

Measurement error, time lag, unmeasured confounding: considerations for longitudinal estimation
of the effect of a mediator in randomised clinical trials

Goldsmith, KA^{1,*}, Chalder, TC², White, PD³, Sharpe, M⁴, Pickles, A¹

1. Department of Biostatistics, King's College London, DeCrespigny Park, London, SE5 8AF.
2. Academic Department of Psychological Medicine, King's College London, Weston Education Centre, Cutcombe Road, London, SE5 9RJ.
3. Centre for Psychiatry, Wolfson Institute of Preventive Medicine, Barts and the London School of Medicine, Queen Mary University, London, EC1M 6BQ.
4. Psychological Medicine Research, Department of Psychiatry, University of Oxford, OX3 7JX.

*Corresponding author

Address for correspondence:

Institute of Psychiatry, Psychology & Neuroscience

King's College London

DeCrespigny Park

London, UK

SE5 8AF

Tel: +44(0)207 848 0847

Fax: +44(0)207 848 0281

Email: kimberley.goldsmith@kcl.ac.uk

Abstract

Clinical trials are expensive and time-consuming and so should also be used to study how treatments work. This would allow evaluation of theoretical treatment models and refinement and improvement of treatments. Treatment processes can be studied using mediation analysis.

Randomised treatment makes some of the assumptions of mediation models plausible, but the mediator – outcome relationship remains one that can be subject to bias. In addition, mediation is assumed to be a temporally ordered longitudinal process, but most mediation studies to date have been cross-sectional and unable to explore this assumption. This study used longitudinal structural equation modelling of mediator and outcome measurements from the PACE trial of rehabilitative treatments for chronic fatigue syndrome (ISRCTN 54285094) to address these issues. In particular, autoregressive and simplex models were used to study measurement error in the mediator, different time lags in the mediator – outcome relationship, unmeasured confounding of the mediator and outcome, and the assumption of a constant mediator – outcome relationship over time. Results showed that allowing for measurement error and unmeasured confounding were important. Concurrent rather than lagged mediator – outcome effects were more consistent with the data, possibly due to the wide spacing of measurements. Assuming a constant mediator-outcome relationship over time increased precision.

Keywords: mediation, longitudinal mediation models, structural equation models, measurement error, clinical trials, chronic fatigue syndrome

Introduction

The primary objective of randomised clinical trials is to answer the question “does the treatment work”? However, researchers generally have other important questions about treatment, such as how treatments work, and the testing of theoretical models of treatment mechanisms – essential knowledge for future therapeutic development. The case is especially strong when a trial shows no treatment differences, when it is often critical to obtain information about why the treatment failed. Trials are expensive and can be demanding for both patients and clinical staff. Answering questions beyond those of treatment effectiveness can be justified in terms of marginal cost, burden and ethics, and in many instances may require little more than additional analyses.. Such investigations often rely heavily on mediation analysis. While mediation analysis is in very common use, with one of the source works by Baron and Kenny having been cited more than 10,000 times, studies applying these methods rarely discuss the problems of variables measured with error and post-randomisation confounding. This paper describes how these problems can be addressed in a structural equation modelling framework.

A mediation model is depicted at the bottom of Figure 1, each path labelled as is traditional for this field. Under certain no unmeasured confounding assumptions, correct model specification and when the mediator and outcome are continuous, mediation can be assessed through a series of ordinary least squares regression equations where Y = outcome, M = mediator and R = randomly assigned treatment ¹:

$$Y_i = \alpha_1 + \beta_1 R_i + \varepsilon_{i1} \quad (1)$$

$$M_i = \alpha_2 + \beta_2 R_i + \varepsilon_{i2} \quad (2)$$

$$Y_i = \alpha_3 + \beta_3 R_i + \gamma M_i + \varepsilon_{i3} \quad (3).$$

β_1 gives the total effect of treatment on outcome, referred to in the mediation literature as the c path. In the past it was suggested that mediation should only be assessed when this total effect was significant, but more recently we and others have argued that mediation should be assessed whether or not β_1 is significant²⁻⁷. In order for there to be mediation, β_2 in equation 2 (a in Figure 1), should be significant; i.e. treatment has an effect on the mediator. In addition, in equation 3 which corresponds to b in Figure 1, γ should be significant, i.e. the mediator has an effect on the outcome. The indirect, or mediated, effect is then given by $\beta_2\gamma$ (or the product a times b), known as the product of coefficients (POC) estimate⁸. Any remaining direct effect of treatment on the outcome is given by β_3 , which is referred to as c' . In the situation where β_1 is not significant, studying these relationships could clarify why the treatment was not effective. The estimated a path shows whether the treatment had any effect on the mediator, while the estimated b path clarifies whether the mediator is predictive of the outcome. Non-significance of either of these estimates could partly explain the ineffectiveness of a treatment. Studying the a , b and c' paths together will also indicate whether the direction of the c' path or direct effect is different from that of the indirect effect. This is referred to as inconsistent mediation or suppression and can be associated with a non-significant overall effect³. The POC mediated effect estimate is a simple example of the application of Sewell Wright's path tracing rules, which takes the products of the estimates along each legal path between two variables and then sums the products across all of the paths^{9,10}. Paths cannot be traced: 1) where an arrow is first traced in a forward direction and the next arrow in a backwards direction, 2) through a variable more than once and 3) through more than one bidirectional arrow. It is important to note that this tracing and indirect effect calculation assumes linear relationships amongst the variables in the model and conditional multivariate normality^{11,12}. The models in (2) and (3) can be fitted simultaneously using the structural equation model (SEM) framework. We focus here on how

the framework allows some progress to be made in relation to two of the likely major sources of bias in mediation effect estimates, namely measurement error in the mediator and confounding of the mediator-outcome relationship.

Classical measurement error in the mediator results in a bias towards zero in the estimate of the mediator, so-called attenuation or dilution¹³. The single mediator/single outcome model of Figure 1 can account for measurement error if a good estimate of the reliability of the mediator is known¹⁴, and other regression-based methods have also recently been applied to address this problem¹⁵.¹⁶ However, SEM allows for the simultaneous estimation of multiple equations and with repeated or multivariate measurement this allows the incorporation of measurement error using latent variables^{11, 12, 17}. The latent or “true score” is a hypothetical quantity relating to the error-free measure of the construct of interest, which is defined in practice from a decomposition of covariances^{11, 12}. In SEM with repeated measures this commonly follows a classical measurement error model, in which the variance of a measure is partitioned into a latent ‘true score’ and an additive error which is assumed conditionally independent¹¹⁻¹³. The error is assumed to be independent of this true latent mediator score given exposure and covariates. The effect of a latent mediator variable on the outcome would provide an estimate of the mediator – outcome relationship disattenuated for measurement error^{11, 12}.

In randomised clinical trials, the mediator is not generally randomised, and as a consequence, the relationship between the mediator and outcome could be subject to confounding^{3, 18-21}. The need to make the assumption of no unmeasured confounding has received much attention in the literature^{2, 3, 22-30}. Extending the range of measured confounders is recommended but without the addition of an instrumental variable, which in practice is rarely available^{4, 14}, we cannot add an error covariance between mediator and outcome to the simple model of Figure 1 as the model will no longer be identified – there would be six parameters to estimate (three regression parameters, two error variances and an error covariance) and only five quantities available for estimation (two variances and three covariances). Quantifying parameter estimate sensitivity to unmeasured confounding provides another option. For example, Vanderweele describes general direct and indirect effects bias formulae for sensitivity analysis²⁷, while Imai and colleagues use the SEM framework as a starting point and provide methods to quantify the sensitivity of the mediator – outcome relationship to the correlation between mediator and outcome error terms^{23, 24}. With repeated measurement, we highlight that relaxation of the conditional independence assumption of SEM measurement models can allow for correlation among the measurement errors of mediator and outcome that some unmeasured confounders may induce. This has allowed us to take a more model selection-based approach using model fit criteria and perusal of the effect of different SEM on the mediator – outcome relationship estimates. Fitting different SEM in this way could be thought of another type of sensitivity analysis. This method has some advantages over other sensitivity analysis solutions that have been offered^{15, 16, 23, 24}, in that it allows for measurement error and unmeasured confounding in a single step, and unlike regression based methods, doesn't require the provision of sensitivity parameters such as measurement reliability and measurement error variance^{15, 16}. Such parameters are likely to be difficult to estimate accurately. De Stavola and colleagues extended the sensitivity analyses described by Imai et al to the situation where there are confounders of the mediator and outcome that are affected by the treatment or “intermediate

confounders”^{22, 25, 26, 28, 29}. We make the assumption for the purposes of this paper that we have no “intermediate confounding”..

The study that motivated this work was the **Pacing, Graded Activity, and Cognitive Behaviour Therapy: A Randomised Evaluation or PACE** study of rehabilitative treatments for Chronic Fatigue Syndrome³¹, which is described in more detail in the Methods section. In an initial study of mediation of the PACE treatments using the POC method^{3,8}, the mid-treatment measurement of the mediators and the follow-up measure of outcomes showed that the effects of CBT and GET were partially mediated through cognitive and behavioural factors³². However, this initial analysis did not take advantage of all the available repeated measures data. Mediation is regarded as a causal process, where the temporal ordering of treatment -> mediator -> outcome is implied. The repeated measures of mediator and outcome data in the PACE trial allow for exploration of time lags in mediated effects in addition to measurement error and unmeasured confounding.

The aim of this work was to address issues affecting estimation and precision of the mediator – outcome relationship in longitudinal mediation models. These issues were: measurement error in the mediator, time lags of mediator – outcome relationship, unmeasured confounding and the assumption of a constant mediator – outcome relationship over time. The repeated measures of mediators and outcomes available in the PACE data were used to address this aim.

Methods

Motivating study

The PACE trial of treatments for chronic fatigue syndrome (CFS) randomised individuals to four treatment groups defined as three therapies each in addition to the control treatment, and the

control treatment alone ³¹. The therapies were cognitive behavioural therapy (CBT), graded exercise therapy (GET), adaptive pacing therapy (APT), with the control being specialist medical care (SMC) delivered by a doctor with experience of treating CFS. CBT was delivered by psychotherapists and focused on examining the link between thoughts, behaviours and symptoms and negotiating behavioural and cognitive changes with the patient as appropriate. It was based on a fear avoidance model of CFS ³². GET was delivered by physiotherapists and involved a mutually agreed gradual increase in activity for the patient. It was based on deconditioning and exercise intolerance models of the illness. APT was delivered by occupational therapists. It was based on the model of a finite and limited amount of energy and advised participants to balance activity with rest. The trial found that CBT and GET were superior to APT and SMC for both physical functioning and fatigue outcomes. Measures of the mediators and outcomes were taken at baseline, mid-treatment (12 weeks post-randomisation), post-treatment (24 weeks post-randomisation) and follow-up (52 weeks post-randomisation). The West Midlands Multicentre Research Ethics Committee (MREC 02/7/89) approved the original PACE study, ISRCTN 54285094. This trial provides a unique opportunity to study longitudinal mediation given that most mediators were measured at all time points, the mediators and outcomes were measured mid-therapy, which is uncommon, and there were four treatment groups rather than the usual two treatment group design. The data on the subset of the participants randomised to receive CBT and APT were used for this analysis for simplicity. Advantages associated with the four treatment group design and longitudinal mediation results for all four treatment groups will be reported elsewhere.

An example mediator, fear avoidance beliefs, and one of the two primary trial outcomes, self-rated physical functioning, were used in this study. Fear avoidance beliefs were measured using the Cognitive Behavioural Responses Questionnaire ^{33, 34}, which is a scale ranging from 0 to 24. Physical functioning was measured with the physical functioning subscale of the SF-36 ^{35, 36}, which ranges

from 0 to 100. Fear avoidance beliefs were found to be the strongest mediator of the effect of CBT and GET as compared to APT in a simple single mediator/outcome POC analysis ³².

Structural Equation Models

Data were standardised using the mean and standard deviation (SD) of the baseline measure, so the units for the mediator – outcome estimates and indirect effects were baseline physical functioning SDs. SEMs were used to model the longitudinal mediation and outcome processes. The only other covariate in the models was treatment group. Mplus version 7.2 was used to fit models; full information maximum likelihood was used, which estimates parameters using information from complete and incomplete records under a missing at random assumption ^{37, 38}. Model fit was assessed using the chi-square test of model fit ¹² and the root mean square error of approximation (RMSEA) ^{39, 40}. Models were compared using informal comparison of the Chi-square statistic, RMSEA and the Bayesian Information Criterion (BIC) ⁴¹.

