


METHODOLOGY

Open Access



Applying a hypothetical strategy to the intercurrent event of non-adherence with the parametric g-formula: a post hoc secondary analysis of the MET-PREVENT randomised controlled trial

Shaun Hiu^{1,5}, Nina Wilson¹, Kevin Wilson², Nan Lin¹, Miles D. Witham^{3,4}, James M. S. Wason^{1*}  and The M. E. T.-PREVENT Study Group

Abstract

Background We conducted a post hoc secondary analysis of the MET-PREVENT randomised placebo-controlled trial to target an estimand that uses a hypothetical strategy on the intercurrent event of non-adherence.

Methods We viewed the targeting of this estimand that uses a hypothetical strategy to handle the intercurrent event of non-adherence as a type of causal mediation problem, where randomised treatment is a binary exposure, adherence is a binary mediator, there is an exposure-mediator interaction, and mediator-outcome confounders that are caused by the exposure (e.g. gastrointestinal symptoms) are present. We used the parametric g-formula to estimate the average controlled direct effect (CDE) of metformin (versus placebo) on 4-m walk speed, which is interpreted as the average treatment effect under the hypothetical scenario that all trial participants adhered to assigned treatment. Variables identified as confounders were informed by a literature review and discussions with an expert; assumptions about the causal structure were represented in a directed acyclic graph. We applied a probabilistic bias analysis (PBA) to understand the potential for bias assuming adherence had been misclassified; observed adherence based on returned tablet count may be an inaccurate version of “true adherence” based on actual consumption.

Results Our sample size was 70 trial participants (34 metformin, 36 placebo). Our estimate of the CDE was 0.072 m/s (percentile-based bootstrap 95% CI –0.292, 0.445). Results from PBA indicated that the greater the extent of misclassification, the more the CDE may be estimated with bias and over-optimistic precision.

Conclusions Our study provided supporting information on metformin’s potential role as a repurposed medication to improve physical performance in nondiabetic older adult patients with physical prefrailty/frailty and probable sarcopenia. Unlike the main trial results, our results do not rule out the possibility of either a meaningful benefit or meaningful harm of metformin, provided that full adherence can be assured. We highlighted the parametric g-formula as a useful method in trials to target estimands with a hypothetical strategy to handle treatment non-adherence.

Trial registration ISRCTN ISRCTN29932357.

*Correspondence:

James M. S. Wason

James.Wason@newcastle.ac.uk

Full list of author information is available at the end of the article



© The Author(s) 2026. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Keywords Estimands, Causal inference, G-formula, Mediation, Secondary analysis, Quantitative bias analysis

Introduction

The de facto position of randomised placebo-controlled superiority trials is to make conclusions from effects that preserve the intention-to-treat (ITT) principle, which is analogous to the treatment policy estimand, followed by accompanying per-protocol (PP) analyses whereby non-adherers are excluded. It is known that such ITT effects are particularly sensitive to intercurrent events since any event occurring after randomisation is essentially ignored in the analysis [1]. Treatment non-adherence (which includes discontinuation) may be such intercurrent events. Intercurrent events are an attribute of the estimands framework [2, 3] and are defined as post-randomisation events that affect the interpretation or existence of the outcome of interest. Differential adherence rates between arms (whereby the intervention arm usually has the lower adherence rate possibly due to the adverse effects of the new treatment) make it challenging to separate between two propositions for observed null findings from ITT results: either the treatment regime under investigation is inefficacious in the target population or it is efficacious but was observed to be inefficacious due to low adherence. Making this distinction is important because adherence is a behaviour that can be improved through intervention.

In this study, we conducted a post hoc secondary analysis of the MET-PREVENT trial to target an estimand that uses a hypothetical strategy to handle the intercurrent event of non-adherence with the parametric *g*-formula to estimate the average controlled direct effect (CDE) of metformin (versus placebo) on 4-m walk speed under the hypothetical scenario that all participants adhered [4].

We viewed the targeting of this estimand as a type of causal mediation problem, where randomised treatment is a binary exposure, adherence is a binary mediator, there is an exposure-mediator interaction (assuming that the active compound in the medicinal product confers treatment benefit, adherence to the medicinal product may induce beneficial changes in physical performance greater than when adhering to an inert placebo), and presence of mediator-outcome confounders that are caused by the exposure (e.g. gastrointestinal symptoms). Causal mediation is an area of research methodology that seeks to address mechanistic ‘how does the intervention or exposure demonstrate its effects’ questions through the lens of the various schools of thought in the field of causal inference. In this paper, we use the potential outcomes framework [5–8]. The parametric *g*-formula (also known as *g*-computation) overcomes

limitations of conventional methods (i.e. the Baron and Kenny approach [9]) in investigating causal effects when adherence plays the role of a mediator. Traditional PP analyses that involve restricting the sample to those who are adherent are at risk of selection bias as an event that happens after randomisation is conditioned upon. Valid estimation of PP effects requires controlling for factors that influence adherence (i.e. mediator-outcome confounders) [10, 11]. However, where there is exposure-mediator interaction and mediator-outcome confounders are caused by the exposure, there is a need to turn to a family of methods known as “*g*-methods” of which the *g*-formula (a generalisation of regression standardisation) is a member [12].

Methods

Details of the trial

MET-PREVENT was a multicentre, proof-of-concept, superiority randomised controlled trial (RCT) of older adults with probable sarcopenia and frailty or prefrailty, but without diabetes mellitus, to investigate if metformin, a glucose-lowering therapy, was efficacious in improving physical performance in this population. The primary analysis was an intention-to-treat (ITT) analysis which found no statistically significant difference between the arms (0.001 (95% CI –0.06, 0.06) m/s). Under the estimand framework, this is akin to using a treatment policy strategy to handle the intercurrent event of non-adherence. The primary results have been published [13]. Adherence to medication was observed only at the end of the trial when tablets were returned and counted; a participant was defined as adherent if they consumed $\geq 80\%$ of the expected number of tablets they should have taken by the final four-month visit. There was a noticeable difference in adherence rates between the intervention (53%) and the placebo (78%) arms, respectively.

