RESEARCH ARTICLE

WILEY **Genetic Epidemiology**

OFFICIAL JOURNAL
INTERNATIONAL GENETIC
EPIDEMIOLOGY SOCIETY
www.geneticepi.org

Semiparametric methods for estimation of a nonlinear exposure-outcome relationship using instrumental variables with application to Mendelian randomization

James R. Staley | **Stephen Burgess** (iD)

Strangeways Research Laboratory, Department of Public Health and Primary Care, Cardiovascular Epidemiology Unit, University of Cambridge, United Kingdom

Correspondence
Stephen Burgess, Strangeways Research Laboratory, Department of Public Health and Primary Care, Cardiovascular Epidemiology Unit, University of Cambridge, Wort's Causeway, Cambridge CB1 8RN, UK.
Email: sb452@medschl.cam.ac.uk

ABSTRACT

Mendelian randomization, the use of genetic variants as instrumental variables (IV), can test for and estimate the causal effect of an exposure on an outcome. Most IV methods assume that the function relating the exposure to the expected value of the outcome (the exposure-outcome relationship) is linear. However, in practice, this assumption may not hold. Indeed, often the primary question of interest is to assess the shape of this relationship. We present two novel IV methods for investigating the shape of the exposure-outcome relationship: a fractional polynomial method and a piecewise linear method. We divide the population into strata using the exposure distribution, and estimate a causal effect, referred to as a localized average causal effect (LACE), in each stratum of population. The fractional polynomial method performs metaregression on these LACE estimates. The piecewise linear method estimates a continuous piecewise linear function, the gradient of which is the LACE estimate in each stratum. Both methods were demonstrated in a simulation study to estimate the true exposure-outcome relationship well, particularly when the relationship was a fractional polynomial (for the fractional polynomial method) or was piecewise linear (for the piecewise linear method). The methods were used to investigate the shape of relationship of body mass index with systolic blood pressure and diastolic blood pressure.

KEYWORDS
causal effects, fractional polynomials, genetic variants, piecewise linear models, UK Biobank

1 | INTRODUCTION

Often the shape of association between an exposure and an outcome is nonlinear. For example, the observed association between body mass index (BMI) and all-cause mortality in a Western context is J-shaped (or U-shaped), as risk of mortality is increased for individuals at both ends of the BMI distribution (Flegal, Kit, Orpana, & Graubard, 2013). However, particularly for underweight individuals, this could reflect either reverse causality or confounding, rather than a true causal effect of low BMI increasing mortality risk. Instrumental variable (IV) methods can be used to distinguish between correlation and causation. However, these methods typically assume that the exposure-outcome relationship is linear when estimating a causal effect (Hernán & Robins, 2006). In many cases, investigating the shape of the exposure-outcome relationship is the primary aim of a study. This can be used to define treatment thresholds for pharmaceutical interventions or health guidelines.

A natural way of tackling the nonlinearity problem in IV analysis is to perform a two-stage analysis similar to the well-known two-stage least squares method, except fitting a

nonlinear function in the second stage (Horowitz, 2011; Newey & Powell, 2003). However, this approach requires the instrument and any covariates included in the first-stage model to explain a large proportion of variance in the exposure, as information for assessing the shape of relationship between the exposure and outcome will only be available for the fitted values of the exposure from the first-stage regression. If the proportion of variance in the exposure explained by the IV is small, then observing nonlinearity for this limited range of values is unlikely. In Mendelian randomization, the use of genetic variants as IV, genetic variants typically only explain a small percentage of the variance in the exposure (usually in the region of 1–4%; Burgess & Thompson, 2015; Ebrahim & Smith, 2008).

Two approaches for addressing nonlinearity in the context of Mendelian randomization have recently been proposed (Burgess, Davies, & Thompson, 2014; Silverwood et al., 2014). Burgess et al. (2014) assessed the consequences of performing a linear IV analysis when the exposure-outcome relationship truly was nonlinear, as well as stratifying individuals using the exposure distribution to obtain IV estimates, referred to as localized average causal effects (LACE), in each stratum. Silverwood et al. (2014) performed metaregression of LACE estimates across strata to examine whether a quadratic rather than a linear model was a better fit for relationships between alcohol consumption and a variety of cardiovascular markers.

In this paper, we present two novel semiparametric methods for investigating the shape of the exposure-outcome relationship using IV analysis developed for use in Mendelian randomization. The first is based on fractional polynomials (Royston & Altman, 1994; Royston, Ambler, & Sauerbrei, 1999), whereas the second fits a piecewise linear function. We also propose a test for nonlinearity based on the fractional polynomial method, and assess the impact of varying the number of strata of the exposure distribution used to test for nonlinearity and to estimate nonlinear relationships. We illustrate the methods using data from UK Biobank (Sudlow et al., 2015), a large UK-based cohort, to investigate the shape of the relationship between BMI and blood pressure using Mendelian randomization.

2 | METHODS

2.1 | Stratifying on the IV-free exposure

We define the exposure-outcome relationship as the function relating the exposure to the expected value of the outcome. We initially assume that this function is homogeneous for all individuals in the population, and return to its interpretation in case of heterogeneity in the discussion.

To assess the shape of association between exposure X and outcome Y using a single instrument G, we first stratify the population using the exposure distribution. If we were to strat-ify on the exposure directly, then an association between the IV and outcome might be induced even if it were not present in the original data, thus invalidating the IV assumptions (Didelez & Sheehan, 2007). This problem can be avoided by instead stratifying on the residual variation in the exposure after conditioning on the IV, assuming that the effect of the IV on the exposure is linear and constant for all individuals across the entire of the exposure distribution (Burgess et al., 2014). In econometrics, this residual is known as a control function (Arellano, 2003). We calculate this residual by performing linear regression of the exposure on the IV, and then setting the value of the IV to 0. We refer to this as the IV-free exposure. It is the expected value of the exposure that would be observed if the individual had an IV value of 0, and can be interpreted as the nongenetic component of the exposure.

In each stratum of the IV-free exposure, we estimate the LACE as a ratio of coefficients: the IV association with the outcome divided by the IV association with the exposure. The assumption that the effect of the IV on the exposure is constant is a stronger version of the monotonicity assumption (Angrist, Imbens, & Rubin, 1996), and hence the LACE are local average treatment effects (also called complier-averaged causal effects; Yau & Little, 2001) for each stratum (Imbens & Angrist, 1994). We then proceed to estimate the exposure-outcome relationship from these LACE estimates using two approaches: the first based on fractional polynomials, and the second a piecewise linear function.

