The GENIUS Approach to Robust Mendelian Randomization Inference

Eric Tchetgen Tchetgen, BaoLuo Sun and Stefan Walter

Abstract. Mendelian randomization (MR) is a popular instrumental variable (IV) approach, in which one or several genetic markers serve as IVs that can sometimes be leveraged to recover valid inferences about a given exposure-outcome causal association subject to unmeasured confounding. A key IV identification condition known as the exclusion restriction states that the IV cannot have a direct effect on the outcome which is not mediated by the exposure in view. In MR studies, such an assumption requires an unrealistic level of prior knowledge about the mechanism by which genetic markers causally affect the outcome. As a result, possible violation of the exclusion restriction can seldom be ruled out in practice. To address this concern, we introduce a new class of IV estimators which are robust to violation of the exclusion restriction under data generating mechanisms commonly assumed in MR literature. The proposed approach named "MR G-Estimation under No Interaction with Unmeasured Selection" (MR GENIUS) improves on Robins' G-estimation by making it robust to both additive unmeasured confounding and violation of the exclusion restriction assumption. In certain key settings, MR GENIUS reduces to the estimator of Lewbel (*J. Bus. Econom. Statist.* **30** (2012) 67–80) which is widely used in econometrics but appears largely unappreciated in MR literature. More generally, MR GENIUS generalizes Lewbel's estimator to several key practical MR settings, including multiplicative causal models for binary outcome, multiplicative and odds ratio exposure models, case control study design and censored survival outcomes.

Key words and phrases: Additive model, confounding, exclusion restriction, G-estimation, instrumental variable, robustness.

1. INTRODUCTION

Mendelian randomization (MR) is an instrumental variable approach with growing popularity in epidemiology studies. In MR, one aims to establish a causal association between a given exposure and an outcome of interest in the presence of possible unmeasured confounding, by leveraging one or more genetic markers defining the IV (Davey Smith and Ebrahim, 2003, 2004, Lawlor et al.,

Eric Tchetgen Tchetgen is Professor, Department of Statistics, The Wharton School of the University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA (e-mail: ett@wharton.upenn.edu). BaoLuo Sun is Assistant Professor, Department of Statistics and Applied Probability, National University of Singapore, Singapore, Republic of Singapore (e-mail: stasb@nus.edu.sg). Stefan Walter is Principal Investigator, Department of Medicine and Public Health, Rey Juan Carlos University, Madrid, Spain (e-mail: stefan.walter@urjc.es).

2008). In order to be a valid IV, a genetic marker must satisfy the following key conditions:

(a) It must be associated with the exposure.

(b) It must be independent of any unmeasured confounder of the exposure-outcome relationship.

(c) There must be no direct effect of the genetic marker on the outcome which is not fully mediated by the exposure in view.

Assumption (c) also known as the exclusion restriction is rarely credible in the context of MR as it requires complete understanding of the biological mechanism by which each marker influences the outcome. Such a priori knowledge may be unrealistic in practice due to the possible existence of unknown pleiotropic effects of the markers (Little and Khoury, 2003, Davey Smith and Ebrahim, 2003, 2004, Lawlor et al., 2008). Violation of assumption (b) can also occur due to linkage disequilibrium or population stratification (Lawlor et al., 2008). Possible vio-

lation or near violation of assumption (a) known as the weak instrumental variable problem also poses an important challenge in MR as individual genetic effects on phenotypes can be fairly weak.