Longitudinal Structural Equation Models for Mediation

There are several longitudinal SEMs that could be applied in a mediation analysis, which have been reviewed by MacKinnon ³. One such model is a first-order autoregressive as shown in Figure 2A for a mediator process where measures have been taken at the time points in the PACE trial. The model assumes that: each variable is a function only of the measure of that variable at the previous time point (plus other covariates such as treatment group), there is no measurement error and correlations between the measurements decrease over time ^{3, 42, 43}. This may or may not be a plausible assumption for the correlations of the mediator and outcome measurements in PACE. When using such models for mediation, an autoregressive structure is fit to both the mediator and outcome processes separately and then these are linked together by estimates we will refer to

hereafter as the b paths³, as shown in Figure 3. The exposure of interest, here R or the randomised treatment group, is allowed to affect different measures of both the mediator and outcome.

We explored the effects of the following on estimation and precision of the b_1 , b_2 and b_3 paths (Figure 3) within the autoregressive model structure:

- 1) allowing for independent measurement error in the mediators to remove attenuation bias,
- 2) allowing different b_1 , b_2 and b_3 path time lags
- 3) allowing for measurement error covariances between the mediator and outcome variables that would be induced by unmeasured confounding and which could attenuate or inflate the b_1 , b_2 and b_3 path estimates and
- 4) assuming equality of the b_1 , b_2 and b_3 paths over time, which would be consistent with the view that no matter how or when the mediator is changed, its relationship with the outcome remains constant.

Before reporting results we detail how each of these issues were addressed in the modelling. All the models leave the randomised treatment group to be uncorrelated with all other variables.

Measurement error - Autoregressive versus simplex models

Autoregressive models can be extended to allow for measurement error by fitting models with a quasi-simplex structure as shown in Figure 2B for the mediator process. The quasi-simplex model (referred to as simplex hereafter) uses each observed value of the variable as a single indicator for a latent 'true score' variable with the autoregressive relationship for the variable then being between these true scores rather than the observed scores⁴²⁻⁴⁴. In this way, classical measurement error is

incorporated¹¹⁻¹³. So if for example we take the baseline mediator measure, M_0 , it is made up of the true measure, FM_0 , plus the conditionally independent error, ε_{M_0} or $M_0 = FM_0 + \varepsilon_{M_0}$. One way to obtain identification of a simplex model with four measurements is to assume parallel measurement and set the factor loadings and error variances equal over time. This was the approach taken here. The importance of allowing for measurement error was assessed by comparing model fit of dual-process autoregressive and simplex models fitted to both the mediator and outcome processes as shown in Figure 3, and by comparing the b path estimates and their standard errors informally (i.e. without using statistical tests). Figure 4 shows the equations associated with the model in Figure 3C, which assume classical measurement error¹³.

Time lags of mediation paths – lagged and contemporaneous

In each case, an autoregressive or simplex model was fitted with either lagged b_1 , b_2 and b_3 paths between the mediator at the time point prior to the outcome at the time point immediately following, or contemporaneous b_1 , b_2 and b_3 paths between mediator and outcome at the same time point. Lagged paths respect the implied temporal ordering of change in mediator prior to change in outcome that would be expected for a causal process. Models with contemporaneous paths were also fitted despite their lack of allowance for temporal ordering as these sorts of relations were plausible, perhaps especially given the considerable lengths of time between measures.

Autoregressive models with these two types of paths are shown in Figures 3A and 3B. Figures 3C and D show dual process simplex models with contemporaneous b_1 , b_2 and b_3 paths (the covariances are discussed in the next section). Models with lagged and contemporaneous b paths were compared using the BIC and the other model fit indices and informal comparison of the b path estimates and their standard errors.

Unmeasured confounding – lagged and contemporaneous measurement error covariances

One plausible extension of the simplex model is to allow for covariance between measurement errors in the mediator and outcome. This covariance could be thought of as encompassing some type of unmeasured confounding that is captured by correlated measurement errors. This could be covariance in errors at the same time point – where factors on a given day are affecting measurement of both mediator and outcome, or covariance between an earlier measure of the mediator and a later measure of the outcome, where there is a more persistent process in play. Such paths can be allowed for when repeated measurements are available, as in Figure 3C and 3D, which show examples of models with contemporaneous and lagged covariances respectively. Models with different types of covariances were compared as described in the previous section on time lag of mediation paths.

Precision of the b paths in longitudinal mediation models

To potentially increase the precision of the b paths and also in the interest of model parsimony, evidence for setting the measurement error covariances to be equal over time and for setting the b_1 , b_2 and b_3 paths to be equal over time was examined. This was done using Wald chi-square tests for model constraints with the degrees of freedom equal to the number of parameter constraints tested. For the covariances this was initially a test of equality of four contemporaneous covariances or two post-baseline lagged covariances. If this test indicated a significant difference, pairwise tests were done to characterize the differences. For the b paths the Wald test was structured as a test of equality of either three contemporaneous paths (b_1 , b_2 and b_3) or two lagged paths (b_2 and b_3).

One other assumption was explored. Theoretically there may be no need for treatment to mediator (a_3 path) and treatment to outcome (c'_3 path) paths beyond the end of treatment. The need for these paths was assessed using a Wald chi-square test for these two paths being equal to zero.

The assumptions were tested in the models in the order: equal measurement error covariances, need for a_3 and c'_3 paths and equality of b paths. If an assumption could be made, it was incorporated in models going forward.

Calculation of indirect effects in longitudinal mediation models

When we extend mediation models to the longitudinal case there are more indirect or mediated paths than the single path in the simple mediation model. In the in the simplex models we have fitted using the SEM framework, the mediated effects are straightforward to calculate using the path tracing rules described previously^{8-10, 21}. When assuming multivariate normality and linear relationships between variables^{11, 12}, this entails finding each path between two variables, multiplying together all parameter estimates on each path and then summing these products together. Cole and Maxwell point out the key issue of choosing the effects of interest in models with several measures of outcome over time, and suggest these will most likely focus on the final outcome time point, given that data have been gathered to that point for a clinical and/or theoretical reason²¹. They describe “time-specific indirect effects”, which might have one of the earlier outcome time points as the time point of interest, and “overall indirect effects” which are the effects up to the final measurement. Indirect effects are those going through any measure of the mediator.

The main indirect/mediated effect of interest in the PACE example here is the overall indirect effect for the 52 week time point. In order to obtain this effect, all of the time-specific indirect effects

would be calculated for that time point and summed together. An example of one of the time-specific effects that would contribute to the overall indirect effect at 52 weeks in the model in Figure 3A would be $R \rightarrow M_{12} \rightarrow M_{24} \rightarrow M_{52} \rightarrow Y_{52}$ (calculated as $a_1 \times m_2 \times m_3 \times b_3$). An example of a direct or non-mediated effect (does not pass through any measure of the mediator) would be $R \rightarrow Y_{12} \rightarrow Y_{24} \rightarrow Y_{52}$ (calculated as $c_1 \times y_2 \times y_3$). So, the total effect of the randomised treatment R on the final measure of Y_{52} would be:

$$Y_{52} = (a_1 R \cdot m_2 M_{12} \cdot m_3 M_{24} \cdot b_3 M_{52}) + \quad (4)$$

$$(a_1 R \cdot m_2 M_{12} \cdot b_2 M_{24} \cdot y_3 Y_{24}) + \quad (5)$$

$$(a_1 R \cdot b_1 M_{12} \cdot y_2 Y_{12} \cdot y_3 Y_{24}) + \quad (6)$$

$$(a_2 R \cdot m_3 M_{24} \cdot b_3 M_{52}) + \quad (7)$$

$$(a_2 R \cdot b_2 M_{24} \cdot y_3 Y_{24}) + \quad (8)$$

$$(a_3 R \cdot b_3 M_{52}) + \quad (9)$$

$$(c_1 R \cdot y_2 Y_{12} \cdot y_3 Y_{24}) + \quad (10)$$

$$(c_2 R \cdot y_3 Y_{24}) + \quad (11)$$

$$(c_3 R) \quad (12)$$

Expressions (4) through (9) are time-specific indirect effects, which summed together, constitute the overall indirect effect. Expressions (10), (11) and (12) are three time-specific direct effects (for the 52 week time point), which summed together, constitute the overall direct effect. The total effect can be obtained as shown by summing all effects, or by summing the overall indirect and direct effects. These effects can be easily extended to the simplex models by replacing the observed variables by their latent true score counterparts. The target parameters are the same in the simplex and

autoregressive models, but the estimates differ; in the case of the simplex model the parameters incorporate a model-based correction for measurement error. The indirect and direct effects can be assumed to be causal if there is no residual unmeasured confounding or measurement error, no “intermediate confounders”, the linearity and multivariate normality assumptions are met and the models are otherwise properly specified. While we focus here on making these assumptions and taking the SEM approach, causal estimators could be elucidated using potential outcomes, following on from the multiple causally ordered mediator estimands put forward by Daniel et al ⁴⁵. CI for these effects were calculated using 1000 repetitions of the bias-corrected bootstrap ^{46, 47}.

Simulation study addressing the consequences of ignoring measurement error and unmeasured confounding

Some model assumptions were addressed using simulations. Data were simulated and analysed using the Mplus MONTECARLO command. Data were generated under the following three models:

1. Simplex model without measurement error or unmeasured confounding (error covariances) to approximate the autoregressive models (as in Figure 6C with $\epsilon = 1 \times 10^{-10}$)
2. Simplex model allowing for measurement error, no unmeasured confounding (error covariances, as in Figure 6C)
3. Simplex model allowing for measurement error and unmeasured confounding (error covariances, as in Figure 6E).

The data generated were then analysed under each of the three types of model in turn. The analyses with equivalent data generation and analysis models served as controls. The true parameter values were obtained by fitting the model of interest to the PACE data, saving the parameter estimates and using them as the true estimates in the MONTECARLO program. The simulation study was based on the two group CBT versus APT comparison, with datasets of $n = 50, 100, 320$ (the number in two

groups in the PACE trial), 640 (the number in four groups in the PACE trial) and 1000 were simulated, with 1500 repetitions in each case⁴⁸. Absolute bias was calculated by subtracting the true parameter values from the average parameter estimate across repetitions. Mean square error (MSE) and coverage across repetitions were obtained from the Mplus output.

Results

The model diagrams are shown in Figures 5 (lagged b paths) and 6 (contemporaneous b paths).

Measurement error

Model fit indices show that the simplex models fitted better than the comparable autoregressive models (Table 1), suggesting it was important to allow for measurement error. For example, for the models with contemporaneous b paths, the simplex model BIC minus the autoregressive model BIC was -24 (comparing Figure 6A and 6B). For the models with lagged b paths, accounting for measurement error had the expected disattenuation effect, i.e. the simplex model b paths were larger in magnitude (in Figure 5B versus Figure 5A and Supplementary Table A). However for models with contemporaneous b paths, the opposite was true (Figure 6B versus Figure 6A and Supplementary Table A). The effect of taking measurement error into account was also seen in the magnitude of paths between the previous measure of the mediator/outcome and the subsequent measure (i.e. the m and y paths in Figure 3C). These paths were of larger magnitude in the simplex models (for example, compare Figures 5A and 5B). Accounting for measurement error led to a small loss in precision (e.g. comparing the b paths from autoregressive model Figure 6A to simplex model Figure 6B, also see Figure 7).

Evidence suggested that the α_3 and c'_3 paths to the final outcome time point could be set equal to zero in the simplex models ($p > 0.43$ for all covariance/ b path lag combinations), so they were set equal going forward.

Time lags of mediation paths

The BIC suggested that models with contemporaneous mediation paths fitted better than models with lagged paths (Table 1). For example, the BIC for the simplex model with contemporaneous b paths minus the BIC for the model with lagged b paths was -31 (comparing Figures 5C and 6C). The contemporaneous b paths were generally larger in magnitude than the lagged b paths (Figure 7 and Supplementary Table A).