Seventy-two participants were randomised 1:1 (using minimisation with a 30% random element) between metformin and matched placebo. The primary outcome was the 4-m walk-speed test at a 4-month visit [14]. A minimal clinically important difference (MCID) of 0.1 m/s was pre-specified. One participant withdrew before receiving their medication and one died during follow-up, leaving $n=70$ with available data on the primary outcome (4-m walk speed). Because there was data on only one death and study withdrawal, respectively, we could not model these events. Scheduled visits occurred at baseline, one month, two months, and

three months after randomisation, with a final visit at four months where the primary outcome was measured. No participant was lost to follow-up. Details on the trial eligibility criteria and design are available in a published protocol [14]. The trial is registered on ISRCTN (ISRCTN29932357).

Treatment assignment is regarded as the exposure in our secondary analysis. Participants were randomised to receive metformin 500 mg or matching placebo tablets. Dosage was three tablets per day. All participants were supplied with 372 tablets at the point of medication reception to last the entire four-month follow-up duration.

True adherence, based on knowledge of actual tablet consumption, was regarded as the mediator. However, in the absence of this knowledge, observed adherence, defined as a medication possession ratio (MPR) $\geq 80\%$ at the final visit when unused tablets were returned and counted, was treated as the closest (but potentially misclassified) surrogate to true adherence in the data. The MPR could not be computed for two participants in the intervention arm as tablets were disposed of by accident and treated as missing data: one participant was classified as non-adherent as they experienced gastrointestinal symptoms leading to treatment discontinuation at 5 days after randomisation, while the other participant was classified as adherent based on mean imputation of the MPR.

The primary outcome was the 4-m walk speed (in m/s) measured at the four-month visit. All participants were tasked to walk 4-m on a measured course and at a pace they normally walk at if they were walking down the

street. Participants were allowed to use their walking aid. The faster of two attempts was the primary outcome.

Estimand attributes

Research question: If there exists a way to intervene on patients’ adherence to medication such that everyone would adhere, what would be the causal effect of taking metformin versus placebo on physical performance?

Population: Physically pre-frail and frail nondiabetic older adults with probable sarcopenia.

Treatment conditions: Metformin and placebo.

Endpoint: The 4-m walk speed (in m/s) measured at four months after randomisation.

Summary measure: Average controlled direct effect (interpretable as a difference in means).

Intercurrent event and strategy: Non-adherence will be handled by a hypothetical strategy which supposes that through some hypothetical and conceivable intervention on adherence, all participants would adhere (MPR $\geq 80\%$) to assigned treatment over a four-month period.

Assumptions about underlying causal relationships

We make explicit our assumptions about the underlying causal relationships in a directed acyclic graph (DAG) (Fig. 1); we did not assume unmeasured exposure-outcome confounding owing to randomisation. Variables identified as mediator-outcome confounders were informed by the intersection between the literature on factors which influence adherence to medication [15–18] and predictors of walk speed amongst older adults, and by discussions with an expert (MDW).

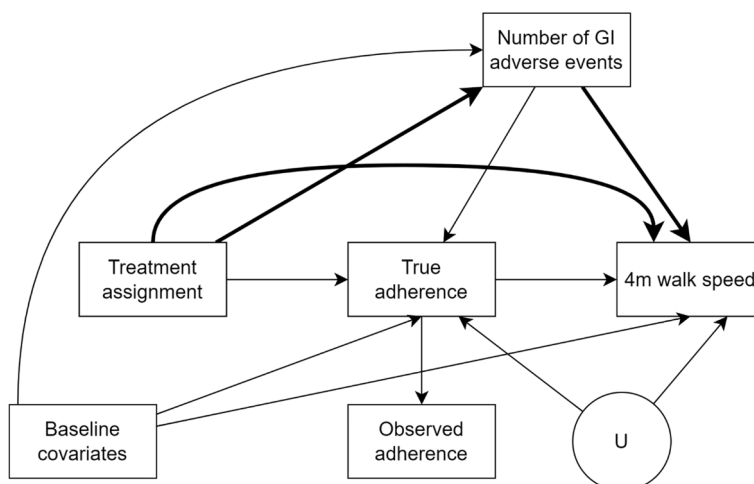


Fig. 1 Directed acyclic graph representing assumptions of causal relationships. The observed adherence variable in our data (which is based on returned tablet counts) may be a misclassified version of “true adherence” (which is based on actual consumption and not known to us). The variable U represents unmeasured mediator-outcome confounder(s) that are not available in the data. The bolded paths collectively make up the CDE of treatment assignment on the outcome

Baseline mediator-outcome confounders included age [19, 20], sex [19], obesity (body mass index ≥ 30 kg/m²) [21, 22], history of depression [23, 24], functional ability as defined by the Nottingham Extended Activities of Daily Living (NEADL) [20], polypharmacy (≥ 5 concomitant medications) [25], and baseline 4-m walk speed. We included baseline walk speed as it is conceivable that impairment influences the ability to adhere. For example, individuals who were more physically impaired may be more motivated to adhere given that participants were informed of the premise of the trial and blinded to treatment allocation. On the other hand, it is also plausible that individuals with greater physical impairment may find it harder to adhere. The mediator-outcome confounder caused by the exposure was the number of gastrointestinal adverse events over follow-up (0, 1, or ≥ 2 events). Metformin is known to potentially cause gastrointestinal (GI) symptoms, such as diarrhoea and nausea, which may discourage adherence [15]. We postulated two ways in which GI symptoms may influence walk speed. Participants who experience GI symptoms may perform worse on physical performance measures, especially if the GI side effects made them less active and they became physically deconditioned. Alternatively, if metformin's effect on skeletal muscle is due to a change in the gut microbiome, it may be that only those who do experience GI symptoms achieve benefit. Adverse events were recorded in an adverse events log and GI adverse events are defined as those which were coded under "gastrointestinal disorders" within the Medical Dictionary for Regulatory Activities (MedDRA[®]) terminology (version 25) [26]. We replaced the missing subscale items on the NEADL with the participant-specific median from the completed items in the subscale, provided less than three items within a subscale were missing.