There has been tremendous interest in the development of statistical methods to detect and account for violation of IV assumptions (a)–(c). The literature addressing violation of assumption (a) is arguably the most developed and extends to possible nonlinear models under a generalized methods of moments framework; notable papers of this rich literature include Staiger and Stock (1997), Stock and Wright (2000), Stock, Wright and Yogo (2002), Chao and Swanson (2005) and Newey and Windmeijer (2009). Methodology to address violations of (b) or (c) is far less developed and constitutes the central focus of this paper. An important approach involves deriving bounds and performing sensitivity analysis for causal estimands which are typically only partially identifiable from the observed data law without further untestable assumptions (Manski and Pepper, 2000, Hirano et al., 2000, Nevo and Rosen, 2012, Conley, Hansen and Rossi, 2012, Flores and Flores-Lagunes, 2013, Mealli and Pacini, 2013, Huber, 2014). The other concerns point identification and estimation of causal effects under violation of either of these assumptions, and three strands of work stand out in recent literature primarily in multiple-IV settings under standard linear outcome and exposure models. In the first strand, Kang et al. (2016) developed a penalized regression approach that can recover valid inferences about the causal effect of interest provided fewer than fifty percent of genetic markers are invalid IVs (known as majority rule); Windmeijer et al. (2019) improved on the penalized approach, including a proposal for standard error estimation lacking in Kang et al. (2016). In an alternative approach, Han (2008) established that the median of multiple estimators of the effect of exposure obtained using one instrument at the time is a consistent estimator also assuming majority rule and that IVs cannot have direct effects on the outcome unless the IVs are uncorrelated. Bowden et al. (2016) explore closely related weighted median methodology. In a second strand of work, Guo et al. (2018) proposed two stage hard thresholding (TSHT) with voting, which is consistent for the causal effect under linear outcome and exposure models, and a plurality rule which can be considerably weaker than the majority rule. The plurality rule is defined in terms of regression parameters encoding the association of each invalid IV with the outcome and that encoding the association of the corresponding IV with the exposure. The condition effectively requires that the number of valid IVs is greater than the largest number of invalid IVs with equal ratio of the regression coefficients given above. Furthermore, they provide a simple construction for 95% confidence intervals to obtain inferences about the exposure effect

which are guaranteed to have correct coverage under the plurality rule. Importantly, in these first two strands of work, a candidate IV may be invalid either because it violates the exclusion restriction, or because it shares an unmeasured common cause with the outcome, that is, either (b) or (c) fails. Both the penalized approach and the median estimator may be inconsistent if majority rule fails, while TSHT may be inconsistent if plurality rule fails. For instance, it is clear that neither approach can recover valid inferences if all IVs violate either assumption (b) or (c). In a third strand of work, Kolesár et al. (2015) considered the possibility of identifying the exposure causal effect when all IVs violate the exclusion restriction (c), provided the effects of the IVs on the exposure are asymptotically orthogonal to their direct effects on the outcome as the number of IVs tends to infinity. A closely related meta-analytic version of their approach known as MR-Egger has recently emerged in the epidemiology literature (Bowden, Davey Smith and Burgess, 2015); they referred to the orthogonality condition as the instrument strength independent of direct effect (InSIDE) assumption. As pointed out by Kang et al. (2016), the orthogonality condition on which these approaches rely may be hard to justify in MR settings as it potentially restricts unknown pleiotropic effects of genetic markers often with little to no biological basis. A notable feature of aforementioned methods is that they are primarily tailored to a multiple-IV setting, in fact methods such as MR-Egger are consistent only under an asymptotic theory in which the number of IVs goes to infinity, together with sample size. It is also important to note that because confidence intervals for the causal effect of the exposure obtained by Windmeijer et al. (2019) and Guo et al. (2018) rely on a consistent model selection procedure, such confidence intervals fail to be uniformly valid over the entire model space (Guo et al., 2018, Leeb and Pötscher, 2008).

Because in practice, it is not possible to ensure that either majority rule or plurality rule holds, it is important to develop causal inference and estimation methods that are robust to possible violation of IV assumptions under alternative conditions. Lewbel (2012, 2018) proposed novel identification and estimation strategies with mismeasured and endogenous regressor models by leveraging a heteroscedastic covariance restriction, which has since been widely applied in econometrics and social sciences. In Section 2, we introduce notation used throughout and provide a review of the invalid IV model considered by Lewbel (2012). We extend Lewbel's identification result in Section 3. The proposed framework which we call "MR G-Estimation under No Interaction with Unmeasured Selection" (MR GENIUS) can also be viewed as a version of Robins' G-estimation (Robins, 1994) that is robust to both