2.2 | Fractional polynomial method

The fractional polynomial method consists of metaregression of the LACE estimates against the mean of the exposure in each stratum in a flexible semiparametric framework (Bagnardi, Zambon, Quatto, & Corrao, 2004; Thompson & Sharp, 1999). Fractional polynomials are a family of functions that can be used to fit complex relationships for a single covariate (Royston & Altman, 1994). The standard powers used when modeling using fractional polynomials are $P = \{-2, -1, -0.5, 0, 0.5, 1, 2, 3\}$, where the power of 0 refers to the (natural) log function. These powers are used throughout this paper. Fractional polynomials of degree 1 are defined as

$$f(x) = \beta_0 + \beta_1 x^p, \tag{1}$$

where $p \in P$. Similarly, fractional polynomials of degree 2 are defined as

$$f(x) = \begin{cases} \beta_0 + \beta_1 x^{p_1} + \beta_2 x^{p_2} & \text{if } p_1 \neq p_2 \\ \beta_0 + \beta_1 x^p + \beta_2 x^p \log(x) & \text{if } p_1 = p_2 = p, \end{cases} \tag{2}$$

where $p_1, p_2 \in P$. In both cases, x^0 is interpreted as $\log(x)$. As fractional polynomials of degree larger than 2 are rarely required in practice, these were not considered in this paper (Royston & Altman, 1994). Because a causal effect is an esti-mate of the derivative of the exposure-outcome relationship

(Small, 2014), we fit the LACE estimates using the derivative of the fractional polynomial function (from either (1) or (2)).

The method proceeds as follows. First, we calculate the IV-free exposure, and stratify the population based on quantiles of its distribution. Second, the LACE estimate is calculated in each stratum as a ratio of coefficients (the LACE estimate for stratum k is $\frac{\hat{\beta}_{Y|G,k}}{\hat{\beta}_{X|G}}$, where $\hat{\beta}_{Y|G,k}$ is the estimated association of the IV with the outcome in stratum k and $\hat{\beta}_{X|G}$ is the estimated association of the IV with the exposure in the whole population), and the standard error of the LACE estimate is computed as $\frac{se(\hat{\beta}_{Y|G,k})}{\hat{\beta}_{X|G}}$ (the first term of the delta method approximation; Thomas, Lawlor, & Thompson, 2007). Third, these LACE estimates are metaregressed against the mean of the exposure in each stratum using the derivative of the fractional polynomial function as the model relating the LACE estimates to the exposure values. The original fractional polynomial function then represents the exposure-outcome relationship. As this function is constructed from the LACE estimates, the intercept of the exposure-outcome curve cannot be estimated and must be set arbitrarily. If it is set to 0 at a reference value (for instance, the mean of the exposure distribution), then the value of the function represents the expected difference in the outcome compared with this reference value when the exposure is set to different values.

Confidence intervals (CIs) for the exposure-outcome curve can be computed arithmetically under a normal assumption either using the estimated standard errors from the metaregression or by bootstrapping the second and third steps from above (we maintain the strata and estimate of the IV on the exposure as in the original data, and estimate the associations of the IV with the outcome in bootstrapped samples for each stratum).

To explore a range of possible parametric forms, we fit all possible fractional polynomial models of degrees 1 and 2, and select the best-fitting one based on the likelihood. A fractional polynomial of degree 2 is preferred over one of degree 1 if the twice the difference in the log-likelihood is greater than the 95th percentile point of a χ_2^2 distribution for the best-fitting fractional polynomial in each class (Royston & Altman, 1994).

2.3 | Piecewise linear method

Another way of estimating the exposure-outcome relationship is to use a piecewise linear approach. The exposure-outcome relationship is estimated as piecewise linear function with each stratum contributing a line segment whose gradient is the LACE estimate for that stratum. The function is constrained to be continuous, so that each line segment begins where the previous segment finished. As in the fractional polynomial method, although the intercept for each line segment is fixed by the previous line segment, the overall intercept of the exposure-outcome curve cannot be estimated and must be set arbitrarily.

CIs are estimated by bootstrapping the IV associations with the outcome as in the fractional polynomial approach. For a 95% CI, the piecewise linear method is performed for each bootstrapped dataset, and then the 2.5th and 97.5th percentiles of the function are taken at selected points across the exposure distribution; we chose the mean exposure values in each of the strata.

2.4 | Tests of nonlinearity

There are already two proposed methods in this framework for testing whether a nonlinear exposure-outcome model fits the data better than a linear model. The first is a heterogeneity test using Cochran's Q statistic to assess whether the LACE estimates differ more than would be expected by chance. The second is a trend test where the LACE estimates are metaregressed against the mean value of the exposure in each strata; this is equivalent to fitting a quadratic exposure-outcome model.

A more flexible version of this method is to test the best-fitting fractional polynomial of degree 1 against the linear model. This can be achieved by comparing twice the difference in the log-likelihood between the linear model and the best-fitting fractional polynomial of degree 1 with a χ_1^2 distribution.

These methods are available in the nlmr R package (https://github.com/jrs95/nlmr).

3 | SIMULATION STUDY

To assess the performance of these methods in realistic scenarios for Mendelian randomization, we performed a simulation study. We simulated data for 10,000 individuals for an IV G, a continuous exposure X that takes only positive values, a continuous outcome Y, and a confounder U (assumed to be unmeasured). The data-generating model for individual i is

$$x_i = 2 + 0.25g_i + u_i + \epsilon_{Xi}$$
$$y_i = h(x_i) + 0.8u_i + \epsilon_{Yi},$$

where $g_i \sim \text{Bin}(2, 0.3)$, $\epsilon_{Xi} \sim \text{Exp}(1)$, $u_i \sim \text{Unif}(0,1)$, $\epsilon_{Yi} \sim N(0, 1)$, and $h(x_i)$ is the function relating the exposure to the outcome (the exposure-outcome relationship). Exposure values were taken to be positive and away from zero so that the outcome takes sensible values for log and negative power functions. The IV explains 2.6% of the variance in the exposure.

3.1 | Choice of exposure-outcome model

For the fractional polynomial method, all possible fractional polynomials of degrees 1 and 2 were considered as the functional form of the exposure-outcome relationship.

Combinations of effect sizes for the β parameters were chosen ranging from 0 to 2. For fractional polynomials of degree 2, we also considered effects in opposing directions for β_1 and β_2; these simulations yielded similar results to those discussed here (results not shown). Fixed-effects metaregression was used in the simulations, however, random-effects metaregression yielded similar results (results not shown).