Statistical analysis

We used the parametric g-formula (also known as g-computation) to identify the CDE. Briefly, the g-formula is a generalisation of regression standardisation [27]. We refer readers to Appendix A where we introduce the notation for potential outcomes [6, 7], introduce the g-formula, and briefly describe g-computation. A stepwise process for the parametric g-formula and our modelling assumptions are described in Appendix B. We reported the estimate of the CDE and a percentile-based 95% confidence interval (CI). We also reported a bias-corrected and accelerated (BCa) bootstrap 95% CI for comparison; the relevance of the BCa bootstrap is included in our discussion but, briefly, it has been proposed to improve the coverage of the parametric g-formula estimator, especially for small sample sizes [28]. The

g-formula algorithm is provided in greater detail elsewhere [29]. Analyses were carried out in R version 4.3.0.

Sensitivity analyses

We assessed how well our model specifications agreed with our assumptions about the underlying causal structure by estimating the natural course [30]. The idea of the natural course is if one were to derive the observed sample mean of the primary outcome through the parametric g-formula, it would require us to specify parametric models for the conditional probability of treatment, adherence, occurrence of GI adverse events, and the conditional mean of the primary outcome. If all these models were correctly specified, then the summary statistics and distributions of the variables that are predicted by the parametric g-formula algorithm should approximately match the observed data. We compared the mean, standard deviation, and empirical cumulative distribution function (eCDF) of the observed primary outcome data against the data predicted by the parametric g-formula, and the means of observed adherence and GI adverse events data against those predicted by the parametric g-formula.

Observed adherence rates based on returned tablet count may overestimate "true adherence" rates based on actual consumption, and there may be concerns over bias due to a misclassified mediator. To understand the impact of misclassifying adherence, we applied a probabilistic bias analysis (PBA) that adjusted the 'naïve' estimate of the CDE for misclassification by incorporating information from the literature surrounding the possible extent of misclassification [31]. The aim of a PBA was to gain insight into the impact that misclassification had on the magnitude and direction of bias in our 'naïve' CDE estimate, as well as on our uncertainty [32]. Unlike quantitative bias analysis which assumes that the single estimate of the degree of misclassification chosen by the investigator is the true value, PBA entertains the possibility that there is likely a range of other plausible values for the extent of misclassification and that some values are more likely than others. We simulated four conditions that represented: 1) a very high certainty in our beliefs that there is a low degree of misclassification; 2) a very high certainty in our beliefs that there is a severe degree of misclassification; 3) more moderate certainty in our beliefs that there is a low degree of misclassification; and lastly 4) more moderate certainty in our beliefs that there is a severe degree of misclassification. Detailed methodology on the PBA is available in Appendix C. We reported the median and percentile-based 95% simulation interval (95% SI) of the misclassification bias-adjusted CDE. The misclassification bias-adjusted point estimate was compared against the naïve point estimate,

as well as the widths of the simulation intervals versus the confidence interval [32].

Results

Summary characteristics of our sample are in Table 1. Thirteen participants in the control arm (36.1%) and 27 participants in the intervention arm (79.4%) experienced at least one GI adverse event over the course of follow-up. Of the 27 in the control arm, 10 discontinued due to GI adverse events (9 participant-led, 1 investigator-led). Of the 13 in the control arm, 1 discontinued due to a GI adverse event which was participant-led. Two participants assigned to metformin discontinued due to elevated lactate concentrations (>4 mmol/L at the 1-month and 3-month visits, respectively) and one participant assigned to placebo discontinued due to hypoglycaemia (<4 mmol/L) at the two-month visit. None discontinued due to severe renal impairment.

Our estimate of the CDE was 0.072 (percentile-based bootstrap 95% CI -0.292, 0.445) m/s. We also reported the BCa bootstrap 95% CI (-0.244, 0.519) for comparison. We conducted a natural course analysis to validate our models used in the parametric g-formula. We observed agreement, to an extent, between the observed data and the data predicted from the parametric g-formula algorithm under the natural course (Table 2). The eCDFs of the observed and predicted primary outcome data are presented in Fig. 2. We observed that there was

Table 2 Comparison of observed data and data predicted by the parametric g-formula estimator in the natural course analysis

	Observed data (n = 70)	Predicted data (n = 10,000)
Number of GI adverse events		
No events, %	42.86%	40.66%
One event, %	42.86%	45.04%
≥ 2 events, %	14.29%	14.30%
Adherence		
Observed adherence, %	67.14%	66.80%
4-m walk speed at four months, m/s		
Mean (SD)	0.58 (0.22)	0.56 (0.20)
1 st percentile	0.20	0.19
2.5th percentile	0.21	0.21
5th percentile	0.22	0.23
25th percentile	0.41	0.41
50th percentile	0.56	0.58
75th percentile	0.75	0.70
95th percentile	0.94	0.94
97.5th percentile	1.02	1.06
99th percentile	1.08	1.07

GI gastrointestinal, SD standard deviation

close agreement on the proportion of GI adverse events, adherence, and the distribution of the primary outcome. Our PBA results are presented in Table 3.