additive unmeasured confounding and violation of IV assumptions, and which unlike the aforementioned methods equally applies whether one has observed a single or many candidate IVs. An important feature of multiple IV MR GENIUS is that the correlation structure for the IVs can essentially remain unrestricted without necessarily affecting identification, this is in contrast with Bowden, Davey Smith and Burgess (2015) who require uncorrelated IVs and Kang et al. (2016) who likewise require IV correlation structure to be somewhat restricted (Windmeijer et al., 2019). Section 4 presents several key generalizations including MR GENIUS under multiplicative or odds ratio exposure models, as well as for right censored time-to-event endpoint under a structural additive hazards model, which extends the recent semiparametric IV estimator of Martinussen et al. (2017) against possible violation of the exclusion restriction assumption. In Section 5, we evaluate the proposed methods and compare them to a number of previous MR methods in extensive simulation studies. In Section 6, we illustrate the methods in an MR analysis of the effect of diabetes on memory in the Health and Retirement Study. Section 7 offers some concluding remarks. We note that although MR is used as motivating example throughout the paper, the development of methodology to adequately address the issue of invalid IVs remains a priority for several disciplines, including biostatistics, epidemiology, econometrics and sociology, in which our results equally apply.

2. NOTATION AND DEFINITIONS

Suppose that one has observed n i.i.d. realizations of a vector (A, G, Y) where A is an exposure, G the candidate IV and Y is the outcome. Let U denote an unmeasured confounder (possibly multivariate) of the effect of A on Y. G is said to be a valid instrumental variable provided it fulfills the following three conditions (Didelez, Meng and Sheehan, 2010):

ASSUMPTION 1. IV relevance: $G \not\perp\!\!\!\perp A | U$;

ASSUMPTION 2. IV independence: $G \perp\!\!\!\perp U$;

ASSUMPTION 3. Exclusion restriction: $G \perp\!\!\!\perp Y | A, U$.

The first condition ensures that the IV is a correlate of the exposure even after conditioning on U. The second condition states that the IV is independent of all unmeasured confounders of the exposure-outcome association, while the third condition formalizes the assumption of no direct effect of G on Y not mediated by A (assuming Assumption 2 holds). The causal diagram in Figure 1a encodes these three assumptions and therefore provides a graphical representation of the IV model. It is well known that while a valid IV satisfying Assumptions 1–3, that is, the causal diagram in Figure 1a, suffices to obtain a valid statistical test of the sharp null hypothesis of no individual causal effect, the population average causal effect is itself not point identified with a valid IV without an additional assumption. Angrist, Imbens and Rubin (1996) showed that the local average treatment effect among compliers can be nonparametrically identified under an additional monotonicity assumption about the effect of Z on D. The additive and multiplicative structural mean models of Robins (1989, 1994) represent another approach in which the average treatment effect on the treated can be identified under effect heterogeneity restrictions. In case of a valid binary IV and binary exposure, Wang and Tchetgen Tchetgen (2018) recently established that the average causal effect β_a of A on Y is nonparametrically identified by the so-called Wald estimand

$$\beta_a = \delta \equiv \frac{E(Y|G=1) - E(Y|G=0)}{E(A|G=1) - E(A|G=0)},$$

if either of the conditions

$$(2.1) \qquad E(Y|A, G, U) = \beta_a A + \beta_u(U),$$

$$(2.2) \qquad E(A|G, U) = \alpha_g G + \alpha_u(U),$$

holds, where the unknown functions $\beta_u(\cdot)$ and $\alpha_u(\cdot)$ are restricted only by natural features of the model, see, for example, such that the outcome and exposure means are bounded between zero and one in the binary case. Suppose that, as encoded in the diagram given in Figure 1b, Assumption 3 does not necessarily hold. Lewbel (2012) considered identification and estimation of β_a under the
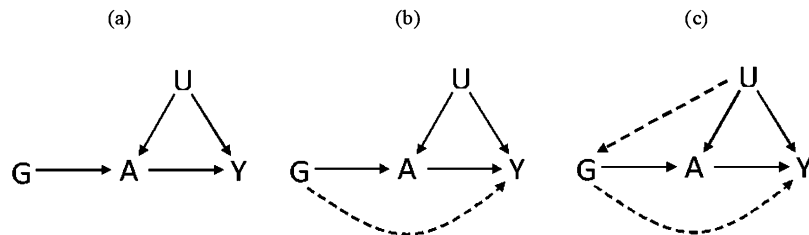