For the piecewise linear method and comparisons between methods, linear, quadratic, square-root, and logarithmic functions were considered as the functional form of the exposure-outcome relationship, as well as a threshold model:

$$h(x_i) = \begin{cases} 0 & \text{if } x_i \leq 3.65 \\ \beta x_i & \text{if } x_i > 3.65. \end{cases}$$

3.2 | Evaluating the performance of the methods

To evaluate the fractional polynomial method, we first fitted the correct fractional polynomial model (i.e., with the correct degree and powers) and assessed the bias and coverage of the effect parameter estimates. Subsequently, we fitted all fractional polynomials of the same degree and selected the best-fitting polynomial based on the likelihood. We assessed the proportion of simulations where the best-fitting fractional polynomial was the correct fractional polynomial. If the correct fractional polynomial was not the best-fitting fractional polynomial, we tested whether it was in the group of fractional polynomials that fit the data almost as well as the best-fitting polynomial; defined as those fractional polynomials where twice the difference in the log-likelihood (compared with the best-fitting polynomial) was less than the 90th percentile point of a χ_m^2 distribution, where $m = 1$ for comparing fractional polynomials of degree 1 and $m = 2$ for comparing polynomials of degree 2.

To evaluate the piecewise linear method, we first compared the outcome estimates at the mean exposure value in each quantile to the values of the true model at the same points. The coverages of the bootstrapped 95% CIs were also evaluated at these points.

For comparing the fit of the fractional polynomial and piecewise linear models, we used the following heuristic function:

$$\sum_{k=1}^{K} \left| \hat{\bar{y}}_k - \bar{y}_k \right|, \tag{3}$$

where summation is across the K quantile groups, and $\bar{y}_k$ is the expected value of the outcome evaluated at the mean value of the exposure in each quantile group.

3.3 | Varying the number of strata

In the initial simulations, the population was split based on the IV-free exposure into decile groups. Further simulations

were performed varying the number of strata using 5, 10, 50, and 100 quantile groups. Tests of nonlinearity were performed to assess the impact of the number of strata on the empirical power of each test. The empirical power of each test was reported as the proportion of simulation replicates with P-value less than 0.05. The heuristic function (3) was calculated based on 10 deciles for each number of strata.

For each simulation and set of parameters, 500 replications were performed. Bootstrap 95% CIs were generated using 500 bootstrap samples. All analyses were performed using R version 3.0.2.

3.4 | Additional simulations to assess impact of violations of assumptions

We performed additional simulations in which the underlying assumptions that the effect of the IV on the exposure and the effect of the exposure on the outcome are fixed and independent were relaxed. In these simulations, we assessed both modeling assumptions by allowing the effect of the IV on the exposure to vary (by drawing the effect parameter from a normal distribution $N(0.25, 0.1^2)$ for each individual in the population), and allowing the exposure-outcome relationship to vary (by drawing the causal parameter from a normal distribution $N(\beta, 0.2^2)$ for each individual in the population). We assessed the impact of allowing each of these parameters to vary separately and both to vary together. In addition, we also allowed variation in both parameters to be correlated by drawing the parameters from a bivariate normal distribution with correlation 0.2. For fractional polynomials of degree 2, only the causal parameter for the second polynomial was allowed to vary across individuals.

We also performed further simulations using a low-frequency genetic variant having a large effect on the exposure (minor allele frequency = 0.03, linear effect on the exposure = 0.75), and using the original genetic variant but having a superadditive (first allele increases exposure by 0.1 units, second by 0.3 units) and a subadditive (first allele increases exposure by 0.3 units, second by 0.1 units) effect on the exposure in the data-generating model.

4 | RESULTS

4.1 | Fractional polynomial method

Comparisons of fractional polynomials for all powers are provided in Table S1 (degree 1) and Table S2 (degree 2); a summary of results for the most commonly encountered powers is given in Table 1.

For fractional polynomials of degree 1, when fitting the correct fractional polynomial model, the causal estimate was

TABLE 1 Simulation results for the fractional polynomial method

(a) Degree 1

p	β	Fitting correct FP		Fitting all FPs	
		Mean (SD) [Mean SE]	Coverage	Powers	
p_1	β_1	$\hat{\beta}_1$	$\hat{\beta}_1$	Correct	Set
0	0	−0.01 (0.22) [0.21]	0.934	-	-
0	1	0.98 (0.22) [0.21]	0.944	0.172	0.918
0	2	1.98 (0.21) [0.21]	0.954	0.386	0.910
0.5	0	0.00 (0.25) [0.23]	0.930	-	-
0.5	1	1.00 (0.25) [0.23]	0.936	0.194	0.892
0.5	2	1.99 (0.24) [0.24]	0.932	0.340	0.904
1	0	0.00 (0.06) [0.06]	0.938	-	-
1	1	1.00 (0.07) [0.07]	0.944	0.748	0.938
1	2	2.00 (0.07) [0.07]	0.938	0.912	0.958
2	0	0.00 (0.01) [0.01]	0.942	-	-
2	1	1.02 (0.01) [0.01]	0.756	1.000	1.000
2	2	2.03 (0.02) [0.02]	0.436	1.000	1.000

(b) Degree 2

p		β		Fitting correct FP		Coverage		Fitting all FPs	
				Mean (SD) [Mean SE]				Powers	
p_1	p_2	β_1	β_2	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$	Correct	Set
0	0.5	0	0	−0.04 (1.98) [1.95]	0.06 (2.20) [2.17]	0.958	0.960	-	-
0	0.5	1	2	1.00 (2.11) [2.03]	1.98 (2.35) [2.26]	0.938	0.934	0.004	0.964
0	0.5	2	1	1.95 (1.99) [2.02]	1.02 (2.23) [2.24]	0.952	0.956	0.002	0.968
0	1	0	0	−0.01 (0.95) [0.94]	0.00 (0.28) [0.28]	0.950	0.956	-	-
0	1	1	2	1.00 (1.15) [1.17]	2.00 (0.35) [0.36]	0.940	0.940	0.012	0.948
0	1	2	1	2.03 (1.09) [1.08]	0.98 (0.33) [0.33]	0.938	0.938	0.016	0.956
0	2	0	0	−0.01 (0.48) [0.45]	0.00 (0.02) [0.02]	0.940	0.936	-	-
0	2	1	2	1.30 (1.25) [1.09]	2.02 (0.06) [0.06]	0.898	0.902	0.026	0.954
0	2	2	1	2.12 (0.84) [0.81]	1.01 (0.04) [0.04]	0.938	0.926	0.040	0.964
0.5	1	0	0	−0.05 (2.06) [1.96]	0.01 (0.55) [0.53]	0.938	0.944	-	-
0.5	1	1	2	1.03 (2.60) [2.56]	2.00 (0.72) [0.71]	0.940	0.938	0.010	0.954
0.5	1	2	1	1.86 (2.27) [2.34]	1.03 (0.62) [0.60]	0.950	0.950	0.006	0.960
0.5	2	0	0	−0.06 (0.65) [0.63]	0.00 (0.02) [0.02]	0.928	0.946	-	-
0.5	2	1	2	1.48 (1.78) [1.63]	2.01 (0.08) [0.07]	0.936	0.934	0.024	0.970
0.5	2	2	1	2.18 (1.25) [1.19]	1.01 (0.05) [0.05]	0.924	0.936	0.036	0.948
1	2	0	0	0.01 (0.24) [0.24]	0.00 (0.03) [0.03]	0.952	0.954	-	-
1	2	1	2	1.23 (0.78) [0.70]	2.00 (0.12) [0.11]	0.912	0.908	0.030	0.960
1	2	2	1	2.08 (0.57) [0.54]	1.00 (0.09) [0.08]	0.942	0.926	0.108	0.962