Unmeasured confounding

In both the cases of the lagged and contemporaneous error covariances, the overall tests indicated that not all covariances could be assumed equal ($p < 0.004$ for lagged and $p < 0.030$ for contemporaneous). For the lagged covariances, there was no evidence that the 12 – 24 week and 24 – 52 week covariances were different ($p > 0.77$), so these were set equal. For the contemporaneous covariances, the general pattern was that the baseline and 24 week covariances were of a smaller magnitude than those at 12 and 52 (data not shown). When considered from the point of view of unmeasured confounding there would seem little basis for a hypothesis of a constant covariance. However, when considered from a measurement model perspective a parsimonious model would suggest uniform covariances as plausible. The RMSEA and AIC were very similar in models with and without these covariances set equal, with the BIC being smaller in the models with equal covariances. For example, the model shown in Figure 6E with equal contemporaneous covariances had a BIC difference of 7 when compared to a model where the covariances were freed (data not

shown). Equal covariances were therefore used in models going forward. The covariances were generally statistically significant in models where they were set equal (see Figures 5D and E and 6D and E).

The inclusion of contemporaneous error covariances had greater impact than the inclusion of lagged error covariances. For example, in the three models with lagged b paths in Figures 5C, D and E, the BICs for the models with no and with lagged error covariances were very similar (Table 1, Figure 5C compared to Figure 5D), however, the differences in BIC for the model with contemporaneous error covariances and the model with none was -64 (Table 1, Figure 5E compared to Figure 5C). There was a similar pattern for models with contemporaneous b paths. The only models with RMSEA values consistent with good model fit were those with contemporaneous error covariances (Table 1). In models with contemporaneous b paths, including contemporaneous covariances decreased the magnitude of the b path somewhat (comparing the models in Figures 6C and 6E, see also Figure 7).

Precision of the b paths

Tests for equality of b paths over time showed that these could be set equal both for lagged paths (for equality of b_2 and b_3 paths, $p = 0.92$ for model with lagged covariances, $p = 0.84$ for contemporaneous covariances) and contemporaneous paths (for equality of b_1 , b_2 and b_3 paths, $p = 0.13$ for model with lagged covariances, $p = 0.12$ for contemporaneous covariances). Making this assumption of equal b paths at each time point gave a large precision gain; for example, comparing Figure 6B to 6C, the smallest standard error in Figure 6B with free b paths was 0.057 and in Figure 6C with a common b path was 0.030, giving a 47% relative increase in precision. Assuming equal error covariances over time also increased the precision of the b path, but by only a small amount in most cases (data not shown).

Type of covariances versus lag of b paths

In the simplex model it appeared that the type of measurement error covariances was more important than the type of mediation paths. For example, in terms of model fit there was little to choose between the two models having contemporaneous covariances but different types of *b* paths in Figure 5E and 6E (Table 1). On the other hand, when comparing models with the same type of *b* paths but contemporaneous versus lagged covariances (for example, Figure 5E versus 5D), it is clear that the former fitted the data much better (Table 1).

Best fitting model

In summary, the most plausible model by RMSEA, AIC and BIC was the simplex with contemporaneous *b* paths and error covariances (Table 1 and Figure 6E). Because this model had no treatment to 52 week mediator or outcome paths it assumed no direct effect of treatment on physical functioning at 52 weeks, and that any residual effects of treatment had persisted from the post-treatment time point (24 weeks).

Indirect effects in longitudinal simplex mediation models

Figure 8 shows the total indirect/mediated effects for each time point, with more detail of the estimates and CI for the various time specific indirect, direct and total and overall effects in Tables B, C and D in the supplementary material. Figure 8 shows the smaller *b* paths in the lagged models led to smaller indirect effects as would be expected. The figure also shows that using path analysis rules to calculate mediated effects from the simplex models gives effects that accumulate over time.

The overall effect of treatment group on physical functioning at 52 weeks in the best fitting model with contemporaneous b paths and error covariances (Figure 6E) was 0.69 physical functioning SD units (95% CI 0.44, 0.95) (Figure 8 and Supplementary Table D). The time specific indirect/mediated effect for this time point indicated that CBT increased physical functioning through the fear avoidance mediator by 0.28 physical functioning SD units (0.15, 0.44) (Figure 8 and Supplementary Table D). This suggested that approximately 41% of the effect of treatment on physical functioning was mediated through fear avoidance when all repeated measures of both mediator and outcome were modelled and measurement error and unmeasured confounding were taken into account.

The common b path assumption that led to greater precision in the b paths also led to greater precision in the indirect effects. In the model shown in Figure 6E, the standard error (SE) for the 52 week indirect effect with this assumption was 0.068, giving a CI of 0.15 to 0.44 (Figure 8). If the common b path assumption was relaxed in this model the SE was 0.073 and CI 0.18 to 0.47.

Consequences of ignoring measurement error and unmeasured confounding

Figures 9 and 10 show the absolute bias, MSE and coverage results from the simulation study. Each simulation took 25 seconds or less to run using the Mplus program. Figure 9 shows effects of ignoring measurement error when it is present on the a_1 , a_2 , b , m_2 , m_3 , y_2 and y_3 paths and Figure 10 shows the effect of ignoring measurement error and confounding when they are present on the same paths (the numeric data are available upon request). Models converged across most repetitions in all conditions, with the exceptions occurring when $n = 50$ (Supplementary Table E). There were problems with the latent variable covariance matrix and the residual covariance matrix for simulations where data were generated and analysed with measurement error, which were alleviated in the case of larger sample sizes. There were also problems with the residual covariance

matrix for simulations where data were generated without and analysed allowing for measurement error that did not dissipate with increasing sample size. There was little evidence of effects on bias, MSE and coverage when allowing for measurement error and unmeasured confounding in the case where these weren't present (Supplementary Figure A and Figure 9 middle row). However, there were issues with the statistical quantities for most of the paths of interest in the models when measurement error was present but not accounted for (Figures 9 and 10, bottom rows) and to a lesser extent when only unmeasured confounding was ignored (Figure 10, middle row). The effects were somewhat worse when both measurement error and confounding were present and ignored (Figure 10, bottom row) as compared to measurement error alone (Figure 9, bottom row). . Effects tended to be worse for the paths connecting the mediator and outcome measures through time, i.e. the m_2 , m_3 , y_2 and y_3 paths as shown in Figure 3C. This is important because these paths contribute to longitudinal mediated effects (see *Calculation of indirect effects in longitudinal mediation models* section).

Except for larger MSE values for the smallest sample sizes, increasing sample size did not generally alleviate problems with the estimates, and where bias did not improve, coverage worsened. It was of note that the b path suffered from very poor coverage both when measurement error and confounding were ignored and also to a lesser extent when just confounding was ignored.

Discussion

The aim of this paper was to examine issues affecting the mediator – outcome relationship in a longitudinal mediational analysis of data from a trial of complex treatments for chronic fatigue syndrome. The paper focused specifically on measurement error in the mediator, time lags, unmeasured confounding and the assumption of a constant mediator – outcome relationship over time. We found that it was clearly important to account for measurement error and unmeasured confounding, otherwise parameter estimates and mediated effects may be biased and inferences about estimates flawed. While lagged mediator – outcome paths would be more consistent with a

causal effect, models with contemporaneous mediator – outcome relationships fitted better.

Assuming a constant mediator – outcome relationship over time was plausible and brought a large gain in precision. Our findings here using longitudinal measures supported our earlier finding using a single measure of both mediator and outcome that fear avoidance mediated the effect of treatment on physical functioning ³².

The superiority of the simplex over the autoregressive models and the results of the simulation study clearly showed it was important to account for measurement error in the mediator, perhaps more so than accounting for unmeasured confounding. Models with lagged mediator – outcome relationships followed the classical measurement error paradigm where error dampens effects and so taking account of it increased the magnitude of effects. On the other hand, the contemporaneous mediator – outcome effects were smaller in the simplex models as compared to the autoregressive models. Complex effects of accounting for measurement error in multi-equation models have been noted previously ¹¹. Measurement error was accounted for in this study by using the simplex models, but it is also important to try to do more to address this issue through improved measurement of mediators and outcomes..

Accounting for measurement error led to a small loss in precision. Instrumental variables analysis (IV) is another method for coping with confounding and measurement error in predictor variables. It has proven difficult to apply IV methods to mediation analysis so far, mainly due to the absence of strong instruments, leading to imprecise mediator – outcome estimates ^{4, 14}. In our experience, the use of these repeated measures measurement error models as an alternative to IV has led to much smaller losses in precision.

Lagged mediator – outcome relationships, which would be more consistent with the temporal ordering of a causal process such as mediation, were not supported in the PACE data. This could have been due to the apparent almost simultaneous change in mediator and outcome in these data³¹. However, it could also be because the first measurement of the mediator was taken too late to capture mediator change prior to change in the outcome. The mid-treatment measurements were taken after participants had received approximately seven sessions of therapy, which may be quite late in the process of change. For example, evidence of gains in the first three sessions of brief psychosocial therapy interventions has been demonstrated for depression⁴⁹. Studies looking more carefully at the trajectories of mediator and outcome change by taking earlier and more frequent measures of the variables, perhaps even at every session of therapy, could clarify optimal timing and number of measurements.

The potential for unmeasured confounding of the mediator – outcome relationship was allowed for in models through covariances between mediator and outcome errors. The best fitting models were those with contemporaneous error covariances, suggesting there were unmeasured confounders of the mediator and outcome variables at the same time point that needed to be taken into account. Lagged covariances would have been more consistent with unmeasured confounding in a typical ‘simple’ mediator model with one measure of the mediator taken earlier acting on a single later measure of the outcome. There was less evidence for this sort of unmeasured confounding in the PACE data, although this may not be the case in other situations. Allowing for unmeasured confounding is desirable given the attention this issue has been given in the literature, with the approach described here providing one option. More generally, in practice there is no single best approach and, for example, the approach here could be extended to incorporate existing sensitivity analysis methods²²⁻²⁴ to quantify the level of confounding that would alter these longitudinal model conclusions.

The assumption of equal mediator – outcome relationships over time led to greater precision in these estimates. The idea that no matter how or when the mediator is changed it will have the same effect on the outcome is a potentially strong and theoretically appealing assumption. This assumption aligns well with a description of mediation used in programme theory and intervention evaluation. These fields have described mediation analysis as evaluating both an ‘action theory’ – the a path in Figure 1 where an intervention seeks to change a mediating variable, and a ‘conceptual theory’ – the b path in Figure 1, which is the causal relationship between the mediator and outcome⁵⁰. Describing the b path as the ‘conceptual theory’ fits with thinking of this as a stable relationship existing in nature that can be manipulated by the ‘action theory’ or intervention. This implies that the ‘conceptual theory’ relationship exists in the absence of the intervention and should exist at different points in time. The support of both action and conceptual theories provides evidence for mediation. From a statistical point of view, this assumption led to a large increase in mediator – outcome effect precision, which in turn led to more precise mediated effects. When plausible, making this assumption could be important given the often low power to detect mediated effects⁵¹.

The study’s strengths lay in the use of high quality data stemming from a rigorously conducted trial, as well as the availability of multiple measurements of mediators and outcomes allowing for more complex models. It was only possible to allow for unmeasured confounding in these models because of the availability of multiple measurements. The single measurement each of mediator and outcome generally used to evaluate mediation do not allow for fitting of a model with mediator - outcome covariance as such a model is not identified. It is also more difficult to account for measurement error in these single measure models, although it can be done if the reliability of the measure is known. At least four measurements are needed for identification of all parameters in the simplex models. Clinical trials often only take baseline, post-treatment and follow-up measurements, but the mid-treatment measures taken in PACE made it possible to allow for more paths and to

explore assumptions in models. Also, given the apparent simultaneous early change in mediators and outcomes in PACE ³² it may be fruitful to collect more measurements earlier in the process to clarify mediator and outcome trajectories. Having additional repeated measures of mediator and outcome and/or different measurements of the mediator and outcome at each time point could allow for the exploration of further model assumptions. Furthermore, the methods used here had some particular strengths. Much of the causal mediation literature has focused on the issue of unmeasured confounding, however, both the simulation results in this study and previous mediation model results ¹⁴ suggest that in mental health measurement error may be of even greater concern than unmeasured confounding. The approach taken here addresses both measurement error and the induced unmeasured confounding that correlated measurement error can give rise to simultaneously, while avoiding the need to estimate complex sensitivity parameters. As such, this approach provides for another type of sensitivity analysis. In addition, it is likely much easier to gain information about the reliability of measurement, such as we have done here using repeated measures, than it is in most situations to identify and measure all important confounders. This being said, we do not see the approach taken here and the sensitivity analyses described in the literature ^{15, 16, 22-24} as mutually exclusive. For example, the approaches using SEM to study sensitivity to unmeasured confounding ²²⁻²⁴ could be incorporated into the sorts of measurement models that we have fitted here.