Table 1 Characteristics of the sample

	Analysis set (n = 70)
Treatment arm, n (%)	
Metformin	34 (48.57)
Placebo	36 (51.43)
Age, years, mean (SD)	80.49 (5.75)
Sex, n (%)	
Female	41 (58.57)
Male	29 (41.43)
History of depression, n (%)	19 (27.14)
Obesity (body mass index ≥ 30 kg/m ²), n (%)	18 (25.71)
Nottingham Extended Activities of Daily Living, mean (SD)	15.53 (4.94)
Polypharmacy (≥ 5 medications), n (%)	60 (85.71)
Baseline 4-m walk speed, m/s, mean (SD)	0.59 (0.22)
Four month 4-m walk speed, m/s, mean (SD)	0.58 (0.22)
Observed adherence, n (%)	47 (67.14)
Number of gastrointestinal adverse events over follow-up, n (%)	
0	30 (42.86)
1	30 (42.86)
2	2 (5.71)
3	3 (7.14)
4	4 (1.43)

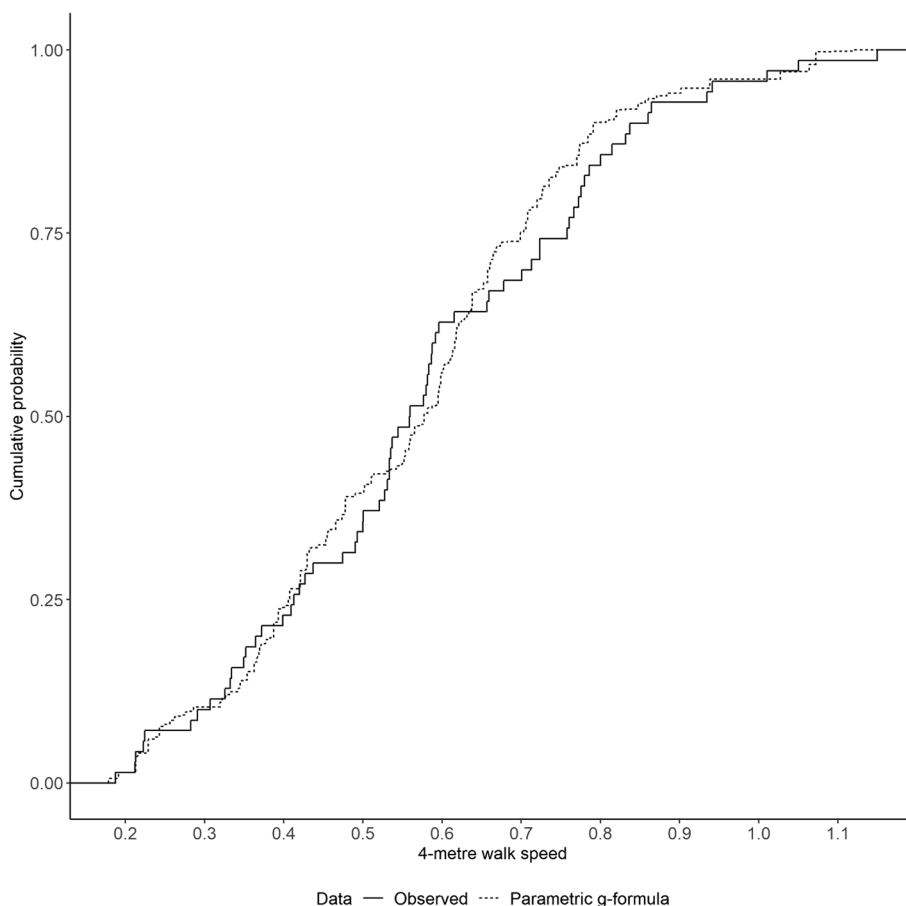


Fig. 2 Empirical cumulative distribution functions of the observed and predicted primary outcome data

Table 3 Results of probabilistic bias analysis

Scenario	Description	Prior Beta distribution of the FPR	Median [95% SI]
1	Very high certainty of low degree of misclassification	Beta(33.3, 231)	0.066 [−0.657, 0.818]
2	Very high certainty of severe degree of misclassification	Beta(355, 205)	0.020 [−0.744, 0.760]
3	Moderate certainty of low degree of misclassification	Beta(4, 26)	0.069 [−0.687, 0.809]
4	Moderate certainty of low degree of misclassification	Beta(39.8, 25)	0.029 [−0.728, 0.779]

FPR false positive rate, SI simulation interval

Discussion

Our post hoc secondary analysis supports the trial’s primary finding of insufficient evidence to conclude a beneficial effect of metformin on older adults with physical pre-frailty/frailty and probably sarcopenia. However, in contrast to the prespecified primary analysis findings, our results do not rule out the possibility of either a meaningful benefit or meaningful harm of metformin, provided that the consumption of ≥ 80% of prescribed tablets can be assured. Both the percentile-based and BCa bootstrap CIs of our CDE included zero as well as

the MCID on both sides of the null, indicating that a null effect, meaningful benefit, and meaningful harm are all consistent with our analysis [33]. It should be reiterated that MET-PREVENT was not originally designed to accommodate causal mediation analyses, so the results presented should be interpreted as exploratory and hypothesis-generating.

Our main analyses assumed that there was no misclassification of treatment adherence and that observed adherence faithfully reflected true adherence. If on the other hand we suspect there to be misclassification, our

sensitivity analyses indicated that the greater the extent of misclassification, the more the CDE may be estimated with bias (the estimate would have been closer to the null if true adherence had been collected and used in the parametric g-formula analysis) and with over-optimistic precision (the confidence intervals would have been wider if true adherence had been collected and used in the parametric g-formula analysis). Because the actual extent of misclassification is not known, we ran our PBA under multiple scenarios to cover different sets of beliefs about the misclassification. It could be argued that the low misclassification scenarios were more plausible than the severe misclassification scenarios. This is based on the premise that participants on clinical trials are generally motivated to follow trial processes especially if an investigational product has the potential to improve the symptoms of a progressive disorder like sarcopenia. Hence, the plausible source of misclassification may be benign forgetfulness such as accidental disposal as opposed to an active intent to defy protocol.

