Supplementary Materials - Simulation studies

The main simulation study reported in the paper evaluated the biases of three competing estimators of mediation effects under six pertinent data generating models assuming effect homogeneity. Further simulations were carried out to evaluate the statistical properties (bias and standard error) of the best performing mediation approach (C) while allowing for effect heterogeneity. This supplementary section describes our overall simulation study design and summarizes the additional findings from the simulations under effect heterogeneity.

Data generating model

We simulated data from a trial of size *n*=500 with 1:1 randomisation ratio using the data generating model described in Figure 1:

Baseline variables

Past predictors of baseline values V_i are assumed to arise from a standard normal distribution. Baseline values are generated from

$M_{i,0} := l_1 V_i + \varepsilon_i^{(1)}$ and

$Y_{i,0} := l_2 V_i + \varepsilon_i^{(2)}$.

standardised to have unit variance; i.e. $\mathrm{var}\left(M_{i,0}\right) = 1$ and $\mathrm{var}\left(Y_{i,0}\right) = 1$. This implies that $\mathrm{var}\left(\varepsilon_i^{(1)}\right) = 1 - l_1^2$ with $|l_1| < 1$ and $\mathrm{var}\left(\varepsilon_i^{(2)}\right) = 1 - l_2^2$ with $|l_2| < 1$. Error terms $\varepsilon_i^{(1)}$ and $\varepsilon_i^{(2)}$ are independently normally distributed with zero means. Parameters l_1 and l_2 are correlation coefficients.

Time 1 potential outcomes

We generate potential outcomes whose variances under the control condition increase by 20% over time; i.e. we ensure that $\mathrm{var}\left[M_{i,1}(R=0)\right]=1.2\,\mathrm{var}\left(M_{i,0}\right)$ and $\mathrm{var}\left[Y_{i,2}(R=0)\right]=1.2\,\mathrm{var}\left(Y_{i,0}\right)$. Applying path tracing rules to Figure 1 this implies that $\mathrm{var}\left[M_{i,1}(0)-M_{i,0}\right]=0.2-2(c_1+l_1l_2d_2)$ with $c_1+l_1l_2d_2<0.1$ and that $\mathrm{var}\left[Y_{i,2}(0)-Y_{i,0}\right]=0.2-2[d_1+c_2l_1l_2+\beta l_1l_2+\beta(d_2+c_1l_1l_2)]$ with $d_1+c_2l_1l_2+\beta l_1l_2+\beta(d_2+c_1l_1l_2)<0.1$.

We define two potential mediator outcomes for $r=0{,}1$ by the linear structural model:

$$\left[M_{i,1}-M_{i,0}\right](r):=\alpha r+c_1M_{i,0}+d_2Y_{i,0}+\varepsilon_i^{(3)}(r) \quad \text{and}$$

$$M_{i,1}(r):=\left[M_{i,1}-M_{i,0}\right](r)+M_{i,0}$$

This implies that $\mathrm{var}\left[\varepsilon_i^{(3)}(0)\right]=0.2-2(c_1+l_1l_2d_2)-c_1^2-d_2^2-2c_1d_2l_1l_2$ with $0.2-2(c_1+l_1l_2d_2)>c_1^2+d_2^2+c_1d_2l_1l_2$. To accommodate this we generate $\varepsilon_i^{(3)}(r):=\mathrm{var}\left[\varepsilon_i^{(3)}(0)\right]^{0.5}[(1-f^2)^{0.5}\varepsilon_{Mi}+f(2r-1)\,\tau_{Mi}]$ with ε_{Mi} and τ_{Mi} independently standard normally distributed. Here $f\in[0{,}1]$ is a variance component reflecting the contribution of (random) heterogeneity across individuals in the effect of *r* to the error variance. Furthermore, parameter α is the target effect and $\alpha/\sqrt{1.2}$ can be interpreted as a standardised difference (Cohen's d). Parameters $c_1<0$ and d_2 determine the effect of the baseline measures on the mediator change scores.