Although there has been some criticism of the use of SEMs to model causal processes, these models have made a large contribution to the modern study of mediation and have been shown here and by others to be a very useful tool for this purpose ⁵². This study showed that modelling of multiple mediator and outcome measurements using SEMs provided flexibility and allowed for exploration of some important assumptions.

There are some limitations of the study. One is that we didn't adjust for other potential baseline confounders of the mediator and outcome besides the baseline measures of mediator and outcome. We included a larger list of measured confounders in the initial PACE mediation analysis and so are confident that the relationship between fear avoidance and physical function is robust to confounding³². We would generally advise the inclusion of measured confounders, which requires consideration at the trial design stage. However, the baseline measures of the mediator and outcome are likely the most important confounders¹⁴ and may also act as proxies for other confounders that haven't been included. The longitudinal models used here included these as well as other measures of the mediator and outcome over time, which are also likely to be important confounders. Given this and the results of the initial mediation analysis³², it seems unlikely the inclusion of other variables would have had a large effect on the mediation estimates. In another analysis (not shown), the variable with the largest effect on the relationship between the fear avoidance mediator and physical functioning outcome was baseline fear avoidance, with baseline work and social adjustment having the second largest effect⁴. The latter variable changed the mediated effect by less than 0.06 units on the standardised scale. Given this, it seems unlikely that including work and social adjustment and other weaker confounders would negate the overall mediation effects found here. It is possible that measurement error covariances could be less important when more measured confounders are included, and it would be important to explore this. A further consideration in the interest of simplicity was the focus of this study on a subset of PACE data also. However, these and other longitudinal models have been used to study the full extent of the PACE trial data and will be described in another paper. Finally, we have assumed that the data are missing at random, an assumption that may not have been met. However, the autoregressive and simplex models likely made this assumption more plausible as earlier measures of the mediator and outcome were predictors of later measures in the models. Future research could evaluate the effect of violation of this assumption on estimates of mediated effects.

The simplex models offer advantages over the autoregressive models, including acknowledging measurement error. However, simplex models make some restrictive assumptions, such as assuming effects between different measurements of the same variable are only from the measure of that variable from the prior time point, as well as an autoregressive correlation structure^{42, 43}. These may or may not be plausible assumptions for the PACE data. In addition, the accumulation of mediated effects over time implied by these models may not be the best fit for the PACE data, where at least on average there was greater change in the mediator and outcome variables up to 12 weeks with a plateau afterwards³¹. These issues may have been reflected in the model fit indices, which suggested that the best model shown here could still be improved upon. We will explore other model types that can allow for different assumptions, such as latent growth models that can allow for a trajectory for each process, and latent change models allowing for potentially independent directions of change between each time point⁴² in another paper.

In conclusion, longitudinal SEMs can account for important sources of potential bias in mediation analysis, such as measurement error in the mediator and unmeasured confounding. Optimal application of these methods requires the availability of repeated measures, necessitating consideration at the study design stage. Assuming that no matter how or when the mediator is changed it has the same effect on the outcome is plausible and gains efficiency. Later phase clinical trials of treatments should aim to address mediation as well as effectiveness hypotheses, keeping in mind these analyses would be enhanced by the inclusion of more and earlier measurements of mediators and outcomes.

Acknowledgements

We acknowledge D. MacKinnon for hosting KG and providing advice on fitting dual process structural equation models for mediation analysis, and the PACE Trial Management Group, which consisted of

the authors of this paper, excluding AP, plus (in alphabetical order): B. Angus, H. Baber, J. Bavinton, M. Burgess, L. V. Clark, D. L. Cox, J. C. DeCesare, A. Johnson, P. McCrone, G. Murphy, M. Murphy, H. O'Dowd, T. Peto, L. Potts, R. Walwyn, and D. Wilks. A. Johnson also read and commented on the paper.

Funding

Funding for the PACE trial was provided by the Medical Research Council, Department for Health for England, The Scottish Chief Scientist Office, and the Department for Work and Pensions. TC, AP and KG were in part supported by the NIHR Biomedical Research Centre for Mental Health at the South London and Maudsley NHS Foundation Trust and Institute of Psychiatry, Kings College London. KG was also funded by an NIHR Doctoral Fellowship, DRF-2011-04-061.

Conflict of Interest Statement

PDW has done voluntary and paid consultancy work for the United Kingdom government and a reinsurance company. TC has received royalties from Sheldon Press and Constable and Robinson. MS has done voluntary and paid consultancy work for the United Kingdom government, has done consultancy work for an insurance company, and has received royalties from Oxford University Press. AP has received royalties from Western Psychological Services, Oxford University Press, Imperial College Press, and Chapman and Hall. KG has no conflicts of interest to declare.

Disclaimer

This report is independent research partly arising from a Doctoral Research Fellowship supported by the National Institute for Health Research. The views expressed in this publication are those of the author(s) and not necessarily those of the NHS, the National Institute for Health Research or the Department of Health.

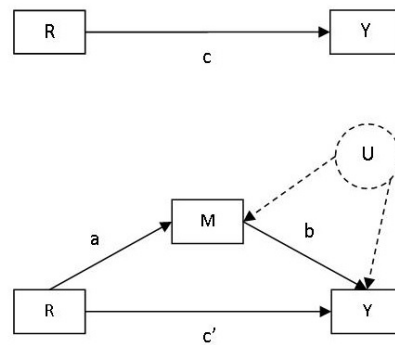
References

1. Baron RM and Kenny DA. The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *J Pers Soc Psychol.* 1986; 51: 1173-82.
2. Emsley R, Dunn G and White IR. Mediation and moderation of treatment effects in randomised controlled trials of complex interventions. *Stat Methods Med Res.* 2010; 19: 237-70.
3. MacKinnon DP. *Introduction to Statistical Mediation Analysis.* Taylor & Francis Group LLC, 2008, p.477.
4. Goldsmith KA. Instrumental variable and longitudinal structural equation modelling methods for causal mediation: the PACE trial of treatments for chronic fatigue syndrome *Biostatistics Department, Institute of Psychiatry, Psychology & Neuroscience.* London, UK: King's College London, 2014.
5. MacKinnon DP and Dwyer JH. Estimating mediated effects in prevention studies. *Evaluation Review.* 1993; 17: 144-58.
6. Shrout PE and Bolger N. Mediation in experimental and nonexperimental studies: New procedures and recommendations. *Psychological methods.* 2002; 7: 422-45.
7. MacKinnon DP and Fairchild AJ. Current directions in mediation analysis. *Current directions in psychological science : a journal of the American Psychological Society.* 2009; 18: 5.
8. MacKinnon DP. Mediating variable. In: Smelser NJ and Baltes PB, (eds.). *International Encyclopedia of the Social and Behavioural Sciences.* 2001, p. 9503-7.
9. Wright S. The relative importance of heredity and environment in determining the piebald pattern of guinea pigs. *Proceedings of the National Academy of Sciences of the United States of America.* 1920; 6: 320-32.
10. Wright S. Correlation and causation Part I. Method of path coefficients. *J Agric Res.* 1920; 20: 0557-85.
11. Bollen KA. *Structural Equations with Latent Variables.* New York, NY: John Wiley & Sons, 1989.
12. Klein RB. *Principles and Practice of Structural Equation Modelling.* New York, NY: The Guildford Press, 2011, p.425.
13. Fleiss JL and Shrout PE. Effects of Measurement Errors on Some Multivariate Procedures. *Am J Public Health.* 1977; 67: 1188-91.
14. Pickles A, Harris V, Green J, et al. Treatment mechanism in the MRC preschool autism communication trial: implications for study design and parent-focussed therapy for children. *J Child Psychol Psychiatry.* 2015; 56: 162-70.
15. le Cessie S, Debeij J, Rosendaal FR, Cannegieter SC and Vandenbroucke JP. Quantification of bias in direct effects estimates due to different types of measurement error in the mediator. *Epidemiology.* 2012; 23: 551-60.
16. Valeri L, Xihong L and Vanderweele TJ. Mediation analysis when a continuous mediator is measured with error and the outcome follows a generalized linear model. *Statistics in medicine.* 2014; 33: 4875-90.
17. Muthen B and Asparouhov T. Causal effects in mediation modelling: An introduction with applications to latent variables. *Struct Equ Modeling.* 2014.
18. Judd CM and Kenny DA. Process analysis - estimating mediation in treatment evaluations. *Evaluation Review.* 1981; 5: 602-19.
19. James LR. The unmeasured variables problem in path analysis. *J Appl Psychol.* 1980; 65: 415-21.
20. James LR and Brett JM. Mediators, moderators, and tests for mediation. *J Appl Psychol.* 1984; 69: 307-21.
21. Cole DA and Maxwell SE. Testing mediational models with longitudinal data: Questions and tips in the use of structural equation modeling. *J Abnorm Psychol.* 2003; 112: 558-77.

22. De Stavola BL, Daniel RM, Ploubidis GB and Micali N. Mediation analysis with intermediate confounding: structural equation modeling viewed through the causal inference lens. *American journal of epidemiology*. 2015; 181: 64-80.
23. Imai K, Keele L and Tingley D. A general approach to causal mediation analysis. *Psychological methods*. 2010; 15: 309-34.
24. Imai K, Keele L and Yamamoto T. Identification, inference and sensitivity analysis for causal mediation effects. *Statistical Science*. 2010; 25: 51-71.
25. Robins JM and Greenland S. Identifiability and exchangeability for direct and indirect effects. *Epidemiology*. 1992; 3: 143-55.
26. VanderWeele TJ. Marginal structural models for the estimation of direct and indirect effects. *Epidemiology*. 2009; 20: 18-26.
27. VanderWeele TJ. Bias Formulas for Sensitivity Analysis for Direct and Indirect Effects. *Epidemiology*. 2010; 21: 540-51.
28. VanderWeele TJ and Vansteelandt S. Conceptual issues concerning mediation, interventions and composition. *Statistics and Its Interface*. 2009; 2: 457-68.
29. VanderWeele TJ, Vansteelandt S and Robins JM. Effect Decomposition in the Presence of an Exposure-Induced Mediator-Outcome Confounder. *Epidemiology*. 2014; 25: 300-6.
30. Vanderweele TJ. *Explanation in Causal Inference: Mediation and Interaction*. Oxford, UK: Oxford University Press, 2015.
31. White PD, Goldsmith KA, Johnson AL, et al. Comparison of adaptive pacing therapy, cognitive behaviour therapy, graded exercise therapy, and specialist medical care for chronic fatigue syndrome (PACE): a randomised trial. *Lancet*. 2011; 377: 823-36.
32. Chalder T, Goldsmith KA, White PD, Sharpe M and Pickles A. Rehabilitative therapies for chronic fatigue syndrome: a secondary mediation analysis of the PACE trial. *Lancet Psychiatry*. 2015; 2: 141-52.
33. Knudsen AK, Henderson M, Harvey SB and Chalder T. Long-term sickness absence among patients with chronic fatigue syndrome. *Br J Psychiatry*. 2011; 199: 430-1.
34. Skerrett TN and Moss-Morris R. Fatigue and social impairment in multiple sclerosis: the role of patients' cognitive and behavioral responses to their symptoms. *Journal of psychosomatic research*. 2006; 61: 587-93.
35. Buchwald D, Pearlman T, Umali J, Schmalting K and Katon W. Functional status in patients with chronic fatigue syndrome, other fatiguing illnesses, and healthy individuals. *Am J Med*. 1996; 101: 364-70.
36. McHorney CA, Ware JE and Raczek AE. The Mos 36-Item Short-Form health survey (Sf-36) 2. Psychometric and clinical tests of validity in measuring physical and mental health constructs. *Med Care*. 1993; 31: 247-63.
37. Enders CK and Bandalos DL. The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Struct Equ Modeling*. 2001; 8: 430-57.
38. Okleshen Peters CL and Enders CK. A primer for the estimation of structural equation models in the presence of missing data: Maximum likelihood algorithms. *Journal of Targeting, Measurement and Analysis for Marketing*. 2002; 11: 81-95.
39. Steiger JH. Structural model evaluation and modification - an interval estimation approach. *Multivar Behav Res*. 1990; 25: 173-80.
40. Steiger JH and Lind JC. Statistically-based tests for the number of common factors. *Annual Spring Meeting of the Psychometric Society*. Iowa City, IA1980.
41. Schwarz G. Estimating dimension of a model. *Ann Stat*. 1978; 6: 461-4.
42. Dunn G, Everitt B and Pickles A. *Modelling Covariances and Latent Variables Using EQS*. London, UK: Chapman & Hall, 1993, p.201.
43. Joreskog KG. Estimation and testing of simplex models. *Brit J Math Stat Psy*. 1970; 23: 121-45.