Strengths

Methods based on the potential outcome framework have utility in overcoming some of the limitations of traditional per-protocol methods which may unintentionally subject themselves to bias [11]. To illustrate this, we compare our estimates with those from the trial's original PP analysis of 0.023 (95% CI -0.053, 0.099) m/s. The noticeable difference between the point estimates may reflect that the original PP analysis was subjected to collider-stratification bias when adherence was conditioned upon (by limiting the analysis to adherers only) [5, 34]. A collider is variable C that is caused by two other variables A and B (hence the effects of the two variables on C 'collide'). The effect of A on B is estimated with (collider-stratification) bias when a model regressing B on A also includes C as a covariate. With reference to Fig. 1, and assuming that it accurately reflects the underlying causal structure, limiting the analysis to adherers only induces a noncausal association between treatment assignment and the outcome through the baseline covariates and number of GI adverse events because adherence is a collider on four backdoor paths through GI adverse events. If one were to adjust for the baseline covariates and number of GI adverse events in a traditional outcome regression model (e.g. analysis of covariance), a portion of the direct effect of treatment assignment on the outcome which goes through number of GI events would be removed. Hence, the need for g-methods to avoid collider-stratification bias, unintended partial subtraction of direct effects, and importantly interpret results as causal effects. A more detailed introduction to directed acyclic

graphs, collider-stratification bias, and backdoor paths is available elsewhere [5].

The strength of the parametric g-formula is that parametric models allow investigators to extrapolate beyond the observed data, thus positivity violations due to sparsity of data are less of an issue compared to methods using inverse probability weights [35]. However, this trade-off means the parametric g-formula relies heavily on correct model specification to extrapolate over strata of covariates with sparse data. We attempted to make appropriate model specifications as much as possible by referencing prior literature and seeking expert opinion on potential mediator-outcome confounders, and relaxing modelling assumptions by allowing for interactions between treatment and all covariates and the inclusion of non-linear terms of continuous covariates in the primary outcome model. On the other hand, baseline cognitive functioning [15, 17, 36, 37] may be an unmeasured mediator-outcome confounder such that higher cognitive functioning may lead to better adherence and walk speed. Under unmeasured confounding, a qualified interpretation of our estimates is necessary. If the confounder(s) exist and increase the likelihood of adherence and the average value of the outcome simultaneously, then our estimate of the CDE may be interpreted as an estimate for the lower bound of the true CDE instead [4]. The impact of this confounder was partially mitigated against as none of the participants was diagnosed with dementia and that baseline walk speed and functional ability are correlated with cognition [38, 39]. Nevertheless, this means that our estimate of 0.072 m/s may be regarded as our best estimate of this lower bound, and the true effect of metformin under unanimous adherence may be larger than this quantity. However, given the confidence interval, a negative-valued lower bound is consistent with our data, which maintains the possibility of the true CDE being zero. Sensitivity analyses such as the use of E-values [40] have been proposed to investigate how large the exposure-unmeasured confounder and outcome-unmeasured confounder associations would minimally have to be to "flip" a statistically significant effect into a non-statistically significant result.

Limitations

The use of the parametric g-formula at small sample sizes carries limitations. Few studies have evaluated the performance of the parametric g-formula estimator in small sample sizes ($n \leq 100$). Tackney and colleagues [28] examined the performance of the parametric g-formula estimator of the average treatment effect (ATE) of a point exposure on a continuous outcome. They examined the performance of the estimator when the sample size was small ($n=50$ and $n=100$) and the model for

the conditional mean of their outcome included a treatment variable, 17 covariates (which were all predictive of the outcome), and interactions between treatment and all covariates. Provided all confounders have been accounted for, their results indicate that whilst the estimator had a mean bias close to zero, but there was suboptimal coverage (approximately 90% at either sample size), which was due to the model-based standard errors underestimating the empirical standard error. Coverage was further reduced when including four additional covariates that were unpredictable of the outcome. These findings regarding suboptimal coverage are supported by Chatton and colleagues [41] who evaluated the performance of the parametric *g*-formula estimator of the average treatment effect of a point exposure when the outcome was binary, the sample size was 100, and the outcome model included 9 covariates (6 were predictive of the outcome) and the treatment variable. Assuming correct model specification, their results indicated that the estimator had a mean bias close to zero but coverage was suboptimal at 91.8%; reducing the number of covariates appeared to shift the coverage probability closer towards the desired 95% probability. Small sample corrections such as BCa bootstrap have thus been proposed to improve coverage [28].

A limitation of the CDE is that all participants are always set to adhere through some conceivable intervention. This may not be realistic as interventions should allow treatment to be discontinued at any point when a participant experiences a clinical event that jeopardises their safety. Relevant events in the context of metformin would be elevated lactate concentrations, hypoglycaemia, and severe renal impairment. One strategy would have been to conceive of a dynamic intervention on adherence that allows for non-adherence when participant safety is compromised at any point during follow-up [42]. Given that the incidence rates of elevated lactate concentrations, hypoglycaemia, and severe renal impairment following metformin use are low based on previous reports and our observations [43–45], it may be reasonable to interpret our effect as a close approximation to the effect under this dynamic intervention on adherence.

We observed a relatively wider confidence interval of the CDE compared to that of the trial's ITT and PP effects. One possible reason could be that controlled direct effects estimated by the parametric *g*-formula have greater sample size requirements for the same level of precision because of the added modelling assumptions. A previous study conducted a secondary analysis of a placebo-controlled trial to investigate the per-protocol effect of complete adherence to low-dose aspirin at pre-conception to improve live births amongst 1227 women with documented pregnancy losses who were actively

trying to conceive [46]. Compared to the trial's ITT effect of a relative risk of live birth of 1.10 (95% CI 0.98, 1.22), the parametric *g*-formula estimate of the per-protocol effect in which all participants adhered was 1.33 (95% CI 1.08, 1.64), an approximately two-fold increase in the width of the confidence interval on the log relative risk scale. Another reason could be that at the given sample size the models were overfitted to the data. However, Tackney and colleagues [28] showed by simulation that even at a fixed sample size of 50, increasing the number of covariates or even adding interaction terms between treatment and all covariates did not substantially change the model-based standard error or empirical standard error of the parametric *g*-formula estimator of the ATE of a point exposure on a continuous outcome. More familiar methods like inverse probability of treatment weighting have been shown to produce biased estimates under small sample sizes and are relatively less powerful than the parametric *g*-formula [28, 41]. Further work could be developed to investigate if these findings also extend to mediation analyses.