Notes. Results for all the fractional polynomials of degree 1 (all effect sizes) and degree 2 (β_1=1 and β_2=2) are presented in Tables S1 and S2; this table is a summary of results for the most commonly encountered powers. p are the powers and β are the effect parameters. Coverage refers to the number of replications where the true value of β was contained within the corresponding 95% CI. The power(s) was correctly chosen (Correct) if the best-fitting fractional polynomial was also the correct fractional polynomial, whil the correct model was within the set of powers that fit the data equally, as well as the best-fitting fractional polynomial (Set) if the difference between twice the log-likelihood for the correct model and the best-fitting model was less than the 90th percentile of the relevant χ^2 distribution. SD, standard deviation; SE, standard error; FP, fractional polynomial; CI, confidence interval.

generally unbiased (Table 1). Coverage estimates were close to the nominal 95% rate, except for fractional polynomials of power 2 (and power 3; Table S1), where causal estimates were slightly biased, and this small bias led to undercoverage. However, under the null, causal estimates were unbiased and correct coverage rates were maintained. For fractional polynomials of degree 2, a similar pattern was observed except that small biases and resulting undercoverage was more common, although correct coverage rate under the null was always maintained.

TABLE 2 Simulation results for the piecewise linear method

Model	β	Parameter	Decile of the IV-free exposure distribution										Heuristic	
			1	2	3	4	5	6	7	8	9	10	FP	PL
Linear	0.5	Correct	0.201	0.357	0.460	0.550	0.641	0.743	0.870	1.040	1.302	1.995		
		Mean	0.197	0.352	0.454	0.543	0.633	0.736	0.862	1.032	1.296	1.961	1.12	1.20
		Coverage	0.958	0.964	0.962	0.962	0.958	0.952	0.940	0.944	0.950	0.952	(0.74)	(0.68)
Quadratic	0.5	Correct	0.891	1.697	2.281	2.826	3.409	4.107	5.027	6.360	8.644	16.002		
		Mean	1.011	1.831	2.426	2.982	3.572	4.281	5.214	6.568	8.891	16.617	1.07	2.94
		Coverage	0.680	0.784	0.786	0.786	0.772	0.758	0.764	0.762	0.750	0.922	(0.77)	(1.43)
Threshold	0.5	Correct	0.000	0.000	0.000	0.000	0.000	0.000	0.054	0.224	0.486	1.179		
		Mean	0.001	0.000	0.001	0.003	0.005	0.020	0.088	0.239	0.499	1.171	1.35	1.25
		Coverage	0.960	0.964	0.952	0.950	0.942	0.930	0.934	0.930	0.940	0.946	(0.46)	(0.73)
Threshold	1	Correct	0.000	0.000	0.000	0.000	0.000	0.000	0.108	0.448	0.971	2.359		
		Mean	0.008	0.012	0.013	0.012	0.014	0.041	0.181	0.493	1.017	2.393	2.31	1.33
		Coverage	0.946	0.940	0.936	0.922	0.916	0.904	0.904	0.916	0.916	0.958	(0.42)	(0.79)

Notes. β is the effect parameter. Mean is the mean value of the outcome at the mean value of the exposure in the deciles of the IV-free distribution. Coverage refers to the number of replications where the correct value of the outcome at the mean value of the exposure in the decile of the IV-free distribution was contained within the corresponding 95% prediction interval. The heuristic statistic (mean (SD) across simulations) is the sum of the absolute values of the predicted value of the outcome minus the correct value of the outcome at the mean value of the exposure in deciles of the IV-free distribution. FP, fractional polynomial; PL, piecewise linear; IV, instrumental variable.

When fitting all the fractional polynomial models, the correct fractional polynomial model was fitted more often for a fractional polynomial of degree 1, and when the power of the fractional polynomial differed substantially from 0. Although the power to detect the correct functional form was low for the logarithmic and square-root functions, this is to some extent an artifact of the choice of powers; if a basis with only one concave function (either a logarithmic or a square-root function) were used, then the correct model would be chosen more often. As the causal parameter in the model increased, the correct model was chosen more frequently. However, in all cases, the correct fractional polynomial was in the set of best-fitting fractional polynomials in at least 89% of simulations. The fractional polynomial test for nonlinearity rejects the null exactly when the linear model is not in the set of best-fitting fractional polynomials. With a null causal effect ($\beta = 0$), the probability of fitting the "correct" fractional polynomial was not estimated as all fractional polynomials with zero coefficients would describe the data equally well. In reality, the true exposure-outcome relationship is unlikely to have an exact functional form, so the ability to estimate the shape of the relationship is more important that the precise identification of the function.

4.2 | Piecewise linear method

The piecewise linear method performed well when the true model was piecewise linear (such as a linear or a threshold relationship), with the predicted mean values of the outcome similar to their true values at the mean value of the exposure within each decile of the IV-free exposure (Table 2). The

bootstrapped CIs also had approximately 95% coverage at these points, except for the quantiles at or either side of the point of inflection of the threshold model. However, when the true model was not piecewise linear (in particular, for a quadratic relationship), estimates were biased and coverage was below nominal levels.

Using the heuristic function (3) to compare between the estimates for the best-fitting fractional polynomial and the piecewise linear model, the models performed similarly under a linear model. For a quadratic model, the fractional polynomial method outperformed the piecewise linear method, whereas the opposite was true for a threshold model. This is unsurprising, as the fractional polynomial method performed best when the true model was a polynomial, and likewise for the piecewise linear method when the model was piecewise linear.

