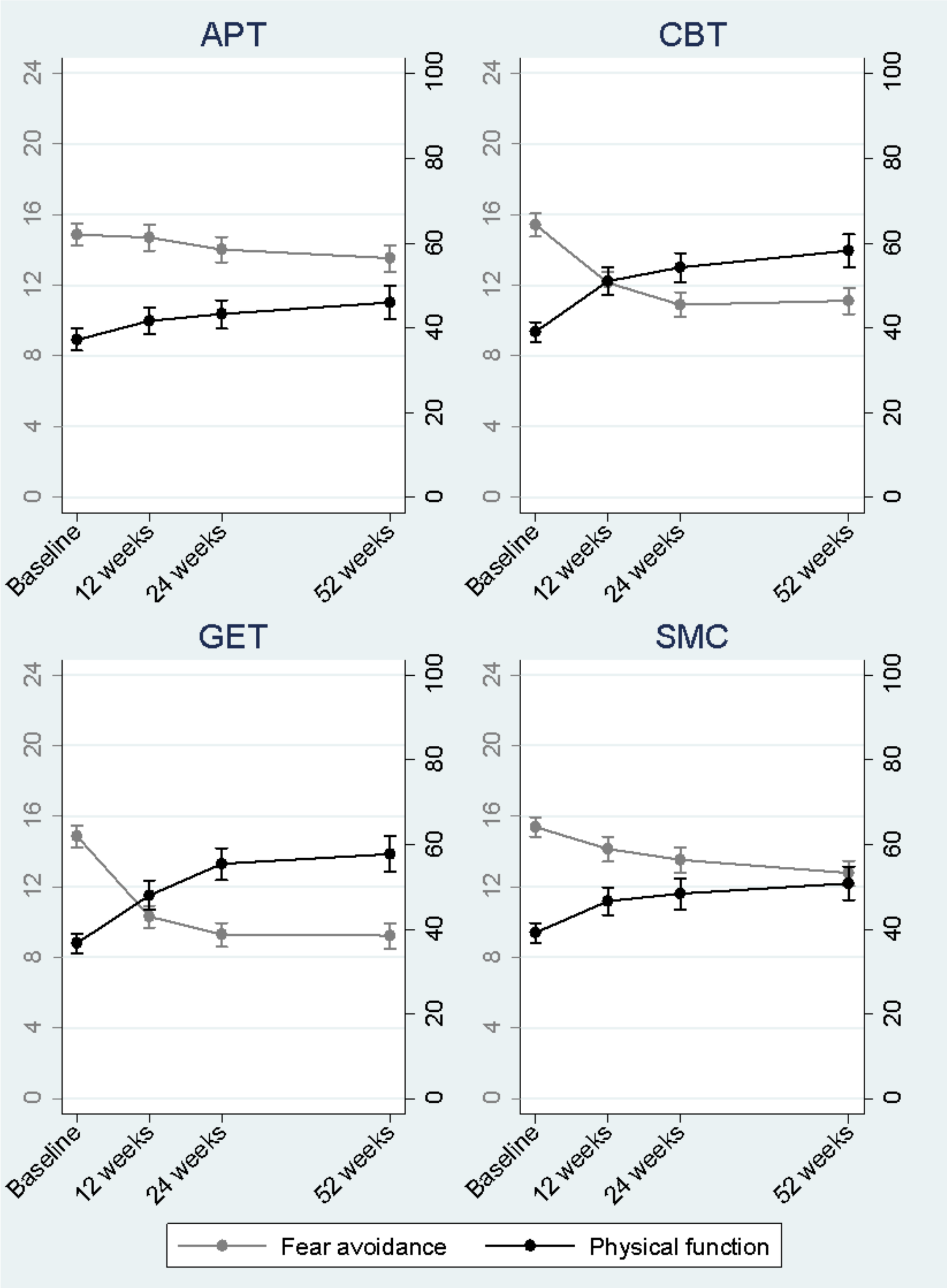
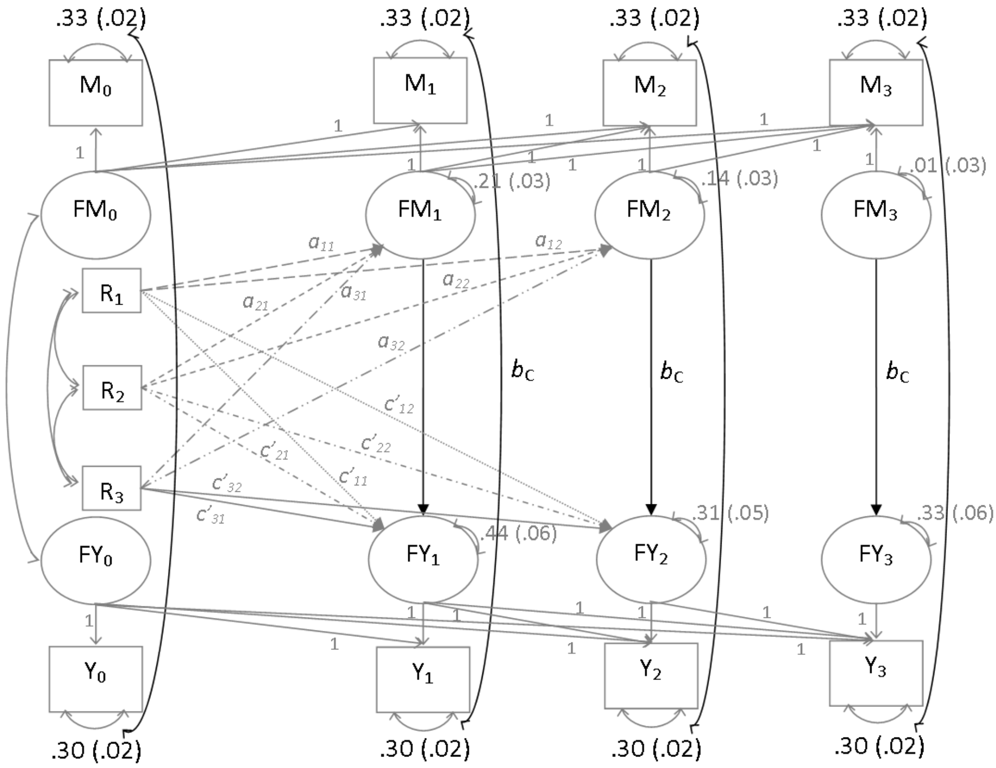