Implications for trial methodologists

Our study allowed us to draw out lessons for trialists. We highlighted the parametric *g*-formula as a useful method in trials to target estimands that are concerned with treatment adherence. Our study shines a light on two contexts in which targeting an estimand based on adherence is meaningful.

Firstly, efficacy trials may benefit most where maximising adherence is important so as not to miss a possible benefit of treatment. Meeting the threshold for adherence for all participants in an efficacy-trial setting could possibly be reached perhaps by introducing an initial titration period [47] plus raising awareness amongst participants about the placebo effect, i.e. a portion of GI adverse events may be attributable to negative expectations about the medicine rather than a pharmacological effect [48]. Other strategies may also include patient education and counselling [15], shared decision-making between clinicians and patients [15], dose adjustment [49], use of extended-release formulations [49], and lifestyle adaptations (e.g. time medication intake around meals [49]). However, in effectiveness trials, it is much more important that the trial conditions reflect how the drug will be used in the real world, so if low adherence is going to be an issue in clinical practice, seeing this in the trial is preferable and thus there may not be much to be gained from targeting an estimand that uses a hypothetical strategy to handle non-adherence.

Secondly, while there may be no substitute for efforts to actually maximise adherence at the design and implementation stages, methods such as the parametric

g-formula to target estimands using a hypothetical strategy to handle non-adherence may be the best alternative when maximising adherence is challenging in practice. If there is an a priori interest in the treatment effect under an ideal circumstance of unanimous adherence, then there may be a need to build in an analysis to estimate the CDE at the trial design stage as it may influence sample size. Generally, closed-form solutions are not readily available for causal effects, thus Monte Carlo simulation will be needed [50, 51]. Investigators should bear in mind that sample size requirements will likely be larger based on our earlier discussion on the relative widths of the confidence intervals, and so the relative merits of a primary research question requiring the CDE and running a larger trial need to be weighed.

Third, the methodology used in the current study bears some connection to principal stratification causal effects but remains distinct from them. Our study's main premise is that there exists a conceivable way to intervene on adherence such that all participants would adhere. In situations where such a hypothetical strategy is not plausible, using a principal stratification strategy to target the complier-averaged causal effects may be more clinically relevant [52, 53]. In this case, inference is on a latent subgroup of participants who would always be adherent to the metformin treatment regimen should they be assigned to the metformin arm.

Lastly, our study adds to the existing literature that ideas in causal inference and the potential outcomes framework are complementary to the estimand framework [54]. Primary estimands in trials will typically adopt a treatment policy strategy to handle intercurrent events. Our study highlights that estimands which use a hypothetical strategy on intercurrent events that can conceivably be intervened on (such as adherence) are useful supplements and may be summarised by the CDE. The CDE may provide additional information on aspects of the disease process by allowing for the exploration of intercurrent events as mediators and honing in on certain pathways that the treatment may have on the outcome.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13063-026-09708-1>.

Supplementary Material 1. Appendix A: Introduction to potential outcomes and the g-formula. Appendix B: Parametric g-formula steps and modelling assumptions. Appendix C: Probabilistic bias analysis (PBA). Appendix D: The MET-PREVENT study group members and affiliations. Appendix E: QQ plot and histogram of standardised bootstrap replicates of the controlled direct effect.

Acknowledgements

We thank the participants in the MET-PREVENT trial and members of the Trial Management Group who are not in the co-authorship list.

Authors' contributions

Conception and design: SH, NW, NL, KW, JMSW. Acquisition, analysis: SH, NW, MDW. Interpretation of data: all authors. Drafting: SH. Critical revision of draft: all authors. All authors have approved the submitted version.

Funding

The MET-PREVENT trial was funded by the NIHR Newcastle Biomedical Research Centre. Additional funding for mechanistic analyses was provided as a philanthropic gift by Mr Alan Halsall. MDW, AAS, and CMCD acknowledge support from the NIHR Newcastle Biomedical Research Centre. JMSW is funded by an NIHR Research Professorship (NIHR301614). The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care.

Data availability

Data can be shared upon reasonable request to the NIHR Newcastle Biomedical Research Centre by contacting Professor Miles Witham (Chief Investigator): Miles.Witham@newcastle.ac.uk.

Declarations

Ethical approval and consent to participate

All participants provided their informed consent. The trial was approved by the UK Health Research Authority North-West—Liverpool Central Research Ethics Committee (20/NW/0470) and the UK Medicines and Healthcare products Regulatory Agency (2020–004023-16). The trial was sponsored by the Newcastle Upon Tyne Hospitals NHS Foundation Trust. The trial complied with the ethical standards laid down in the 1964 Declaration of Helsinki and its later amendments.

Consent for publication

Not applicable.

Competing interests

None to declare.

Author details

¹Population Health Sciences Institute, Newcastle University, Newcastle Upon Tyne, UK. ²School of Mathematics, Statistics and Physics, Newcastle University, Newcastle Upon Tyne, UK. ³AGE Research Group, Faculty of Medical Sciences, Translational and Clinical Research Institute, Newcastle University, Newcastle Upon Tyne, UK. ⁴NIHR Newcastle Biomedical Research Centre, Newcastle Upon Tyne Hospitals NHS Foundation Trust, Cumbria, Northumberland, Tyne and Wear NHS Foundation Trust and Newcastle University, Newcastle Upon Tyne, UK. ⁵Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, UK.