4.3 | Varying the number of strata

The best-fitting fractional polynomial method had a similar or slightly better model fit (judged by the heuristic function) when a greater number of strata were used (Table 3). However, the piecewise linear method fitted the data better when fewer strata were used. Although the fractional polynomial method ensures that the estimate of the exposure-outcome relationship is a smooth function regardless of the number of strata, the estimate from the piecewise linear method becomes increasingly jagged as the number of strata increases.

The coverage under the null (i.e., a linear model) was not overly inflated for any of the tests. In general, the

TABLE 3 Varying the number of strata and tests of nonlinearity

Model	β	Number of strata	Heuristic		Power of test		
			FP method	PL method	Quad	Q	FP test
Linear	1	5	1.18 (0.88)	1.24 (0.70)	0.076	0.054	0.050
		10	1.11 (0.86)	1.30 (0.76)	0.074	0.040	0.046
		50	1.06 (0.84)	1.48 (0.93)	0.064	0.064	0.040
		100	1.08 (0.85)	1.55 (0.97)	0.062	0.062	0.036
Logarithm	2	5	1.34 (0.79)	1.35 (0.76)	0.486	0.342	0.502
		10	1.31 (0.79)	1.37 (0.78)	0.488	0.264	0.518
		50	1.30 (0.80)	1.56 (0.91)	0.504	0.164	0.544
		100	1.31 (0.80)	1.62 (0.96)	0.504	0.124	0.530
Square root	2	5	1.21 (0.80)	1.23 (0.73)	0.166	0.102	0.170
		10	1.22 (0.80)	1.33 (0.78)	0.156	0.084	0.166
		50	1.21 (0.78)	1.56 (0.91)	0.164	0.072	0.176
		100	1.20 (0.77)	1.63 (0.96)	0.164	0.104	0.178
Quadratic	0.1	5	1.10 (0.77)	1.37 (0.74)	0.618	0.422	0.608
		10	1.03 (0.76)	1.42 (0.77)	0.710	0.392	0.674
		50	0.90 (0.74)	1.52 (0.82)	0.830	0.226	0.774
		100	0.87 (0.72)	1.59 (0.88)	0.874	0.186	0.818
Threshold	0.5	5	1.37 (0.47)	1.20 (0.71)	0.868	0.816	0.804
		10	1.35 (0.44)	1.28 (0.76)	0.862	0.698	0.778
		50	1.36 (0.46)	1.48 (0.89)	0.864	0.364	0.758
		100	1.38 (0.50)	1.58 (0.97)	0.862	0.284	0.742

Notes. β is the effect parameter. The heuristic statistic (mean (SD) across simulations) is the sum of the absolute values of the predicted value of the outcome minus the correct value of the outcome at the mean value of the exposure in deciles of the IV-free distribution. The heuristic measure for the fractional polynomial model was from the best-fitting fractional polynomial for the threshold model. SD, standard deviation; SE, standard error; Quad, quadratic test for assessing nonlinearity; Q, Cochran Q test; FP, fractional polynomial; PL, piecewise linear; IV, instrumental variable.

fractional polynomial and quadratic tests were more powerful than the Cochran Q test across the simulations. The power of the Cochran Q test also decreased as the number of strata increased, whereas the power of the other tests either remained the same or increased. The quadratic test slightly outperformed the fractional polynomial test when the true model was a quadratic or a threshold model; the fractional polynomial test was slightly superior when the true model was a logarithm or a square-root model.

4.4 | Additional simulations to assess impact of violations of assumptions

In the simulations where we relaxed the assumptions that the IV–exposure and the exposure-outcome effects are the same for all individuals, we found that the fractional polynomial models of degree 1 and the piecewise linear method both performed well in terms of bias and coverage (Tables S3 and S4). The only concern was that tests of nonlinearity had slightly inflated Type I error rates when the IV-exposure and exposure-outcome effects were varied in a correlated way;

Type I error rate inflation was not observed when the effects were varied either separately or independently.

In the simulations with a low-frequency genetic variant having a large effect on the exposure (Tables S5 and S6), there was some bias in estimates. This is likely to be the result of weak instrument bias (Burgess & Thompson, 2011): with a low-frequency variant, the variation in instrument strength between the strata is much larger, and so the chances of weak instrument bias affecting the results in specific strata are increased. However, nominal Type I error rates for tests of nonlinearity were maintained. With a superadditive or subadditive model for the genetic association with the exposure (Tables S5 and S6), estimates of the causal parameter in the fractional polynomial method were unbiased, but the power to detect nonlinearity was somewhat reduced. Estimates in the piecewise linear method for a threshold exposure-outcome relationship were somewhat biased. This finding is consistent with previous work on measurement error in the independent variable: here, the IV-free exposure is estimated with error resulting from misspecification of the genetic association with the exposure (see Section 6).

5 | APPLICATION OF METHODS TO THE RELATIONSHIP BETWEEN BMI AND BLOOD PRESSURE IN UK BIOBANK

We illustrate the methods proposed in this paper in an applied example considering the shape of relationship between BMI and blood pressure in the UK Biobank study. The observational relationship between BMI and blood pressure has been investigated previously in a variety of contexts: populations of lean individuals (Kaufman et al., 1997), Danish adolescents (Nielsen & Andersen, 2003), and Iranian adolescents (Hosseini et al., 2010). The relationship has been demonstrated to be monotonically increasing, with inconclusive evidence for or against nonlinearity due to limited sample sizes.

UK Biobank is a prospective cohort study of 502,682 participants recruited at 22 assessment centers across the United Kingdom between 2006 and 2010 (Sudlow et al., 2015). Participants were aged between 40 and 69 at baseline. Extensive health, lifestyle, biological, and genetic measurements were taken on all participants. At the time of writing this paper, genetic information was only available for 133,687 individuals of European ancestry. For individuals on antihypertensive medication, 15/10 mmHg were added to their SBP/DBP (where SBP is systolic blood pressure and DBP is diastolic blood pressure) measurement, respectively. A sensitivity analysis was performed in individuals who had no history of hypertension.

To create an allele score (also called a genetic risk score) of variants related to BMI to be used as an IV, we extracted the 97 variants previously associated with BMI at a genome-wide level of significance by the GIANT consortium (Locke et al., 2015). A proxy variant (rs751414; $r^2 = 0.99$) was used instead of rs2033529, as this variant was not available in UK Biobank; the linkage disequilibrium information was calculated using the European samples from 1000 Genomes (Abecasis et al., 2012). All of the variants were either directly genotyped or well-imputed (INFO > 0.9). The allele score for each individual was computed by multiplying the number of BMI-increasing alleles for each variant by the effect of the variant on BMI (as estimated in the GIANT consortium) and summing across the 97 variants (Burgess & Thompson, 2013). Overall, this score explained 1.7% of the variance in BMI. We performed both fractional polynomial and piecewise linear methods for estimating the relationships of BMI with SBP and DBP. The fractional polynomial method was implemented using 100 strata, whereas the piecewise linear method was implemented using 10 strata to avoid the exposure-outcome curve being overly jagged. The reference point was set at 25 kg/m².