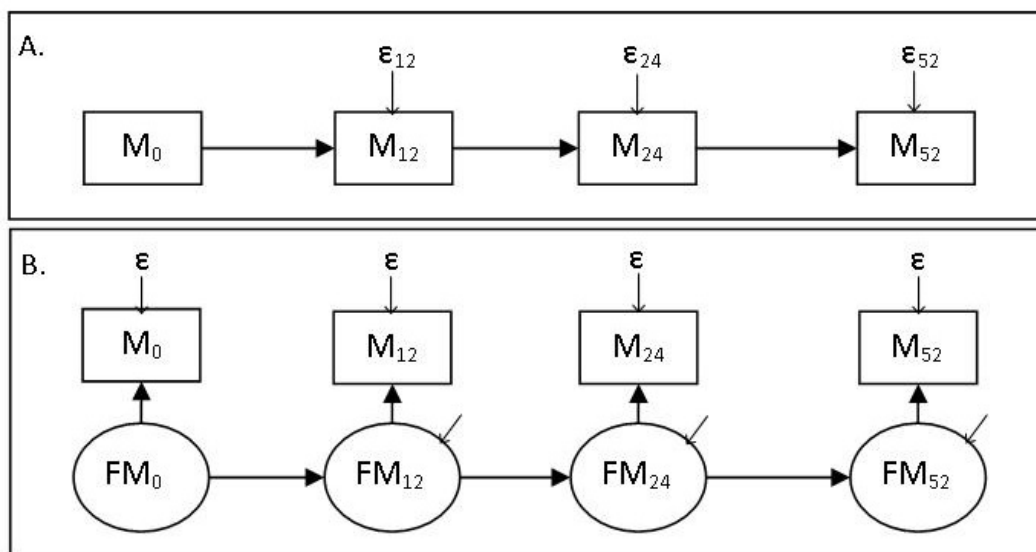
44. Marsh HW. Stability of individual-differences in multiwave panel studies - comparison of simplex models and one-factor models. *J Educ Meas.* 1993; 30: 157-83.
45. Daniel RM, De Stavola BL, Cousens SN and Vansteelandt S. Causal mediation analysis with multiple mediators. *Biometrics.* 2015; 71: 1-14.
46. Efron B and Tibshirani R. *An Introduction to the Bootstrap.* Boca Raton, FL: Chapman and Hall/CRC, 1993.
47. MacKinnon DP, Lockwood CM and Williams J. Confidence limits for the indirect effect: distribution of the product and resampling methods. *Multivariate Behav Res.* 2004; 39: 99.
48. Burton A, Altman DG, Royston P and Holder RL. The design of simulation studies in medical statistics. *Statistics in medicine.* 2006; 25: 4279-92.
49. Delgadillo J, McMillan D, Lucock M, Leach C, Ali S and Gilbody S. Early changes, attrition, and dose-response in low intensity psychological interventions. *Brit J Clin Psychol.* 2014; 53: 114-30.
50. Chen H. Action theory and conceptual theory: summatively diagnosing the intervention program. *Theory-Driven Evaluations.* Newbury Park, CA, USA: Sage Publications, Inc., 1990.
51. Fritz MS and MacKinnon DP. Required sample size to detect the mediated effect. *Psychol Sci.* 2007; 18: 233-9.
52. Bollen KA and Pearl J. Eight myths about causality and structural equation models. In: Morgan SL, (ed.). *Handbook of Causal Analysis for Social Research.* Dordrecht, Netherlands: Springer, 2013, p. 301-28.

Figure 1. Treatment effect and mediation path diagrams



R = randomised treatment, Y = outcome, M = mediator, U = unmeasured confounders

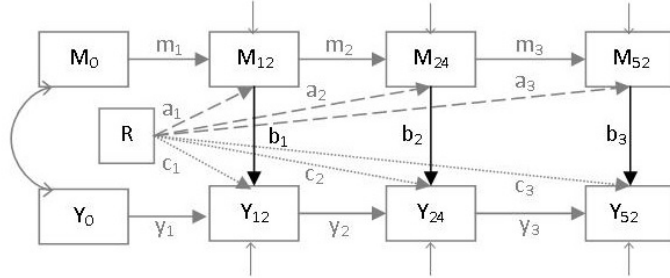
Figure 2. Autoregressive and simplex model examples



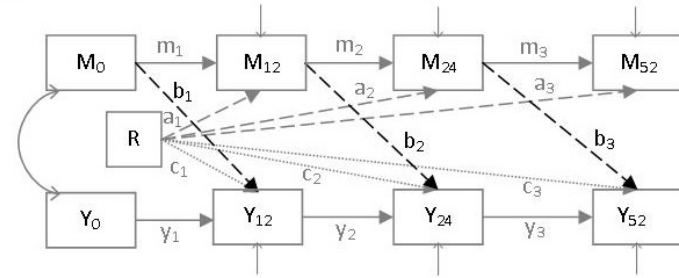
M_0 = mediator at baseline, M_{12} = mediator at 12 weeks, M_{24} = mediator at 24 weeks, M_{52} = mediator at 52 weeks, FM_0 = latent true mediator score at baseline, FM_{12} = latent true mediator score at 12 weeks, FM_{24} = latent true mediator score at 24 weeks, FM_{52} = latent true mediator score at 52 weeks. One method for obtaining identification of the model in B is shown in the figure: all $\text{var}(\epsilon)$ are set equal.

Figure 3. Autoregressive models with contemporaneous and lagged b paths and simplex models with contemporaneous and lagged covariances using PACE example time points

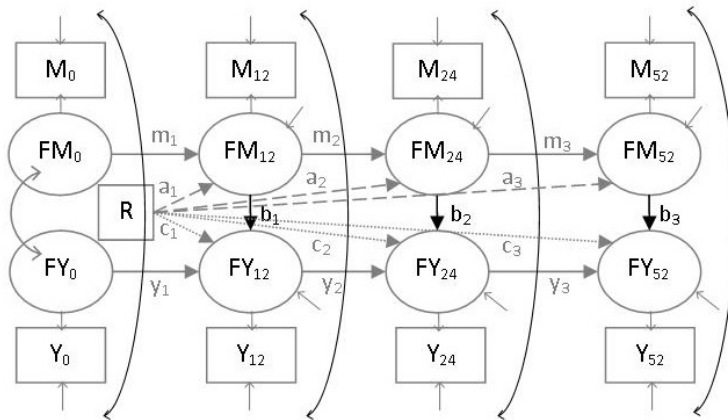
A. Contemporaneous b paths



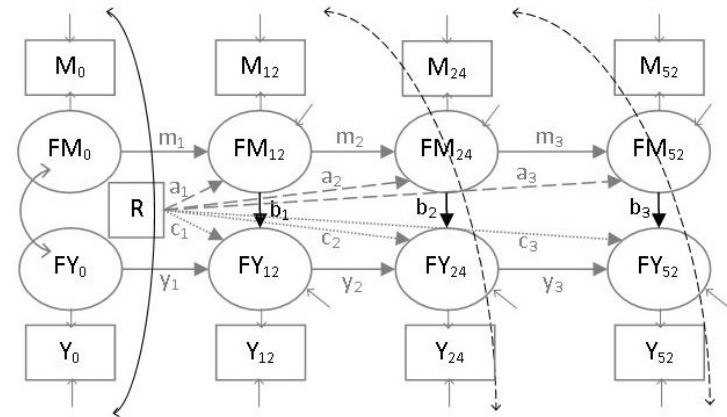
B. Lagged b paths



C. Contemporaneous covariances



D. Lagged covariances



$M_0, M_{12}, M_{24}, M_{52}$ – mediator measurements taken at baseline, 12 weeks, 24 weeks and 52 weeks post-randomisation. $Y_0, Y_{12}, Y_{24}, Y_{52}$ – outcome measurements taken at the same time points.