Time 2 potential outcomes

We define four (counterfactual) clinical outcomes for $r_1=0{,}1$; $r_2=0{,}1$ by the linear structural model:

$$\left[Y_{i,2}-Y_{i,0}\right]\left[r_1,M_{i,1}(r_2)\right]:=\gamma r_1+\beta M_{i,1}(r_2)+d_1Y_{i,0}+c_2M_{i,0}+\varepsilon_i^{(4)}\left[r_1,M_{i,1}(r_2)\right] \quad \text{and}$$

$$Y_{i,2}\left[r_1,M_{i,1}(r_2)\right]:=\left[Y_{i,2}-Y_{i,0}\right]\left[r_1,M_{i,1}(r_2)\right]+Y_{i,0}$$

This then implies that $\mathrm{var}\left\{\varepsilon_i^{(4)}\left[0, M_{i,1}(0)\right]\right\} = 0.2 - 2[d_1 + c_2 l_1 l_2 + \beta l_1 l_2 + \beta(d_2 + c_1 l_1 l_2)] - d_1^2 - c_2^2 - 2c_2 d_1 l_1 l_2 - \beta^2 - c_2 d_1 l_1 l_2 - 2\beta d_1(d_2 + c_1 l_1 l_2 + l_1 l_2) - 2\beta c_2(1 + c_1)$ with $0.2 - 2[d_1 + c_2 l_1 l_2 + \beta l_1 l_2 + \beta(d_2 + c_1 l_1 l_2)] > d_1^2 + c_2^2 + 2c_2 d_1 l_1 l_2 + \beta^2 + c_2 d_1 l_1 l_2 + 2\beta d_1(d_2 + c_1 l_1 l_2 + l_1 l_2) + 2\beta c_2(1 + c_1)$. And we generate $\varepsilon_i^{(4)}\left[r_1, M_{i,1}(r_2)\right] := \mathrm{var}\left\{\varepsilon_i^{(4)}\left[0, M_{i,1}(0)\right]\right\}^{0.5}\left[(1 - z^2 - g^2)^{0.5}\varepsilon_{Yi} + zM_{i,1}(r_2)\zeta_{Yi} + g(2r_1 - 1)\tau_{Yi}\right]$ with $(1 - z^2 - g^2) \geq 0$ and $\varepsilon_{Yi}, \zeta_{Yi}$ and τ_{Yi} independently standard normally distributed. Here $z, g \in [0,1]$ are variance components reflecting the contribution of (random) heterogeneity in the effect of $M_{i,1}(r_2)$ or of r_1 respectively to the error variance. Furthermore, parameter γ is the natural direct effect with $\gamma/\sqrt{1.2}$ representing a standardised difference. Parameter β is the causal effect of the mediator on the clinical outcome expressed as a standardised regression coefficient. Parameters d_1 with $d_1 + d_2\beta < 0$ and c_2 determine the effect of the baseline measures on the clinical change scores.

Observed variables

We then map the potential outcomes onto observable variables by calculating

$M_{i,1} = (1 - R_i)M_{i,1}(0) + R_i M_{i,1}(1)$

and $Y_{i,2} = (1 - R_i)Y_{i,2}\left[0, M_{i,1}(0)\right] + R_i Y_{i,2}\left[1, M_{i,1}(1)\right]$

We note that, strictly speaking, we needed only to generate two potential outcomes, namely $Y_{i,2}\left[0, M_{i,1}(0)\right]$ and $Y_{i,2}\left[1, M_{i,1}(1)\right]$. However, we generate all four potential outcomes to enable us to validate our simulations by confirming values for the intention-to-treat effect and and the causal mediation effects.