Figure S3. Profile plot of fear avoidance mediator and physical function outcome by treatment group in PACE data. APT = adaptive pacing therapy, CBT = cognitive behavioral therapy, GET = graded exercise therapy, SMC = specialist medical care.



a_{11}	**-.83 [-1.02, -.63]**	a_{12}	.03 [-.14, .20]	c'_{11}	.04 [-.31, .33]	c'_{12}	.18 [-.03, .40]
a_{21}	.05 [-.13, .23]	a_{22}	.03 [-.14, .21]	c'_{21}	.05 [-.18, .28]	c'_{22}	.14 [-.08, .36]
a_{31}	**-.96 [-1.17, -.78]**	a_{32}	.09 [-.08, .26]	c'_{31}	-.16 [-.54, .13]	c'_{32}	**.29 [.08, .52]**
b_C	**-.68 [-.99, -.45]**						

Paths and effects for third time point	Parameters	R1 Estimate	R2 Estimate	R3 Estimate
R -> fm_1 -> fy_1 -> y_3	$a_{(\#)1} \times b_C \times 1$	**0.56 [0.33, 0.88]**	-0.04 [-0.17, 0.09]	**0.65 [0.41, 0.99]**
R -> fm_2 -> fy_2 -> y_3	$a_{(\#)2} \times b_C \times 1$	-0.02 [-0.14, 0.10]	-0.02 [-0.15, 0.10]	-0.06 [-0.20, 0.06]
Overall indirect effect		**0.54 [0.34, 0.84]**	-0.06 [-0.20, 0.07]	**0.59 [0.37, 0.90]**
R -> fy_1 -> y_3	$c'_{(\#)1} \times 1$	0.04 [-0.31, 0.33]	0.05 [-0.18, 0.28]	-0.16 [-0.54, 0.13]
R -> fy_2 -> y_3	$c'_{(\#)2} \times 1$	0.18 [-0.03, 0.40]	0.14 [-0.08, 0.36]	**0.29 [0.08, 0.51]**
Overall direct effect		0.23 [-0.13, 0.53]	0.20 [-0.07, 0.45]	0.13 [-0.25, 0.44]
Total effect		**0.77 [0.53, 1.01]**	0.14 [-0.12, 0.41]	**0.72 [0.47, 0.98]**

Figure S4. Four group dual process latent change score model with contemporaneous mediation and residual covariance paths. Numbers in round brackets are standard errors, numbers in square brackets are 95% confidence intervals. The lower table shows indirect and direct effect estimates for the third post-randomization time point. Significant effects shown in bold font, R_1 R_2 and R_3 = dummy variables for randomized treatment group, M_0, M_1, M_2, M_3 = mediator measurements taken at baseline, 1st follow-up time point, 2nd follow-up time point and 3rd follow-up time point post-randomization, Y_0, Y_1, Y_2, Y_3 = outcome measurements taken at the same time points, FM_0 = true latent mediator score at baseline, FM_1, FM_2, FM_3 = true latent mediator change between each time point and the previous time point, FY_0 = true latent outcome score at baseline FY_1, FY_2, FY_3 = true latent outcome change between each time point and the previous time point, *(r#)* in the table indicates that the number of the treatment group of interest (R1, R2 or R3) should be substituted.