Figure 4. Equations describing dual-process simplex model shown in Figure 3C

$$M_0 = FM_0 + \varepsilon$$

$$M_{12} = FM_{12} + \varepsilon$$

$$M_{24} = FM_{24} + \varepsilon$$

$$M_{52} = FM_{52} + \varepsilon$$

$$Y_0 = FY_0 + \varepsilon$$

$$Y_{12} = FY_{12} + \varepsilon$$

$$Y_{24} = FY_{24} + \varepsilon$$

$$Y_{52} = FY_{52} + \varepsilon$$

$$FM_{12} = R + FM_0 + \varepsilon$$

$$FM_{24} = R + FM_{12} + \varepsilon$$

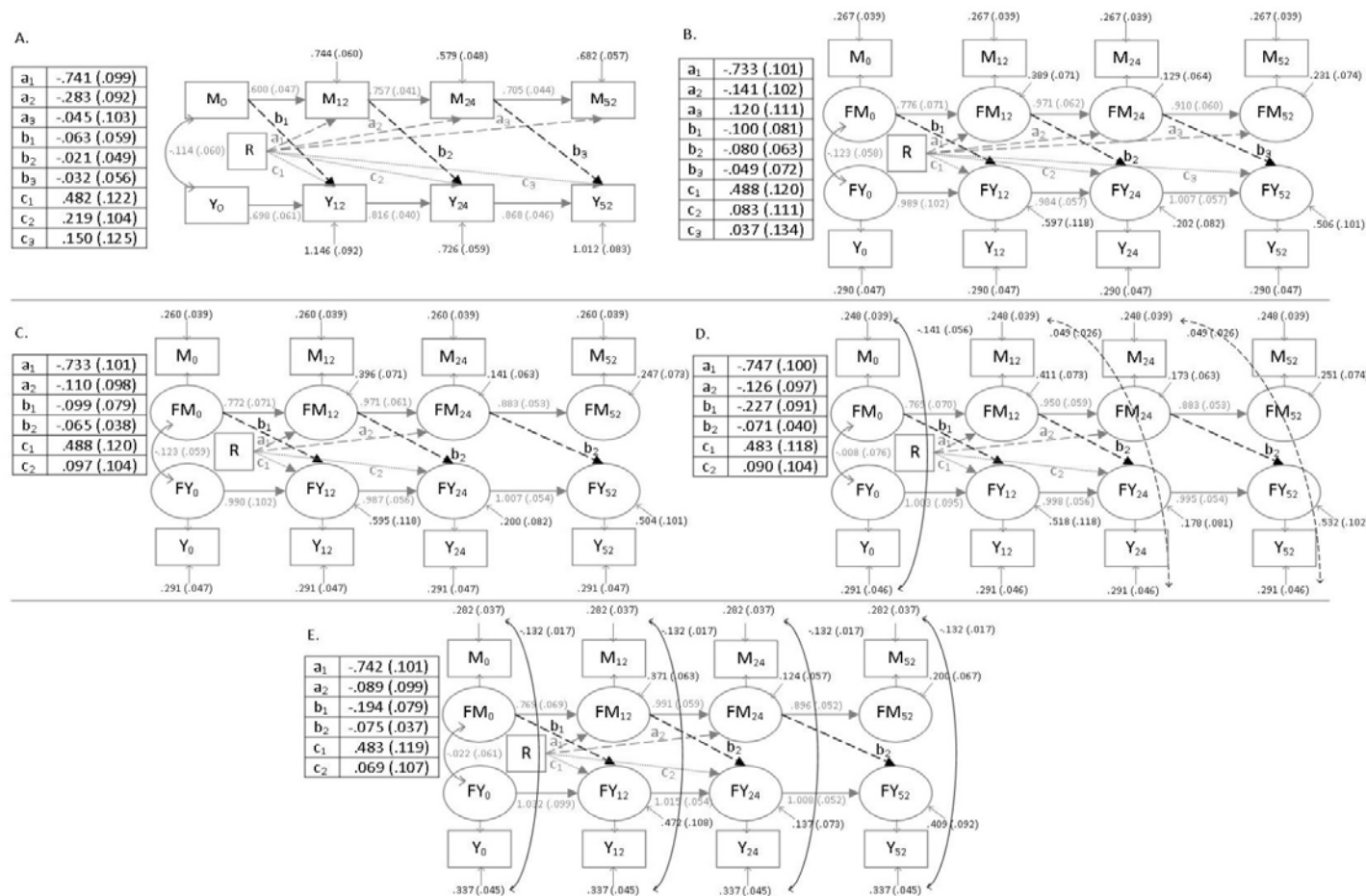
$$FM_{52} = R + FM_{24} + \varepsilon$$

$$FY_{12} = R + FY_0 + FM_{12} + \varepsilon$$

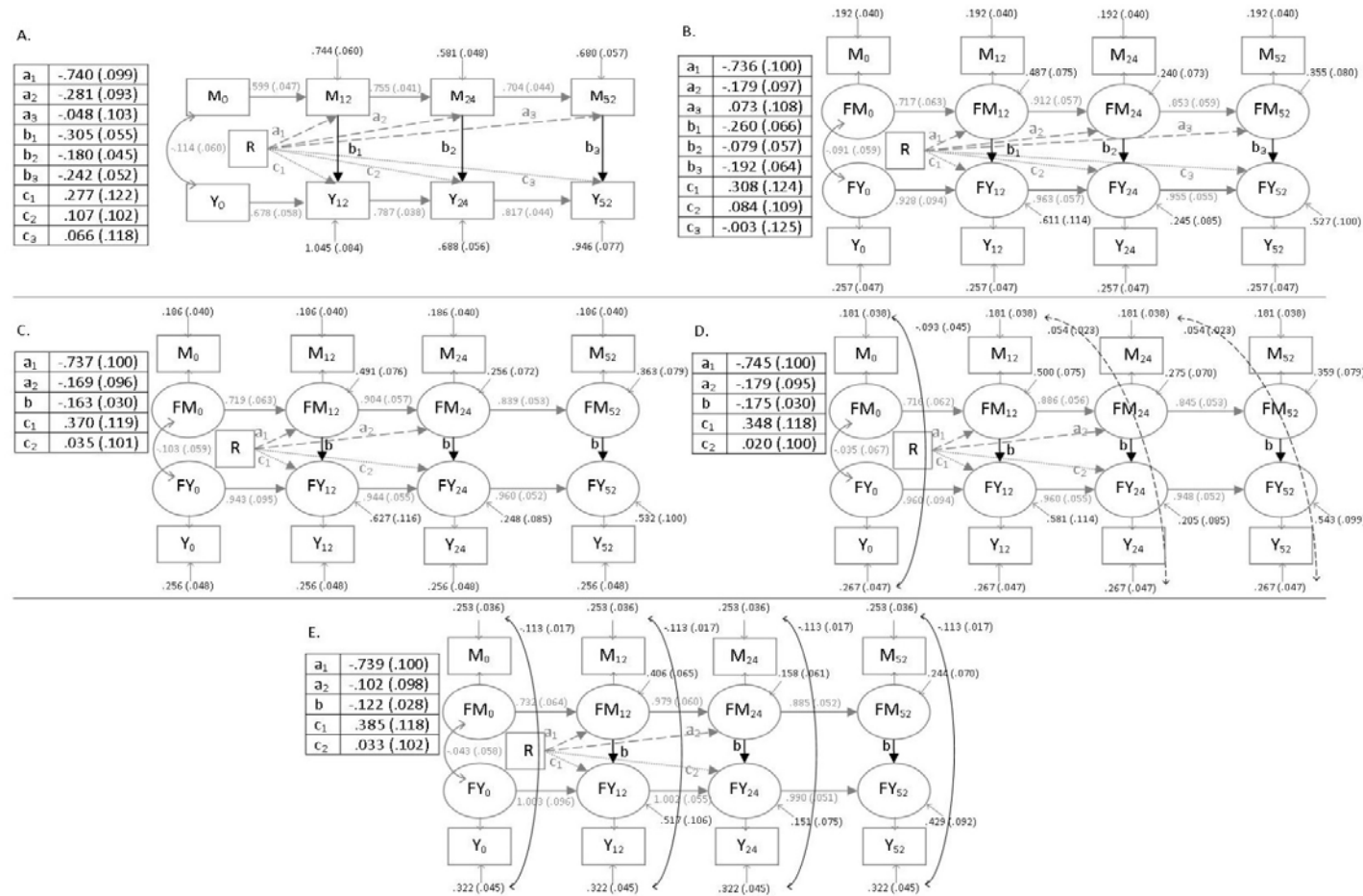
$$FY_{24} = R + FY_{12} + FM_{24} + \varepsilon$$

$$FY_{52} = R + FY_{24} + FM_{52} + \varepsilon$$

Models were fitted with mediator (M) error variances set equal (ε_M) and outcome (Y) error variances set equal (ε_Y)

Figure 5. Longitudinal mediator – outcome process models with lagged b paths

Numbers in tables are estimate (standard error). R = randomised treatment group, M_0 , M_{12} , M_{24} , M_{52} – mediator measurements taken at baseline, 12 weeks, 24 weeks and 52 weeks post-randomisation, Y_0 , Y_{12} , Y_{24} , Y_{52} – outcome measurements taken at the same time points, FM_0 , FM_{12} , FM_{24} , FM_{52} – true mediator scores at baseline, 12 weeks, 24 weeks, 52 weeks post-randomisation, FY_0 , FY_{12} , FY_{24} , FY_{52} – true outcome scores at the same time points. Paths between observed and latent variables for both the mediator and the outcome process all have factor loadings of 1.

Figure 6. Longitudinal mediator – outcome process models with contemporaneous *b* paths

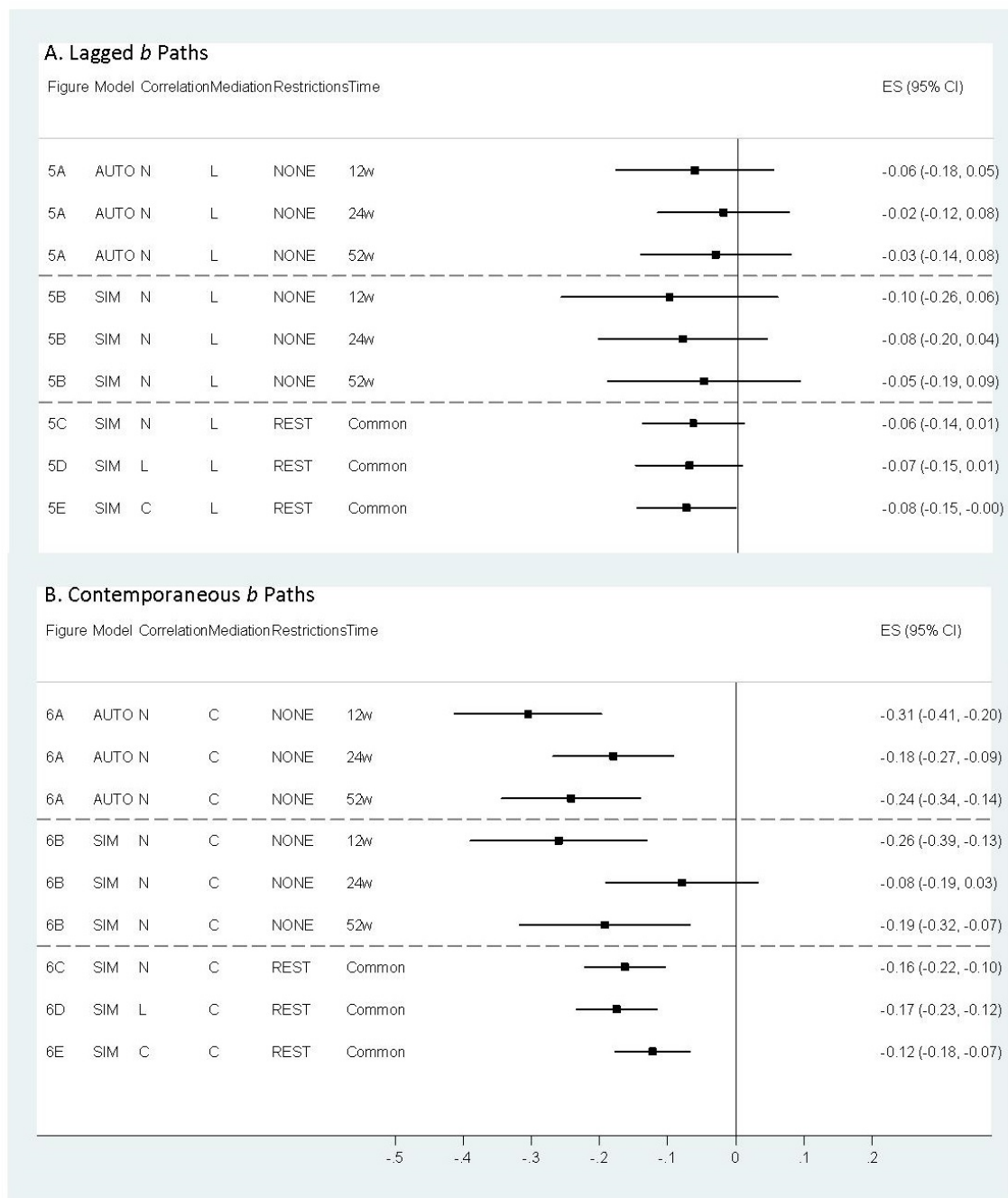
Numbers in tables are estimate (standard error). R = randomised treatment group, M_0 , M_{12} , M_{24} , M_{52} – mediator measurements taken at baseline, 12 weeks, 24 weeks and 52 weeks post-randomisation, Y_0 , Y_{12} , Y_{24} , Y_{52} – outcome measurements taken at the same time points, FM_0 , FM_{12} , FM_{24} , FM_{52} – true mediator scores at baseline, 12 weeks, 24 weeks, 52 weeks post-randomisation, FY_0 , FY_{12} , FY_{24} , FY_{52} – true outcome scores at the same time points. Paths between observed and latent variables for both the mediator and the outcome process all have factor loadings of 1.