Please note the following points regarding our simulation models: (i) We chose to set all intercepts to zero for simplicity. (ii) We inflated the variance of outcomes under the control condition by 20% over time. A change score model that holds the variance constant implies that baseline measures and change scores of a variable are negatively correlated. Some variance inflation needed to be allowed to ensure that individual outcome values can vary over time even if $c_1 = 0$ or $d_1 + d_2\beta = 0.$ (iii) We used a random coefficient model to simulate effect heterogeneity.

Main simulation study

The main simulation study evaluated the statistical properties of the competing estimators assuming effect homogeneity; i.e. we set $f = g = z.$ We simulated from the six different data generating models listed in Table 1. This was achieved by following our general data generating model but imposing parameter restrictions as shown in Table 1 to reflect the specific confounding process. The remaining simulation parameters were set to typical values in mental health trials. These value choices and their rationales are summarized in Table S1.

Table S1 Parameter value choices and their rationales.

Parameter	Interpretation	Value(s)	Rationale
l_1	Correlation between latent common cause and baseline measure of putative mediator.	0.5	"Moderate size" correlation between baseline measures of mediator and clinical outcome
l_2	Correlation between latent common cause and baseline measure of clinical outcome.	0.5	As above.
f	Square is error variance component reflecting treatment offer effect heterogeneity.	0, $\sqrt{1/6}$, $\sqrt{1/3}$	Values chosen to reflect treatment effect heterogeneity contributing up to a third of error variance.
α	Target effect	0.5	"Moderate size" Cohen's d.

c_1	Effect of baseline measure of mediator on change in mediator.	-0.5	Chosen to provide approx. 0.5 correlation between baseline and time 1 measures in the control group (typical for mental health trials).
d_2	Effect of baseline measure of clinical outcome on change in mediator.	0.15	Chosen to be of smaller absolute size than the effect of the baseline measure of the mediator itself.
g	Square is error variance component reflecting direct treatment offer effect heterogeneity.	0, $\sqrt{1/6}$, $\sqrt{1/3}$	Values chosen to reflect treatment effect homogeneity up to "strong" effect heterogeneity.
z	Square is error variance component reflecting mediator effect heterogeneity.	0, $\sqrt{1/6}$, $\sqrt{1/3}$	Values chosen to reflect mediator effect homogeneity up to "strong" effect heterogeneity.
γ	Direct effect of treatment offer on clinical outcome	0.375	"Small/moderate size" Cohen's d.
β	Causal effect of mediator on clinical outcome (implies NIE=0.125, $\text{ATE}_Y = 0.5$) .	0.25	"Moderate size" standardised regression coefficient.
d_1	(Direct) effect of baseline measure of clinical outcome on change in this outcome.	-0.5	Chosen as providing approx. 0.5 correlation between baseline and time 2 measures in the control group (typical for mental health trials).
c_2	(Direct) effect of baseline measure of mediator on clinical outcome not operating via changing the mediator at time 1	-0.10	Chosen so that the total effect size for the path $M_0 \rightarrow Y_2$ is the same as that for the path $Y_0 \rightarrow M_1$ (cf Figure 1).

We repeatedly sampled from these models to generate mediator and clinical outcome variables; construct respective estimators and mimic sampling distributions. The bias findings of the main simulation study have been summarized in our paper.

Additional simulation study

We carried out further simulations to study the impact of random effect heterogeneity. Only approach (C) was able to provide unbiased estimates under all models considered assuming effect homogeneity (cf Table 2). Thus only the statistical properties of this approach were evaluated further. For this assessment we simulated from the general data generating model without imposing further parameter restrictions – thus effectively allowing for all confounding processes involving baseline measures to operate. Importantly, we now varied the parameters that determined the effect variability across individuals. The variabilities of individual treatment effects are determined by the parameters f (inter-individual differences in the treatment effect on the mediator variable) and g (inter-individual differences in the direct treatment effect on the clinical outcome). The variability of individual mediator effects is determined by parameter z (inter-individual differences in the effect of the mediator variable on the clinical outcome). We considered 9 models reflecting combinations of three levels of treatment effect heterogeneity (“effect homogeneity”, “mild effect heterogeneity” and “strong effect heterogeneity”) and mediator effect heterogeneity (also “homogeneity”, “mild heterogeneity” and “strong heterogeneity”).