Received: 15 February 2025 Accepted: 2 April 2026

Published online: 10 April 2026

References

- Ye CL, Beyene J, Browne G, Thabane L. Estimating treatment effects in randomised controlled trials with non-compliance: a simulation study. *Bmj Open*. 2014;4(6):e005362.
- Kahan BC, Hindley J, Edwards M, Cro S, Morris TP. The estimands framework: a primer on the ICH E9(R1) addendum. *BMJ*. 2024;384:e076316.
- European Medicines Agency. ICH E9 (R1) addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials. 2020. https://www.ema.europa.eu/en/documents/scientific-guideline/ich-e9-r1-addendum-estimands-sensitivity-analysis-clinical-trials-guideline-statistical-principles_en.pdf.
- Vanderweele TJ. Controlled direct and mediated effects: definition, identification and bounds. *Scand J Stat*. 2011;38(3):551–63.
- Igelström E, Craig P, Lewsey J, Lynch J, Pearce A, Katikireddi SV. Causal inference and effect estimation using observational data. *J Epidemiol Commun Health*. 2022;76(11):960–6.

6. Neyman J. Sur les applications de la thar des probabilités aux expériences Agaricales: Essay des principe. [English translation of excerpts by Dabrowska, D., Speed, T.J. *Stat Sci.* 1923;5:463–72.
7. Rubin DB. Estimating causal effects of treatments in randomized and nonrandomized studies. *J Educ Psychol.* 1974;66(5):688–701.
8. VanderWeele TJ. Mediation analysis: a practitioner's guide. *Ann Rev Public Health.* 2016;37(Volume 37, 2016):17–32.
9. Baron RM, Kenny DA. The moderator–mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *J Pers Soc Psychol.* 1986;51:1173–82.
10. Banack HR, Mayeda ER, Naimi A, Fox MP, Whitcomb BW. Collider stratification bias I: principles and structure. *Am J Epidemiol.* 2024;193(2):238–40.
11. Hernán MA, Robins JM. Per-protocol analyses of pragmatic trials. *N Engl J Med.* 2017;377(14):1391–8.
12. Naimi AI, Cole SR, Kennedy EH. An introduction to g methods. *Int J Epidemiol.* 2017;46(2):756–62.
13. Witham MD, McDonald C, Wilson N, Rennie KJ, Bardgett M, Bradley P, et al. Metformin and physical performance in older people with probable sarcopenia and physical frailty or frailty in England (MET-PREVENT): a double-blind, randomised, placebo-controlled trial. *Lancet Healthy Longev.* 2025;6(3):100695.
14. Rennie KJ, Witham M, Bradley P, Clegg A, Connolly S, Hancock HC, et al. MET-PREVENT: metformin to improve physical performance in older people with sarcopenia and physical frailty/frailty - protocol for a double-blind, randomised controlled proof-of-concept trial. *Bmj Open.* 2022;12(7):e061823.
15. Christofides EA. Practical insights into improving adherence to metformin therapy in patients with type 2 diabetes. *Clin Diabetes.* 2019;37(3):234–41.
16. Lee DSU, Lee H. Adherence and persistence rates of major antidiabetic medications: a review. *Diabetol Metab Syndr.* 2022;14(1):12.
17. Smaje A, Weston-Clark M, Raj R, Orlu M, Davis D, Rawle M. Factors associated with medication adherence in older patients: a systematic review. *Aging Med (Milton).* 2018;1(3):254–66.
18. Yap AF, Thirumoorthy T, Kwan YH. Systematic review of the barriers affecting medication adherence in older adults. *Geriatr Gerontol Int.* 2016;16(10):1093–101.
19. Andrews AW, Vallabhajosula S, Boise S, Bohannon RW. Normal gait speed varies by age and sex but not by geographical region: a systematic review. *J Physiother.* 2023;69(1):47–52.
20. Busch TD, Duarte YA, Nunes DP, Lebrao ML, Naslavsky MS, Rodrigues AD, et al. Factors associated with lower gait speed among the elderly living in a developing country: a cross-sectional population-based study. *Bmc Geriatr.* 2015;15:35.
21. Figgins E, Pieruccini-Faria F, Speechley M, Montero-Odasso M. Potentially modifiable risk factors for slow gait in community-dwelling older adults: a systematic review. *Ageing Res Rev.* 2021;66:101253.
22. Han E, Sohn HS, Jang S. Health behavior and medication adherence. *Value Health.* 2014;17(7):A492–A.
23. Buchner DM, Cress ME, Esselman PC, Margherita AJ, deLateur BJ, Campbell AJ, et al. Factors associated with changes in gait speed in older adults. *J Gerontol a-Biol.* 1996;51(6):M297–302.
24. Demakakos P, Cooper R, Hamer M, de Oliveira C, Hardy R, Breeze E. The bidirectional association between depressive symptoms and gait speed: evidence from the English Longitudinal Study of Ageing (ELSA). *Plos One.* 2013;8(7):e68632.
25. Montero-Odasso M, Sarquis-Adamson Y, Song HY, Bray NW, Pieruccini-Faria F, Speechley M. Polypharmacy, gait performance, and falls in community-dwelling older adults. Results from the gait and brain study. *J Am Geriatr Soc.* 2019;67(6):1182–8.
26. Brown EG, Wood L, Wood S. The medical dictionary for regulatory activities (MedDRA). *Drug Saf.* 1999;20(2):109–17.
27. Keil AP, Edwards JK, Richardson DB, Naimi AI, Cole SR. The parametric g-formula for time-to-event data. *Epidemiology.* 2014;25(6):889–97.
28. Tackney MS, Morris T, White I, Leyrat C, Diaz-Ordaz K, Williamson E. A comparison of covariate adjustment approaches under model misspecification in individually randomized trials. *Trials.* 2023;24(1):14.