FIG. 1. *Directed acyclic graph depicting* (a) *a valid instrument G which satisfies Assumptions* 1–3, (b) *the situation in which exclusion restriction* (*Assumption* 3) *does not necessarily hold* (*the dashed line indicates possible direct effect of G on outcome Y*), *and* (c) *the situation in which IV independence* (*Assumption* 2) *and exclusion restriction* (*Assumption* 3) *do not necessarily hold* (*the dashed lines indicate possible direct effects of U on G, and of G on Y*).

invalid IV model

$$(2.3) \quad E(Y|A, G, U) = \beta_a A + \beta_g(G) + \beta_u(U),$$

$$(2.4) \quad E(A|G, U) = \alpha_g(G) + \alpha_u(U),$$

where $\beta_g(G)$ encodes the direct effect of G on Y, with $\beta_g(0) = \alpha_g(0) = 0$. Note that the models for $E(Y|A, G, U)$ and $E(A|G, U)$ considered by Bowden, Davey Smith and Burgess (2015) satisfy (2.3) with $\beta_g(\cdot)$ and $\alpha_g(\cdot)$ linear functions, while Kang et al. (2016) specified models implied by these two restrictions. With binary exposure A, we consider identification of the average causal effect for the following generalization of the invalid IV model considered in (2.3).

ASSUMPTION 4.

(4a) For unknown functions $\beta_g(\cdot)$ and $\alpha_g(\cdot)$ of (U, G) where $\beta_g(U, 0) = \alpha_g(U, 0) = 0$, and unknown function $\beta_a(\cdot)$ of U,

$$(2.5) \quad \begin{aligned} E(Y|A, G, U) &= \beta_a(U)A + \beta_g(U, G) + \beta_u(U), \\ E(A|G, U) &= \alpha_g(U, G) + \alpha_u(U). \end{aligned}$$

(4b) The orthogonality conditions

$$\begin{aligned} &\text{cov}\{\beta_g(U, G), \alpha_g(U, G)|G\} \\ &= \text{cov}\{\beta_g(U, G), \alpha_u(U)|G\} = 0; \end{aligned}$$

$$(2.6) \quad \begin{aligned} &\text{cov}\{\beta_a(U), \alpha_g(U, G)|G\} \\ &= \text{cov}\{\beta_a(U), \alpha_u(U)|G\} = 0; \\ &\text{cov}\{\alpha_g(U, G), \beta_u(U)|G\} = 0, \end{aligned}$$

hold with probability 1.

Under Assumption 4a, the average causal effect of binary A on Y is given by $\mu \equiv E\{\beta_a(U)\}$. The model $\beta_g(U, G)$ encodes the direct effect of G on Y, with potential effect modification by unmeasured confounders U. Assumption 4b does not imply orthogonality of $\beta_u(U)$ and $\alpha_u(U)$ and therefore the degree of unmeasured confounding is not restricted by these orthogonality conditions. In contrast the degree of common effect modifiers in the outcome and exposure models is effectively restricted by (2.6).

As pointed out by a reviewer, it is instructive to study the effect heterogeneity restrictions (2.6) in the context of the following linear structural models with random coefficients (Wooldridge, 2003, 2010),

$$(2.7) \quad \begin{aligned} Y_i &= \beta_{ai} A_i + \beta_{gi} G_i + \beta_{ui} + v_i; \\ A_i &= \alpha_{gi} G_i + \alpha_{ui} + w_i, \quad i = 1, \ldots, n, \end{aligned}$$

where $E(v_i|A_i, U_i, G_i) = E(w_i|U_i, G_i) = 0$, $\beta_{ai} \equiv \beta_a(U_i)$, $\beta_{gi} \equiv \beta_g(U_i)$, $\beta_{ui} \equiv \beta_u(U_i)$, $\alpha_{gi} \equiv \alpha_g(U_i)$ and $\alpha_{ui} \equiv \alpha_u(U_i)$. Because the random coefficient β_{ai} is correlated with A_i by virtue of unmeasured confounding, (2.7) is also known as a correlated random coefficient