Table S2: Simulation results for approach (C) under effect heterogeneity: Expected values of estimators based on s=10000 simulations. (Where available, estimator SEs and closed-form estimated values are shown in square and curly brackets respectively. Biases are indicated in italics.)

Effect heterogeneity			**True estimand value**				
f	g	z	α=0.5	β=0.25	NIE=0.125	NDE=0.375	$\mathrm{ATE}_Y = 0.5$
0	0	0	0.500 [0.085] {0.085}	0.250 [0.045] {0.044}	0.125	0.376 [0.088] {0.087}	0.501
$\sqrt{1/6}$	$\sqrt{1/6}$	0	0.500 [0.084] {0.084}	0.250 [0.044] {0.044}	0.125	0.375 [0.087] {0.087}	0.500
$\sqrt{1/3}$	$\sqrt{1/3}$	0	0.500 [0.084] {0.084}	0.250 [0.044] {0.044}	0.125	0.375 [0.087] {0.087}	0.500
0	0	$\sqrt{1/6}$	0.500 [0.084] {0.084}	0.250 [0.051] {*0.045*}	0.125	0.376 [0.088] {0.088}	0.501
$\sqrt{1/6}$	$\sqrt{1/6}$	$\sqrt{1/6}$	0.500 [0.084] {0.084}	0.251 [0.050] {*0.045*}	0.126	0.375 [0.088] {0.088}	0.501
$\sqrt{1/3}$	$\sqrt{1/3}$	$\sqrt{1/6}$	0.500 [0.084] {0.084}	0.251 [0.050] {*0.045*}	0.126	0.375 [0.088] {0.088}	0.501
0	0	$\sqrt{1/3}$	0.500 [0.084] {0.084}	0.251 [0.056] {*0.046*}	0.126	0.376 [0.089] {0.089}	0.501
$\sqrt{1/6}$	$\sqrt{1/6}$	$\sqrt{1/3}$	0.500 [0.084] {0.084}	0.251 [0.056] {*0.046*}	0.126	0.375 [0.088] {0.089}	0.501
$\sqrt{1/3}$	$\sqrt{1/3}$	$\sqrt{1/3}$	0.500 [0.084] {0.084}	0.251 [0.056] {*0.046*}	0.126	0.375 [0.088] {0.089}	0.501

Table S2 shows the findings from the additional set of simulations. The expected values of the model-based SE estimators (shown in curly brackets) can be contrasted with the true estimator SEs (shown in square brackets) to assess bias in variance estimators. We observed the following:

- Approach (C) continued to provide unbiased estimators of all mediation parameters under effect heterogeneity.
- The (true) variance of the target effect estimator was not affected by the level of treatment effect heterogeneity (f). (It cannot be affected by choices for g or z.) This variance value was estimated without bias under all scenarios considered here.
- The variance of the natural direct effect estimator remained constant across different levels of direct treatment effect heterogeneity (g), mediator effect heterogeneity (z) as well as of f. This variance value was estimated without bias under all scenarios considered here.
- The variance of the β estimator increased with increasing heterogeneity of the mediator effect (z). When such effect heterogeneity was present ($z \neq 0$) this extra variance was not captured and the variance estimator provided by Approach (C) was downwardly biased.

We suggest that the reason for the unaccounted variance inflation in the ANCOVA estimator of β is that the sample distribution of the mediator variable varies across repeated samples while that of the randomisation variable is held constant by our trial design. This extra variability is not captured by an estimator based on the conditional distribution of the clinical outcome given the mediator. However, we would expect a variance estimator constructed by nonparametric bootstrapping to capture this extra variability since bootstrapping is trying to mimic the full data generating process.