29. Lin SH, Young J, Logan R, Tchetgen EJT, VanderWeele TJ. Parametric mediational g-formula approach to mediation analysis with time-varying exposures, mediators, and confounders. *Epidemiology.* 2017;28(2):266–74.
30. Rudolph JE, Cartus A, Bodnar LM, Schisterman EF, Naimi AI. The role of the natural course in causal analysis. *Am J Epidemiol.* 2022;191(2):341–8.
31. Fox MP, MacLehose RF, Lash TL. Probabilistic bias analysis for simulation of record-level data. In: *Applying quantitative bias analysis to epidemiologic data.* Switzerland: Springer Cham. 2nd. 2021:291–327.
32. Hunnicutt JN, Ulbricht CM, Chrysanthopoulou SA, Lapane KL. Probabilistic bias analysis in pharmacoepidemiology and comparative effectiveness research: a systematic review. *Pharmacoepidem Dr S.* 2016;25(12):1343–53.
33. Hawkins AT, Samuels LR. Use of confidence intervals in interpreting nonstatistically significant results. *JAMA J Am Med Assoc.* 2021;326(20):2068–9.
34. Loeys T, Moerkerke B, Raes A, Rosseel Y, Vansteelandt S. Estimation of controlled direct effects in the presence of exposure-induced confounding and latent variables. *Struct Equ Modeling.* 2014;21(3):396–407.
35. Chatton A, Rohrer JM. The causal cookbook: recipes for propensity scores, g-computation, and doubly robust standardization. *Adv Meth Pract Psych.* 2024;7(1):25152459241236149.
36. Atkinson HH, Rapp SR, Williamson JD, Lovato J, Absher JR, Gass M, et al. The relationship between cognitive function and physical performance in older women: results from the women's health initiative memory study. *J Gerontol a-Biol.* 2010;65(3):300–6.
37. Watson NL, Rosano C, Boudreau RM, Simonsick EM, Ferrucci L, Sutton-Tyrrell K, et al. Executive function, memory, and gait speed decline in well-functioning older adults. *J Gerontol a-Biol.* 2010;65(10):1093–100.
38. Garcia-Pinillos F, Cozar-Barba M, Munoz-Jimenez M, Soto-Hermoso V, Latorre-Roman P. Gait speed in older people: an easy test for detecting cognitive impairment, functional independence, and health state. *Psychogeriatrics.* 2016;16(3):165–71.
39. Rodakowski J, Skidmore ER, Reynolds CF, Dew MA, Butters MA, Holm MB, et al. Can performance on daily activities discriminate between older adults with normal cognitive function and those with mild cognitive impairment? *J Am Geriatr Soc.* 2014;62(7):1347–52.
40. VanderWeele TJ, Ding P. Sensitivity analysis in observational research: introducing the E-value. *Ann Intern Med.* 2017;167(4):268–74.
41. Chatton A, Le Borgne F, Leyrat C, Gillaizeau F, Rousseau C, Barbin L, et al. G-computation, propensity score-based methods, and targeted maximum likelihood estimator for causal inference with different covariates sets: a comparative simulation study. *Sci Rep-Uk.* 2020;10(1):9219.
42. Young JG, Cain LE, Robins JM, O'Reilly EJ, Hernan MA. Comparative effectiveness of dynamic treatment regimes: an application of the parametric g-formula. *Stat Biosci.* 2011;3(1):119–43.
43. Bodmer M, Meier C, Krahenbuhl S, Jick SS, Meier CR. Metformin, sulfonylureas, or other antidiabetes drugs and the risk of lactic acidosis or hypoglycemia: a nested case-control analysis. *Diabetes Care.* 2008;31(11):2086–91.
44. Christiansen CF, Ehrenstein V, Heide-Jorgensen U, Skovbo S, Norrelund H, Sorensen HT, et al. Metformin initiation and renal impairment: a cohort study in Denmark and the UK. *BMJ Open.* 2015;5(9):e008531.
45. Salpeter SR, Greyber E, Pasternak GA, Salpeter EE. Risk of fatal and nonfatal lactic acidosis with metformin use in type 2 diabetes mellitus. *Cochrane Db Syst Rev.* 2010(4):CD002967.
46. Naimi AI, Perkins NJ, Sjaarda LA, Mumford SL, Platt RW, Silver RM, et al. The effect of preconception-initiated low-dose aspirin on human chorionic gonadotropin-detected pregnancy, pregnancy loss, and live birth per protocol analysis of a randomized trial. *Ann Intern Med.* 2021;174(5):595–+.
47. Bonnet F, Scheen A. Understanding and overcoming metformin gastrointestinal intolerance. *Diabetes Obes Metab.* 2017;19(4):473–81.
48. Michnevich T, Pan Y, Hendi A, Oechsle K, Stein A, Nestoriuc Y. Preventing adverse events of chemotherapy for gastrointestinal cancer by educating patients about the nocebo effect: a randomized-controlled trial. *BMC Cancer.* 2022;22(1):1008.
49. Flory JH, Keating S, Guelce D, Mushlin AI. Overcoming barriers to the use of metformin: patient and provider perspectives. *Patient Prefer Adher.* 2019;13:1433–41.
50. Landau S, Stahl D. Sample size and power calculations for medical studies by simulation when closed form expressions are not available. *Stat Methods Med Res.* 2013;22(3):324–45.

51. Rudolph KE, Goin DE, Stuart EA. The peril of power: a tutorial on using simulation to better understand when and how we can estimate mediating effects. *Am J Epidemiol.* 2020;189(12):1559–67.
52. Frangakis CE, Rubin DB. Principal stratification in causal inference. *Biometrics.* 2002;58(1):21–9.
53. Imbens GW, Angrist JD. Identification and estimation of local average treatment effects. *Econometrica.* 1994;62(2):467–75.
54. Drury T, Bartlett JW, Wright D, Keene ON. The estimand framework and causal inference: complementary not competing paradigms. *Pharm Stat.* 2025;24(5):e70035.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.