Table 1. Comparison of fit statistics across autoregressive and simplex models

General type	Figure #	Corr'n type	Med'n type	Model χ^2	RMSEA (90% CI)	AIC	BIC
Models without parameter restrictions ^a							
Autoregressive	5A	N	L	187.943 df = 20 p < 0.001	0.162 (0.141, 0.184) p < 0.001	6633.276	6753.863
Simplex	5B	N	L	115.696 df = 18 p < 0.001	0.130 (0.108, 0.153) p < 0.001	6565.030	6693.153
Autoregressive	6A	N	C	123.004 df = 20 p < 0.001	0.127 (0.106, 0.149) p < 0.001	6568.337	6688.924
Simplex	6B	N	C	87.706 df = 18 p < 0.001	0.110 (0.088, 0.134) p < 0.001	6537.039	6665.162
Models with parameter restrictions ^a							
Simplex	5C	N	L	117.023 df = 21 p < 0.001	0.120 (0.099, 0.141) p < 0.001	6560.356	6677.174
Simplex	6C	N	C	91.916 df = 22 p < 0.001	0.100 (0.079, 0.121) p < 0.001	6533.249	6646.299
Simplex	5D	L	L	106.322 df = 19 p < 0.001	0.120 (0.098, 0.143) p < 0.001	6553.655	6678.010
Simplex	5E	C	L	47.051 df = 20 p = 0.001	0.065 (0.041, 0.089) p = 0.14	6492.384	6612.970
Simplex	6D	L	C	82.252 df = 20 p < 0.001	0.099 (0.077, 0.121) p < 0.001	6527.585	6648.171
Simplex	6E	C	C	41.383 df = 21 p = 0.005	0.055 (0.030, 0.080) p = 0.34	6484.717	6601.534

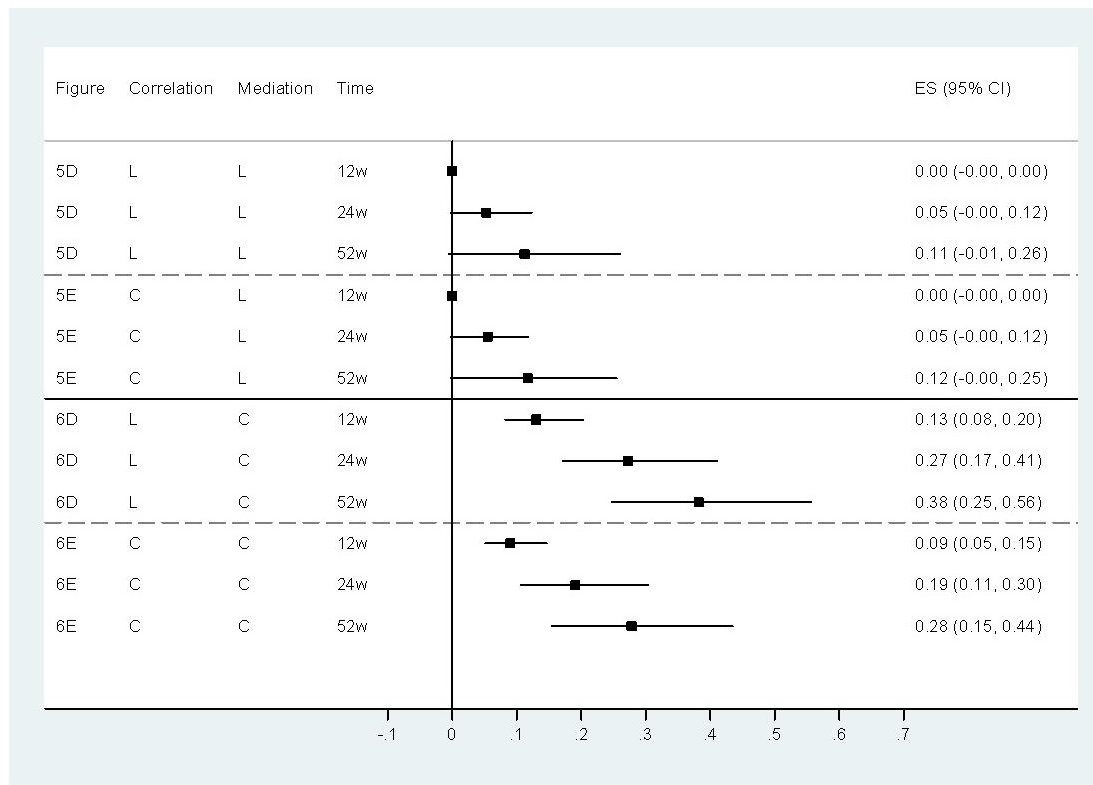
χ^2 = chi-square model fit statistic, RMSEA = root mean square error of approximation, CI = confidence interval, AIC = Akaike's Information Criterion, BIC = Bayesian Information Criterion, N = no post-baseline covariances, C = contemporaneous, L = lagged, df = degrees of freedom

^aParameter restrictions are equal error variances for each mediator measure and for each outcome measure, treatment to mediator at 52 weeks and treatment to outcome at 52 weeks = 0 and mediator to outcome *b* paths equal over time

Figure 7. *b* paths (95% CI) for autoregressive and simplex models

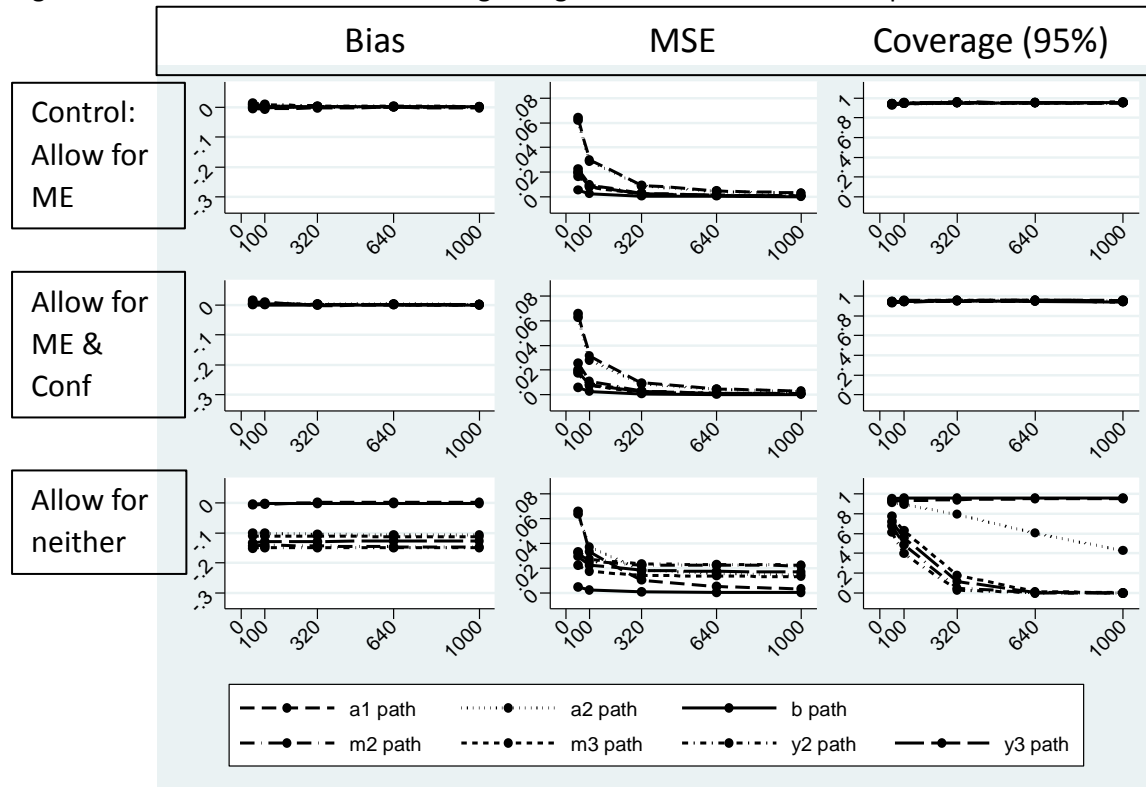
AUTO = autoregressive model, SIM = simplex model, N = no post-baseline covariances, L = lagged, C = contemporaneous, NONE = covariances free, treatment to 52 week paths present, mediation *b* paths not set equal, REST = covariances set equal, treatment to 52 week paths = 0, mediation *b* paths set equal, w = weeks post-randomisation, ES = estimate, CI = confidence interval

Figure 8. Overall indirect/mediated effects (95% bias-corrected bootstrap CI) for simplex models



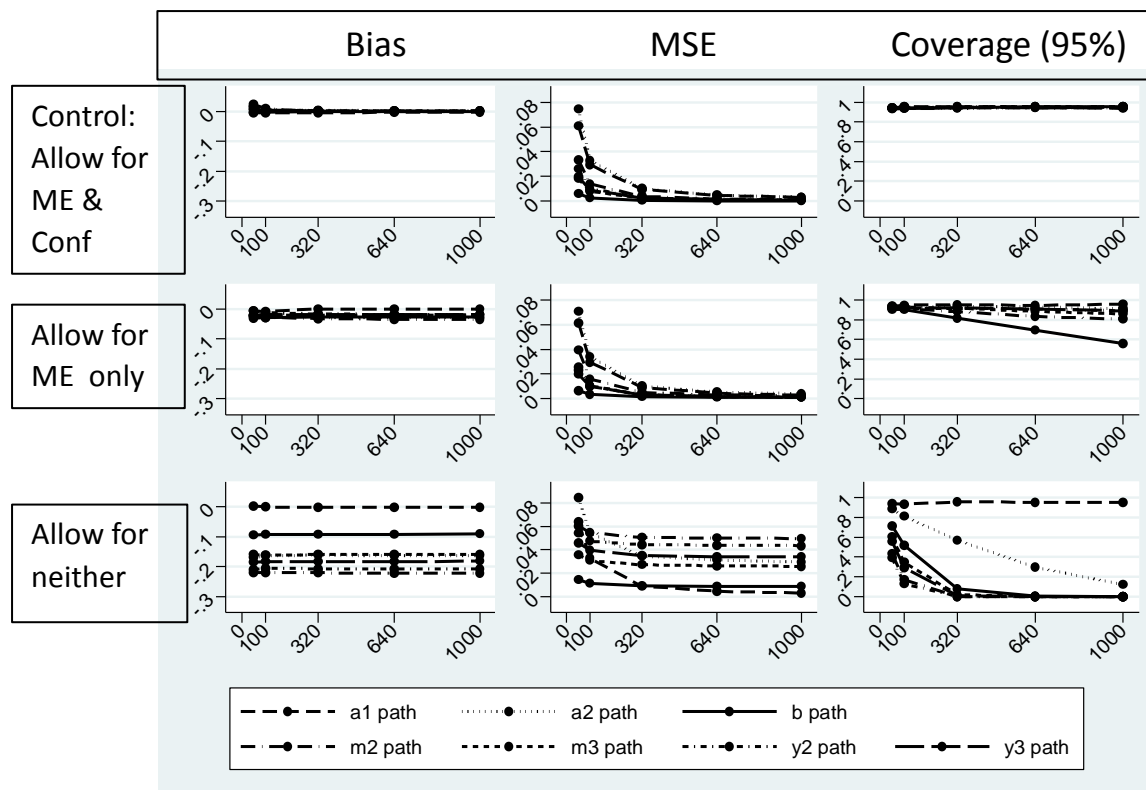
L = lagged, C = contemporaneous, w = weeks post-randomisation, ES = estimate, CI = 95% bias-corrected bootstrap confidence interval

Figure 9. Simulation results – effect of ignoring measurement error when present



All data generated using simplex models and 1500 repetitions for each sample size, allowing for measurement error (ME), but not for measurement error covariances representing unmeasured confounding (Conf). MSE = mean square error, Control = generated and analysed allowing for ME only when only ME present, Allow for ME & Conf = analysed allowing for both ME and Conf when only ME present, Allow for neither = analysed without allowing for ME or Conf when only ME present. .

Figure 10. Simulation results – effect of ignoring measurement error and confounding when present



All data generated using simplex models and 1500 repetitions for each sample size allowing for measurement error (ME) and for measurement error covariances representing unmeasured confounding (Conf). MSE = mean square error, Control = generated and analysed with ME and Conf when both ME & Conf present, Allow for ME only = analyse allowing ME only when both ME & Conf present, Allow for neither = analysed without allowing for ME or Conf when both ME & Conf present.

Supplementary Material

Table A. Comparison of *b* path estimates (SE) across autoregressive and simplex models

Figure #	Corr'n type	Med'n type	BL -> 12 weeks	<i>b</i> path 12 -> 24 weeks	<i>b</i> path 24 -> 52 weeks	<i>b</i> path 12 -> 12 weeks	<i>b</i> path 24 -> 24 weeks	<i>b</i> path 52 -> 52 weeks
Autoregressive								
Figure 5A	N	L	-0.063 (0.059)	-0.021 (0.049)	-0.032 (0.056)	---	---	---
Simplex								
Figure 5B	N	L	-0.100 (0.081)	-0.080 (0.063)	-0.049 (0.072)	---	---	---
Autoregressive								
Figure 6A	N	C	---	---	---	-0.305 ^a (0.055)	-0.180 ^a (0.045)	-0.242 ^a (0.052)
Simplex								
Figure 6B	N	C	---	---	---	-0.260 ^a (0.066)	-0.079 (0.057)	-0.192 ^a (0.064)
Simplex with restrictions								
Figure 5C	N	L	-0.099 (0.079)	-0.065 (0.038)	-0.065 (0.038)	---	---	---
Figure 6C	N	C	---	---	---	-0.163 ^a (0.030)	-0.163 ^a (0.030)	-0.163 ^a (0.030)
Figure 5D	L	L	-0.227 ^a (0.091)	-0.071 (0.040)	-0.071 (0.040)	---	---	---
Figure 5E	C	L	-0.194 ^a (0.079)	-0.075 ^a (0.037)	-0.075 ^a (0.037)	---	---	---
Figure 6D	L	C	---	---	---	-0.175 ^a (0.030)	-0.175 ^a (0.030)	-0.175 ^a (0.030)
Figure 6E	C	C	---	---	---	-0.122 ^a (0.028)	-0.122 ^a (0.028)	-0.122 ^a (0.028)

SE = standard error, BL = baseline, N = no post-baseline covariances, L = lagged, C = contemporaneous

^aestimate significant at $p \leq 0.05$.