To account for the multiple centers, we standardized the measure of BMI by stratifying individuals based on their residual value of BMI (the IV-free exposure) after regression of BMI on the allele score, age, sex, and center (as a categorical variable). Adjustment for age, sex, and center was also made in the regressions to obtain the LACE estimates in each quantile group. If additional population stratification were expected, we could additionally adjust for genetic principal components to minimize the effect of population stratification in biasing IV estimates.

To assess the assumption that the effect of the IV on BMI is constant over the entire distribution of BMI, we also considered BMI as the outcome and calculated the associations of the IV with BMI in each of the strata. We then conducted tests (trend and Cochran Q tests) to investigate heterogeneity in the IV associations with BMI in different strata.

5.1 | Results of applied example

The exposure-outcome relationships for BMI with SBP and DBP estimated using the fractional polynomial and piecewise linear methods are presented in Figure 1. There were strong causal effects of BMI on both SBP and DBP (*P*-value $< 1 \times 10^{-5}$ for the causal estimates differing from zero in the fractional polynomial methods). For comparison, the standard two-stage least squares linear estimate was 0.527 mmHg per 1 kg/m² increase in BMI (95% CI: 0.363, 0.691) for SBP and 0.433 mmHg (95% CI: 0.338, 0.528) for DBP.

There was strong evidence that the association between BMI and SBP was nonlinear, with the quadratic test yielding a *P*-value of 0.0026 (fractional polynomial test *P*-value = 0.0164, Cochran Q test *P*-value = 0.0346). The best-fitting fractional polynomial of degree 1 for the relationship between BMI and SBP had power −0.5, and there was no evidence to suggest that a fractional polynomial of degree 2 fitted the data better (*P*-value = 0.135). The estimate of the exposure-outcome relationship from the piecewise linear method visually suggested a threshold–type relationship, with a steep slope up to a BMI value of about 32 kg/m², and a slightly negative slope from 32 kg/m² onwards. The relationship between BMI and SBP was similar in individuals with no history of hypertension (Fig. S1).

The association between BMI and DBP was also nonlinear (quadratic test *P*-value = 0.0005, fractional polynomial test *P*-value = 0.0114, Cochran Q test *P*-value = 0.0049), and there was strong evidence that the best-fitting fractional polynomial of degree 2 (with p_1 and $p_2 = 3$) fitted the data better than the best-fitting fractional polynomial of degree 1 (*P*-value = 0.0062). There was no evidence of a different relationship between BMI and DBP for underweight individuals, with the exposure-outcome curve increasing almost linearly up to a BMI of around 40 kg/m². But for hyperobese
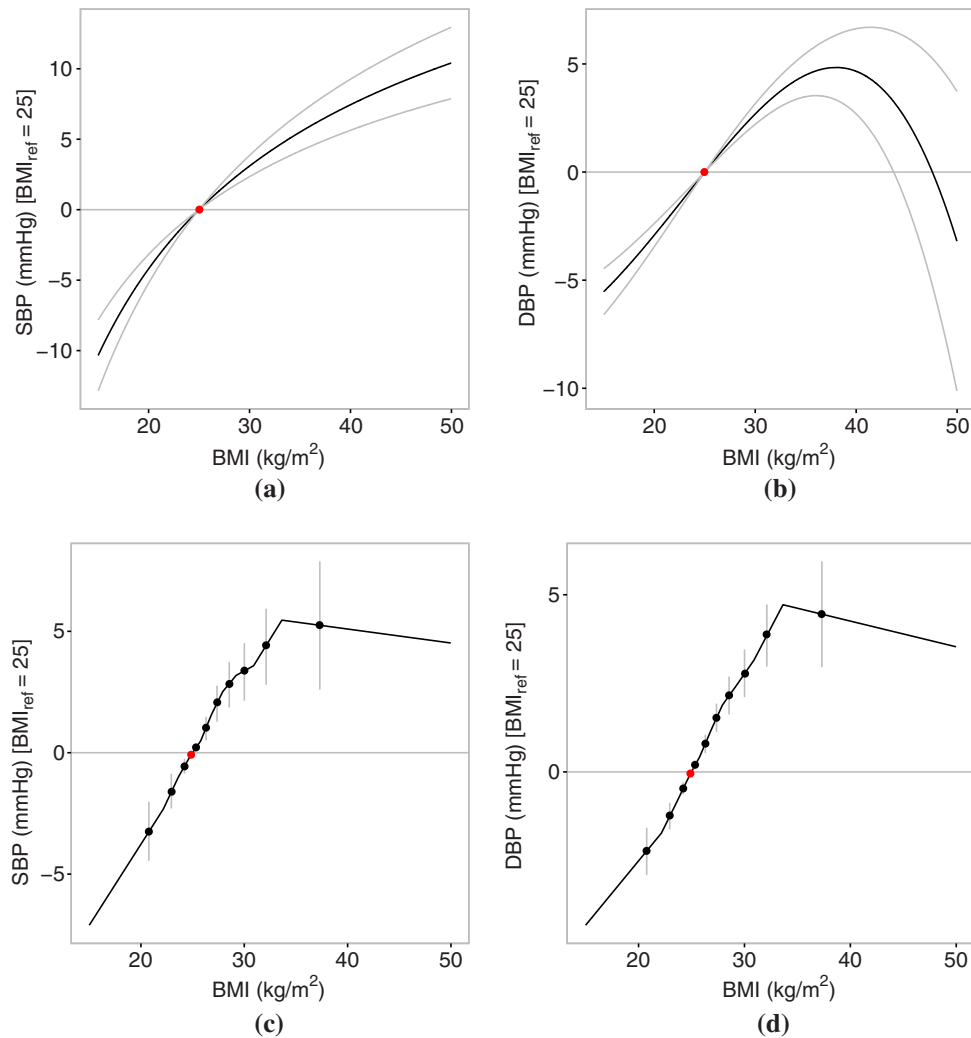
FIGURE 1 Causal effects of body mass index (BMI) on blood pressure (systolic blood pressure, SBP; diastolic blood pressure, DBP) using the fractional polynomial and piecewise linear methods on data from UK Biobank: (a) SBP (fractional polynomial method), (b) DBP (fractional polynomial method), (c) SBP (piecewise linear method), and (d) DBP (piecewise linear method). The red point represents the reference point of BMI of 25 kg/m². Gray lines represent 95% CIs. The fractional polynomial method used 100 strata.

individuals (BMI > 40 kg/m²), DBP seemed to decrease sharply. This was particularly evident in the fractional polynomial method, which used a greater number of strata and hence had more resolution to consider the shape of the exposure-outcome relationship at the extremes of the BMI distribution. One potential reason for this finding is that hyperobese individuals with high DBP are less likely to be enrolled in UK Biobank, perhaps due to differential survival probability. Another reason could be the difficulties in estimating blood pressure in hyperobese individuals (Leblanc et al., 2013). However, there was no evidence that the relationship between BMI and DBP was nonlinear in individuals with no history of hypertension (P-value >0.05 for all tests; Fig. S1).