model (Heckman and Vytlacil, 1998). It is convenient to rewrite the outcome model as $Y_i = \mu A_i + \epsilon_i'$, where $\epsilon_i' = \{\beta_{ai} - \mu\}A_i + \beta_{gi} G_i + \beta_{ui} + v_i$. Then a sufficient set of conditions for identification of μ is $G_i \perp\!\!\!\perp U_i$, $\beta_{gi} = 0$ and either $\beta_{ai} = \mu$ or $\alpha_{gi} = \alpha_g$ with probability 1 so that $E\{\epsilon_i'\{G_i - E(G_i)\}\} = 0$. When $\beta_{gi} \neq 0$, then the proof of Lemma 3.1 in the next section shows that $E\{\epsilon_i'\{G_i - E(G_i)\}\{A - E(A|G)\}\} = 0$ under the restrictions in (2.6), where $\{G_i - E(G_i)\}\{A - E(A|G)\}$ is the constructed instrument proposed by Lewbel (2012). As a special case, the conditions in (2.6) are satisfied if $\beta_g(U, G) = \beta_g(G)$, $\alpha_g(U, G) = \alpha_g(G)$ and $\beta_a(U) = \beta_a$ with probability 1, which is the scenario considered in Lewbel (2012); this is equivalent to setting $\beta_{gi} = \beta_g$, $\alpha_{gi} = \alpha_g$ and $\beta_{ai} = \beta_a$ under the linear random coefficients model (2.7). Intuitively, in this case the constructed instrument works because its exposure regression residual is orthogonal to $\beta_g G_i$ in ϵ_i'. Assumption 4b generalizes the identification result and may hold even if assumptions made by Lewbel (2012) do not, as illustrated by the following example. Suppose that $\alpha_u(U) - E[\alpha_u(U)] = \sum_{k=1}^K \gamma_k \times \phi_k(U)$ for a vector of zero-mean functions $\mathbf{\Phi} = \{\phi_k : k = 1, \ldots, K\}$, and likewise $\beta_u(U) - E[\beta_u(U)] = \sum_{j=1}^J \theta_j \times \tau_j(U)$ for $\mathbf{T} = \{\tau_j : j = 1, \ldots, J\}$. Denote the linear vector space spanned by the vector $\mathbf{\Omega}$ of functions in U to be $S(\mathbf{\Omega})$, and Π to be the least squares projection operator. Let

$$\begin{aligned} &\alpha_g(U, G) \\ &= \alpha_{g0}(G) + \sum_{k=1}^K \alpha_{gk}(G)\{\phi_k(U) - \Pi[\phi_k(U)|S(\mathbf{T})]\}, \end{aligned}$$

$\alpha_{gk}(0) = 0$ for $k = 0, 1, \ldots, K$,

$$\begin{aligned} &\beta_g(U, G) \\ &= \beta_{g0}(G) + \beta_{g1}(G)\{\omega(U) - \Pi[\omega(U)|S(\mathbf{\Phi}) \\ &\quad + S(\phi_k(U) - \Pi[\phi_k(U)|S(\mathbf{T})], \\ &\quad k = 1, 2, \ldots, K)]\}, \end{aligned}$$

for arbitrary function $\omega(U)$ with $\beta_{g0}(0) = \beta_{g1}(0) = 0$, and

$$\begin{aligned} &\beta_a(U) \\ &= \zeta(U) - \Pi[\zeta(U)|S(\mathbf{\Phi}) \\ &\quad + S(\phi_k(U) - \Pi[\phi_k(U)|S(\mathbf{T})], \\ &\quad k = 1, 2, \ldots, K)] \end{aligned}$$

for arbitrary function $\zeta(U)$. Note that $\mathbf{\Phi}$ and $\mathbf{T}$ can be of different dimension. Then the orthogonality conditions of Assumption 4b are satisfied.

Recent work by Shardell and Ferrucci (2016) proposes using baseline covariates interacted with G (or certain non-linear transformations of G) as instrumental variables when G violates Assumption 3, a method which was first

introduced in the causal mediation literature (Ten Have et al., 2007, Albert, 2008, Emsley, Dunn and White, 2010, Small, 2012, Zheng and Zhou, 2015). However, the approach generally does not allow for identification in the absence of covariates when G is binary. For example in the following outcome model in (2.7) with nonrandom coefficients,

$$Y_i = \beta_a A_i + \beta_g G_i + \underbrace{\beta_{ui} + v_i}_{\equiv \epsilon_i^{\dagger}},$$

the method fails to yield the two instrumental variables orthogonal to $\epsilon_i^{\dagger}$ that are necessary to identify (β_a, β_g).