Table B. Time-specific indirect, direct and total effects at 12 weeks (95% bias-corrected bootstrap CI) in fear avoidance/physical functioning simplex models with restrictions

	Figure 5D	Figure 5E	Figure 6D	Figure 6E
Covariance type	L	C	L	C
Mediation type	L	L	C	C
Indirect effects:				
trt -> fm12 -> fy12			0.130 (0.082, 0.203)	0.090 (0.051, 0.146)
Total 12 week indirect effect			0.130 (0.082, 0.203)	0.090 (0.051, 0.146)
Direct effects:				
trt -> fy12	0.483 (0.229, 0.749)	0.483 (0.231, 0.745)	0.348 (0.117, 0.607)	0.385 (0.152, 0.628)
Total 12 week direct effect	0.483 (0.229, 0.749)	0.483 (0.231, 0.745)	0.348 (0.117, 0.607)	0.385 (0.152, 0.628)
Total 12 week effect	0.483 (0.229, 0.749)	0.483 (0.231, 0.745)	0.478 (0.219, 0.735)	0.475 (0.223, 0.725)

L = lagged, C = contemporaneous, fm12 = latent true mediator score at 12 weeks, fy12 = latent true outcome score at 12 weeks

Table C. Time-specific indirect, direct and total effects at 24 weeks (95% bias-corrected bootstrap CI) in fear avoidance/physical functioning simplex models with restrictions

	Figure 5D	Figure 5E	Figure 6D	Figure 6E
Covariance type	L	C	L	C
Mediation type	L	L	C	C
Indirect effects:				
trt -> fm12 -> fy24	0.053 (-0.002, 0.123)	0.055 (-0.002, 0.118)		
trt -> fm12 -> fy12 -> fy24			0.125 (0.080, 0.184)	0.090 (0.050, 0.142)
trt -> fm12 -> fm24 -> fy24			0.115 (0.073, 0.177)	0.088 (0.047, 0.141)
trt -> fm24 -> fy24			0.031 (-0.001, 0.077)	0.012 (-0.009, 0.045)
Total 24 week indirect effect	0.053 (-0.002, 0.123)	0.055 (-0.002, 0.118)	0.272 (0.171, 0.410)	0.191 (0.106, 0.304)
Direct effects:				
trt -> fy12 -> fy24	0.482 (0.232, 0.755)	0.490 (0.236, 0.772)	0.334 (0.110, 0.604)	0.385 (0.159, 0.664)
trt -> fy24	0.090 (-0.119, 0.282)	0.069 (-0.150, 0.270)	0.020 (-0.186, 0.213)	0.033 (-0.179, 0.229)
Total 24 week direct effect	0.572 (0.308, 0.831)	0.560 (0.298, 0.828)	0.354 (0.100, 0.612)	0.419 (0.170, 0.663)
Total 24 week effect	0.625 (0.385, 0.882)	0.615 (0.364, 0.867)	0.625 (0.385, 0.880)	0.610 (0.365, 0.862)

L = lagged, C = contemporaneous, fm12 = latent true mediator score at 12 weeks, fm24 = latent true mediator score at 24 weeks, fy12 = latent true outcome score at 12 weeks, fy24 = latent true outcome score at 24 weeks

Table D. Time-specific indirect, direct and total effects at 52 weeks (95% bias-corrected bootstrap CI) in fear avoidance/physical functioning simplex models with restrictions

	Figure 5D	Figure 5E	Figure 6D	Figure 6E
Covariance type	L	C	L	C
Mediation type	L	L	C	C
Indirect effects:				
trt -> fm12 -> fy24 -> fy52	0.053 (-0.002, 0.121)	0.056 (-0.001, 0.117)		
trt -> fm12 -> fm24 -> fy52	0.050 (-0.002, 0.127)	0.055 (-0.0002, 0.128)		
trt -> fm24 -> fy52	0.009 (-0.002, 0.036)	0.007 (-0.005, 0.034)		
trt -> fm12 -> fy12-> fy24 -> fy52			0.119 (0.079, 0.171)	0.089 (0.051, 0.137)
trt -> fm12 -> fm24-> fy24 -> fy52			0.109 (0.071, 0.157)	0.087 (0.048, 0.136)
trt -> fm12 -> fm24-> fm52 -> fy52			0.098 (0.062, 0.149)	0.078 (0.042, 0.127)
trt -> fm24-> fy24 -> fy52			0.030 (-0.0004, 0.069)	0.012 (-0.009, 0.044)
trt -> fm24-> fm52 -> fy52			0.026 (-0.001, 0.063)	0.011 (-0.009, 0.038)
Total 52 week indirect effect	0.112 (-0.006, 0.260)	0.117 (-0.002, 0.255)	0.382 (0.247, 0.556)	0.278 (0.154, 0.435)
Direct effects:				
trt -> fy12 -> fy24 -> fy52	0.479 (0.236, 0.780)	0.494 (0.239, 0.794)	0.317 (0.109, 0.582)	0.382 (0.160, 0.666)
trt -> fy24 -> fy52	0.090 (-0.118, 0.285)	0.070 (-0.146, 0.276)	0.019 (-0.175, 0.201)	0.033 (-0.171, 0.232)
Total 52 week direct effect	0.569 (0.306, 0.841)	0.564 (0.305, 0.840)	0.335 (0.092, 0.594)	0.415 (0.160, 0.655)

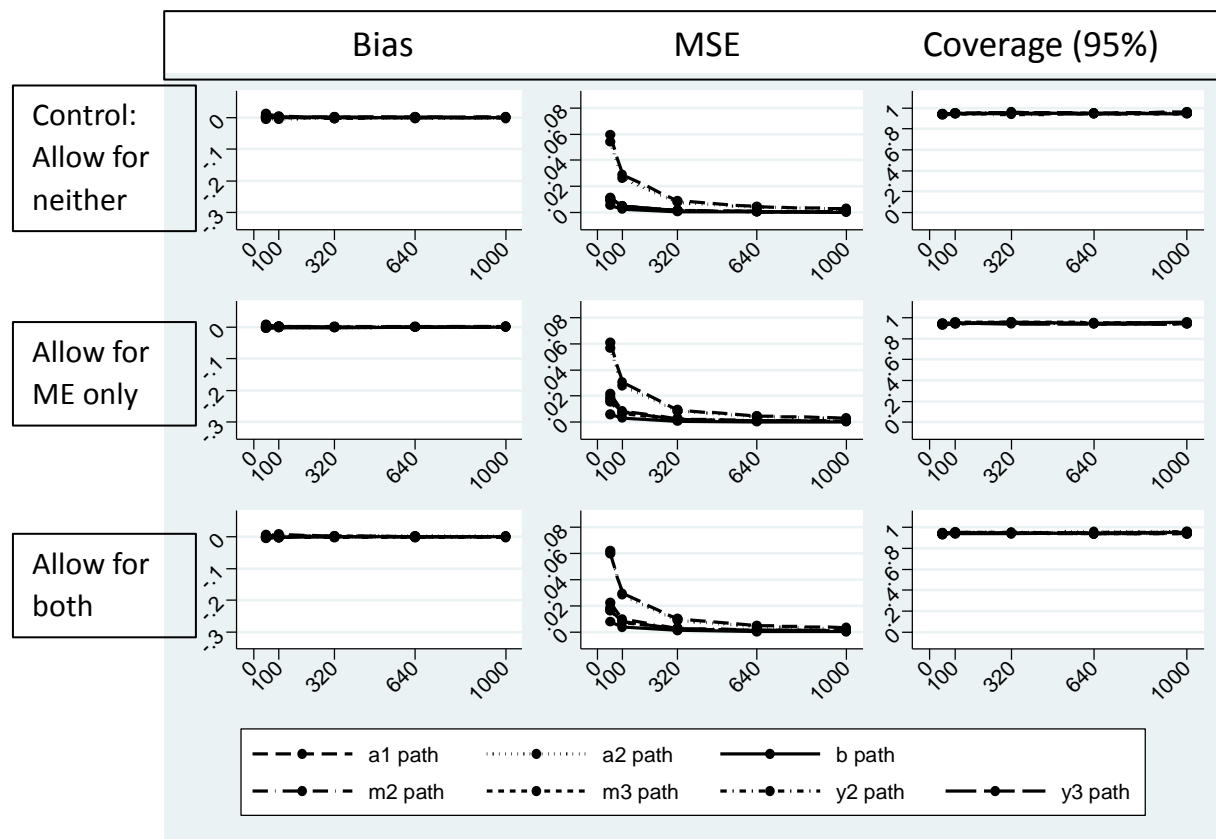
	Figure 5D	Figure 5E	Figure 6D	Figure 6E
Covariance type	L	C	L	C
Mediation type	L	L	C	C
Total 52 week effect	0.681 (0.432, 0.942)	0.682 (0.427, 0.944)	0.717 (0.490, 0.998)	0.692 (0.441, 0.953)

L = lagged, C = contemporaneous, fm12 = latent true mediator score at 12 weeks, fm24 = latent true mediator score at 24 weeks, fm52 = latent true mediator score at 52 weeks, fy12 = latent true outcome score at 12 weeks, fy24 = latent true outcome score at 24 weeks, fy52 = latent true outcome score at 52 weeks

Table E. Simulation convergence results Generating model		Analysis model		Sample size	Convergence	Covariance matrix issues
Measurement error	Unmeasured confounding	Measurement error	Unmeasured confounding			
N	N	N	N	50	1500/1500	0
				100	1500/1500	0
				320	1500/1500	0
				640	1500/1500	0
				1000	1500/1500	0
N	N	Y	N	50	1500/1500	Most
				100	1500/1500	Most
				320	1500/1500	Most
				640	1500/1500	Most
				1000	1500/1500	Most
N	N	Y	Y	50	1499/1500	Most
				100	1500/1500	Most
				320	1500/1500	Most
				640	1500/1500	Most
				1000	1500/1500	Most
Y	N	N	N	50	1500/1500	0
				100	1500/1500	0
				320	1500/1500	0
				640	1500/1500	0
				1000	1500/1500	0
Y	N	Y	N	50	1499/1500	Most
				100	1500/1500	Many
				320	1500/1500	1 (0.1%)
				640	1500/1500	0
				1000	1500/1500	0
Y	N	Y	Y	50	1499/1500	Most

Table E. Simulation convergence results Generating model		Analysis model		Sample size	Convergence	Covariance matrix issues
Measurement error	Unmeasured confounding	Measurement error	Unmeasured confounding			
				100	1500/1500	Many
				320	1500/1500	0
				640	1500/1500	0
				1000	1500/1500	0
Y	Y	N	N	50	1500/1500	0
				100	1500/1500	0
				320	1500/1500	0
				640	1500/1500	0
				1000	1500/1500	0
Y	Y	Y	N	50	1496/1500	Most
				100	1500/1500	Many
				320	1500/1500	18 (1.2%)
				640	1500/1500	1 (0.1%)
				1000	1500/1500	0
Y	Y	Y	Y	50	1498/1500	Most
				100	1500/1500	Many
				320	1500/1500	25 (1.7%)
				640	1500/1500	0
				1000	1500/1500	0

Figure A. Simulation results – effect of accounting for measurement error and confounding when not present



All data generated using simplex models not allowing for measurement error (ME), nor for measurement error covariances representing unmeasured confounding (Conf). MSE = mean square error, Control = generated and analysed without ME and Conf when both ME & Conf absent, Allow for ME only = analyse allowing ME only when both ME & Conf absent, Allow for both = analysed without allowing for ME or Conf when both ME & Conf absent.