There was no evidence that the associations of the IV with BMI varied between the different strata (trend test P-value = 0.135, Cochran Q test P-value = 0.901).

6 | DISCUSSION

In this paper, we have proposed and tested two novel methods for examining the relationship between an exposure and an outcome using IV analysis in the context of Mendelian randomization. Both methods rely on stratifying the population based on the IV-free exposure; the exposure minus the effect of the IV. A causal effect, referred to as a LACE, is estimated in each stratum of population. The first method performs metaregression on these LACE estimates using fractional polynomials. The second method estimates a continuous piecewise linear function, the gradient of which in each stratum is the LACE estimate for that stratum. Both methods were demonstrated in a simulation study to estimate the true exposure-outcome relationship well when its functional form corresponded to the form of the estimate from each method (i.e., when the exposure-outcome relationship was a fractional

polynomial for the fractional polynomial method, and when the relationship was piecewise linear for the piecewise linear method), with causal estimates being close to unbiased and coverage rates generally maintaining nominal levels (in particular, coverage rates were always correct under the null). Additionally, tests of nonlinearity were provided and their performance was assessed. The quadratic and fractional polynomial tests had the best performance in terms of Type I error rate and power.

6.1 | Comparison of methods

The recommendation as to which method to use depends on the aim of the investigation. The fractional polynomial method will always provide a smooth estimate of the exposure-outcome relationship, and as such has more consistent performance when a large number of strata are chosen (i.e., when the shape of the relationship is considered over a wider and more detailed range of the exposure distribution). Fractional polynomials of degree 1 had better performance than those of degree 2 in terms of bias and coverage of effect estimates. However, fractional polynomials of degree 1 are less flexible and would not be able to model complex exposure-outcome relationships. Additionally, they tend to smooth over discrepancies in the data. For example, if the LACE estimate for individuals in the lowest quantile group for BMI was substantially different to the other LACE estimates, then both this difference and any uncertainty in the LACE estimate would be smoothed over somewhat in the fractional polynomial estimate. Preference between the methods therefore comes down to a question of prior belief: if one truly believes the true exposure-outcome relationship to be smooth, and that estimates in the surrounding quantiles should be used to model the LACE in the target quantile, then the fractional polynomial method should be preferred. However, if one does not want to smooth over estimates, then the piecewise linear method should be preferred; however, the estimate of the exposure-outcome relationship will be more jagged and variable.

The number of strata and divisions between strata should ideally be specified before the analysis according to practical considerations (e.g., previously determined categories of BMI) and the sample size available. If the number of strata is too large, then each stratum will have a small sample size. The LACE estimate in a small stratum will be imprecise and may be susceptible to weak instrument bias (Burgess & Thompson, 2011), particularly if the genetic variant is rare.

6.2 | Interpretation of the exposure-outcome relationship

If the function relating the exposure to the average value of the outcome is homogeneous across the population, then the methods provided in this paper estimate this function (the exposure-outcome relationship) even if there is unmeasured confounding. If the function is heterogeneous, then the situation is more complicated (Small, 2014). For example, taking BMI as the exposure, if the subject-specific effect curve (as defined by Small, 2014) is linear for all individuals in the population, but the magnitude of effect is greater for overweight individuals, then the exposure-outcome relationship will be quadratic (or at least convex and positive) rather than linear. The exposure-outcome curve at low values of the exposure is only estimated using underweight individuals, and at high values of the exposure only using overweight individuals. However, this is perhaps the most relevant way to express the exposure-outcome relationship, as the causal effect of reducing one's BMI from 20 to 18 kg/m^2 is not so relevant for someone with a BMI of 40 kg/m^2. Hence, we do not claim any global interpretation of the exposure-outcome relationship as estimated in this paper apart from in the unlikely case that the functional relationship is homogeneous in the population. It is better interpreted as a series of local estimates, which are graphically connected in order to compare and contrast trends in these local estimates at different values of the exposure, and to compare the relative benefit of intervening on the exposure for individuals with different values of the exposure, but which does not necessarily reflect the effect of intervening on the exposure to take any value in its distribution for any single individual.

6.3 | Measurement error in the exposure

As has been noted in other contexts, estimates of nonlinear relationships are sensitive to measurement error in the exposure (Keogh, Strawbridge, & White, 2012). The standard "triple whammy" of measurement error is likely to apply here: measurement error biases parameter estimates, reduces power, and obscures important features in the shape of relationships (Carroll, Ruppert, Stefanski, & Crainiceanu, 2006). For example, with a threshold relationship, measurement error in the exposure would mean that the point of inflexion in the exposure-outcome relationship would be less sharply evident. In the case of BMI, measurement error is not such an issue, as height and weight can be measured precisely, and neither variable experiences substantial diurnal or seasonal variation. However, for other exposures, measurement error may affect results more severely. As noted in the additional simulation analyses, bias due to measurement error can also occur if the model for the genetic association with the exposure is misspecified.

6.4 | Requirement of concomitant and individual-level data

Many recent advances in Mendelian randomization have enabled investigations to be performed using summarized

data on the genetic associations with the exposure and with the outcome only, and/or in a two-sample setting in which genetic associations with the exposure and with the outcome are estimated on separate groups of individuals (Burgess, Butterworth, & Thompson, 2013; Burgess et al., 2015). However, estimation of the exposure-outcome relationship requires both individual-level data and a one-sample setting (otherwise neither stratification of the population nor the estimation of genetic associations with the outcome in the strata is possible). Although large cohorts with concomitant data on genetic variants, exposures, and outcomes are becoming more widely available, particularly in the form of biobanks such as UK Biobank.

In conclusion, these two novel methods are useful in investigating nonlinear exposure-outcome relationships. The methods allow for easy graphical assessment of the shape of the relationship, and allied with tests of nonlinearity, provide an effective tool for assessing nonlinear exposure-outcome relationships using IV analysis for Mendelian randomization.