3. IDENTIFICATION AND ESTIMATION UNDER VIOLATION OF EXCLUSION RESTRICTION

We consider identification of μ within the large class of data generating mechanisms that satisfy Assumptions 1, 2 and 4, which is given in the following lemma.

LEMMA 3.1. *Suppose that Assumptions* 1, 2 *and* 4 *hold, then*

$$\mu = \frac{E[\{G - E(G)\}\{A - E(A|G)\}Y]}{E[\{G - E(G)\}\{A - E(A|G)\}A]}$$

provided that

(3.1) $$\phi \equiv \text{cov}\{G, \text{var}(A|G)\} \neq 0.$$

The proof for Lemma 3.1 is given in Appendix A.1. In particular, for binary G, $\phi = \text{var}(G)\{\text{var}(A|G=1) - \text{var}(A|G=0)\}$ so that (3.1) is satisfied if and only if $\text{var}(A|G=1) - \text{var}(A|G=0) \neq 0$. Lemma 3.1 provides an explicit identifying expression for the average causal effect of A on Y in the presence of additive confounding, which leverages a candidate IV G that may or may not satisfy the exclusion restriction. In order for μ to be well defined, we require a slight strengthening of the IV relevance Assumption 1, that is, that $\text{var}(A|G)$ must depend on G. It is key to note that this assumption is empirically testable, and will typically hold for binary A, other than at certain exceptional laws. To illustrate, let $\pi(g) = \text{Pr}(A = 1|G = g)$ and suppose that Assumptions 1, 2 and 4 hold, however $\pi(1) = 1 - \pi(0)$, in which case (3.1) fails because $\text{var}(A|G = g) = \pi(g)(1 - \pi(g)) = \pi(1)(1 - \pi(1)) = \pi(0)(1 - \pi(0))$ does not depend on g and therefore the identifying expression given in the Lemma does not apply despite the candidate IV satisfying IV relevance Assumption 1, that is, $\pi(1) \neq \pi(0)$. Below, we extend Lemma 3.1 to allow for possible violation of both Assumptions 2 and 3.

The lemma motivates the following MR estimator, which is guaranteed to be consistent under Assumptions 1, 2, 4 and equation (3.1) irrespective of whether or not Assumption 3 also holds:

(3.2) $$\widehat{\beta}_a = \frac{\mathbb{P}_n[\{G - \mathbb{P}_n(G)\}\{A - \widehat{E}(A|G)\}Y]}{\mathbb{P}_n[\{G - \mathbb{P}_n(G)\}\{A - \widehat{E}(A|G)\}A]},$$

where $\mathbb{P}_n = n^{-1}\sum_{i=1}^{n}[\cdot]_i$ and $\widehat{E}(A|G = g) = \mathbb{P}_n[A_i \times 1(G_i = g)]/\mathbb{P}_n[1(G_i = g)]$. This estimator is the simplest instance of MR GENIUS estimation. The asymptotic distribution of the estimator is described in Appendix B.1. We note that (3.2) is equivalent to Lewbel's estimator which can be implemented as follows (Lewbel, 2018):

1. Obtain the estimated residuals $\hat{\epsilon}_a = A - \hat{\theta}^T G$ from ordinary least squares regression of A on G.

2. Estimate β_a by two-stage least squares regression of Y on A, using $(G - \bar{G})\hat{\epsilon}_a$ as the instrument, where $\bar{G}$ is the sample mean of G.

The above estimator assumes $E(A|G) = \theta^T G$, although the first step can be extended for some nonlinear, possibly unknown exposure model. Lewbel showed that $\hat{\beta}_a$ is consistent for the average causal effect which is parameterized by the scalar β_a under model (2.3) and condition (3.1) as well as $\text{cov}\{G, \epsilon_a \epsilon_y\} = 0$, where $\epsilon_y = Y - \beta_a A$.

3.1 Continuous Exposure

Consider the following stronger version of Assumption 4 in which $\beta_a(U) = \beta_a$ with probability 1,

ASSUMPTION 4*.