ACKNOWLEDGMENTS

REFERENCES

Abecasis G. R., Auton A., Brooks L. D., DePristo M. A., Durbin R. M., Handsaker R. E., … McVean G. A. (2012). An integrated map of genetic variation from 1,092 human genomes. *Nature*, *491*(7422), 56–65.

Angrist, J., Imbens, G., & Rubin, D. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, *91*(434), 444–455.

Arellano, M. (2003). *Advances in economics and econometrics, theory and applications, Eight World Congress* (Vol. II, pp. 358–364). Cambridge: Cambridge University Press.

Bagnardi, V., Zambon, A., Quatto, P., & Corrao, G. (2004). Flexible meta-regression functions for modeling aggregate dose–response data, with an application to alcohol and mortality. *American Journal of Epidemiology*, *159*(11), 1077–1086.

Burgess, S., Butterworth, A., & Thompson, S. G. (2013). Mendelian randomization analysis with multiple genetic variants using summarized data. *Genetic Epidemiology*, *37*(7), 658–665.

Burgess, S., Davies, N. M., & Thompson, S. G. (2014). Instrumental variable analysis with a nonlinear exposure–outcome relationship. *Epidemiology*, *25*(6), 877–885.

Burgess, S., Scott, R. A., Timpson, N. J., Davey Smith, G., Thompson, S. G. (2015). Using published data in Mendelian randomization: a blueprint for efficient identification of causal risk factors. *European Journal of Epidemiology*, *30*(7), 543–52.

Burgess, S., & Thompson, S. G. (2011). Bias in causal estimates from Mendelian randomization studies with weak instruments. *Statistics in Medicine*, *30*(11), 1312–1323.

Burgess, S., & Thompson, S. G. (2013). Use of allele scores as instrumental variables for Mendelian randomization. *International Journal of Epidemiology*, *42*(4), 1134–1144.

Burgess, S., & Thompson, S. G. (2015). *Mendelian randomization: Methods for using genetic variants in causal estimation*. Boca Raton, FL: CRC Press.

Carroll, R. J., Ruppert, D., Stefanski, L. A., Crainiceanu, C. M. (2006). *Measurement error in nonlinear models: A modern perspective*. Boca Raton, FL: CRC Press.

Didelez, V., & Sheehan, N. (2007). Mendelian randomization as an instrumental variable approach to causal inference. *Statistical Methods in Medical Research*, *16*(4), 309–330.

Ebrahim, S., & Smith, G. D. (2008). Mendelian randomization: Can genetic epidemiology help redress the failures of observational epidemiology? *Human Genetics*, *123*(1), 15–33.

Flegal, K. M., Kit, B. K., Orpana, H., & Graubard, B. I. (2013). Association of all-cause mortality with overweight and obesity using standard body mass index categories: A systematic review and meta-analysis. *Journal of the American Medical Association*, *309*(1), 71–82.

Hernán, M. A., & Robins, J. M. (2006). Instruments for causal inference: An epidemiologist's dream? *Epidemiology*, *17*(4), 360–372.

Horowitz, J. L. (2011). Applied nonparametric instrumental variables estimation. *Econometrica*, *79*(2), 347–394.

Hosseini, M., Ataei, N., Aghamohammadi, A., Yousefifard, M., Taslimi, S., Ataei, F. (2010). The relation of body mass index and blood pressure in iranian children and adolescents aged 7–18 years old. *Iranian Journal of Public Health*, *39*(4), 126–134.

Imbens, G. W., & Angrist, J. D. (1994). Identification and estimation of local average treatment effects. *Econometrica*, *62*(2), 467–475.

Kaufman, J. S., Asuzu, M. C., Mufunda, J., Forrester T., Wilks, R., Luke, A. … Cooper R. S. (1997). Relationship between blood pressure and body mass index in lean populations. *Hypertension*, *30*(6), 1511–1516.

Keogh, R. H., Strawbridge, A., & White, I. R. (2012). Effects of classical exposure measurement error on the shape of exposure–disease associations. *Epidemiologic Methods*, *1*, 13–32.

Leblanc, M. É., Croteau, S., Ferland, A., Bussières J., Cloutier L., Hould F. S., … Poirier P. (2013). Blood pressure assessment in severe obesity: Validation of a forearm approach. *Obesity*, *21*(12), E533–E541.

Locke A. E., Kahali B., Berndt S. I., Justice A. E., Pers T. H., Day F. R., … Speliotes E. K. (2015). Genetic studies of body mass index yield new insights for obesity biology. *Nature*, *518*(7538), 197–206.

Newey, W. K., & Powell, J. L. (2003). Instrumental variable estimation of nonparametric models. *Econometrica*, *71*(5), 1565–1578.

Nielsen, G. A., & Andersen, L. B. (2003). The association between high blood pressure, physical fitness, and body mass index in adolescents. *Preventive Medicine*, *36*(2), 229–234.

Royston, P., & Altman, D. G. (1994). Regression using fractional polynomials of continuous covariates: Parsimonious parametric modelling. *Journal of Applied Statistics*, *43*(3), 429–467.

Royston, P., Ambler, G., & Sauerbrei, W. (1999). The use of fractional polynomials to model continuous risk variables in epidemiology. *International Journal of Epidemiology*, *28*(5), 964–974.

Silverwood R. J., Holmes M. V., Dale C. E., Lawlor D. A., Whittaker J. C., Smith G. D., … Dudbridge F. (2014). Testing for non-linear causal effects using a binary genotype in a Mendelian randomization study: Application to alcohol and cardiovascular traits. *International Journal of Epidemiology*, *43*(6), 1781–1790.

Small, D. S. (2014). Commentary: Interpretation and sensitivity analysis for the localized average causal effect curve. *Epidemiology*, *25*(6), 886–888.

Sudlow C., Gallacher J., Allen N., Beral V., Burton P., Danesh J., … Collins R. (2015). UK Biobank: An open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med*, *12*(3), e1001779.

Thomas, D., Lawlor, D., & Thompson, J. (2007). Re: Estimation of bias in nongenetic observational studies using "Mendelian triangulation" by Bautista et al. *Annals of Epidemiology*, *17*(7), 511–513.

Thompson, S. G., & Sharp, S. J. (1999). Explaining heterogeneity in meta-analysis: a comparison of methods. *Statistics in Medicine*, *18*(20), 2693–2708.

Yau, L. H., & Little, R. J. (2001). Inference for the complier-average causal effect from longitudinal data subject to noncompliance and missing data, with application to a job training assessment for the unemployed. *Journal of the American Statistical Association*, *96*(456), 1232–1244.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.