(4a*)

(3.3) $$E(Y|A, G, U) = \beta_a A + \beta_g(U, G) + \beta_u(U),$$

(3.4) $$E(A|G, U) = \alpha_g(U, G) + \alpha_u(U).$$

(4b*) The orthogonality conditions

$$\text{cov}\{\beta_g(U, G), \alpha_g(U, G)|G\}$$
$$= \text{cov}\{\beta_g(U, G), \alpha_u(U)|G\} = 0;$$
$$\text{cov}\{\alpha_g(U, G), \beta_u(U)|G\} = 0,$$

hold with probability 1.

Then for continuous A, Lemma 3.1 continues to hold under Assumptions 1, 2 and 4*, where $\beta_a = \mu$ now encodes the causal effect on the outcome mean upon increasing the exposure by one unit. Condition (3.1) implies that the residual error $\varepsilon_A = A - E(A|G)$ must be heteroscedastic, that is, $\text{var}(A|G) = E(\varepsilon_A^2|G)$ depends on G. In the next section, we generalize this identification result in several important directions particularly relevant to MR studies.

We note that as previously stated while $\text{var}(A|G)$ will generally depend on G for binary or discrete A, this may not always be the case for continuous A. However in this case, the assumption can be motivated under an underlying model for A with latent heterogeneity in the effect of G on A. Specifically, suppose that $A = \alpha_g^*(G, \varepsilon_g) + U_a + \varepsilon_a^*$ and $E(\varepsilon_a^*) = 0$, where ε_g and ε_a^* are unobserved random disturbances independent of (G, U); the disturbance ε_g may be viewed as unobserved genetic or environmental factors independent of G, that may however interact

with G to induce additive effect heterogeneity of G–A associations, for example, $\alpha_g^*(G, \epsilon_g) = \alpha_g^* G + \epsilon_g G$. Then, one can verify that the model in the above display implies that $A = \alpha_g(G) + \varepsilon_a$ where $\alpha_g(G) = E(\alpha_g^*(G, \varepsilon_g)|G) + E(U_a)$ and $\text{var}(\varepsilon_a|G) = \text{var}(\{U_a + \varepsilon_a^* + \alpha_g^*(G, \varepsilon_g) - E(\alpha_g^*(G, \varepsilon_g)|G)\}|G)$, which clearly depends on G, provided $\alpha_g^*(g, \epsilon_g) - \alpha_g^*(0, \epsilon_g)$ depends on ϵ_g for a value of g, therefore implying condition (3.1). A model for exposure which incorporates latent heterogeneity in the effects of G is quite natural in the MR context because such a model is widely considered a leading contestant to explain the mystery of missing heritability (Manolio et al., 2009).

Condition (3.1) is also related to the identification assumptions underlying an important class of bias-adjusted estimators of causal effects which leverage on gene-environment interactions when exclusion restriction of the IV is violated (Spiller et al., 2018). For example, in Spiller et al. (2018), a genetic risk score for body mass index (BMI) is shown to interact with a measure of social class (Townsend Deprivation Index, TDI). The genetic risk score explains a higher proportion of variance in BMI for people with high TDI values, and therefore condition (3.1) holds. However, unlike Spiller et al. (2018), we do not require that one observes ϵ_g in order to identify β_a, which is a key advantage.

3.2 Identification Under Violation of IV Independence

In this section, we aim to relax the IV independence Assumption 2, by allowing for dependence between U and G as displayed in Figure 1c. Therefore, we consider replacing Assumption 2 with the following weaker condition:

ASSUMPTION 2*. $\text{cov}(\beta_u(U), \alpha_u(U)|G) = \rho$ does not depend on G.

To illustrate Assumption 2* it is instructive to consider the following submodels of Assumption 4*: $\beta_u(U) = \beta_0 + \beta_u U$ and $\alpha_u(U) = \alpha_0 + \alpha_a U$. Then Assumption 2* implies $\text{var}(U|G) = \rho/(\beta_u \alpha_a)$, that is, the unmeasured confounder U has homoscedastic variance. Under Assumption 2*, $E(U|G)$ is left unrestricted therefore Assumption 2 may not hold. We have the following result.

LEMMA 3.2.   *Suppose that Assumptions 1, 2*, 4* hold, then $\beta_a = \mu$ provided that condition (3.1) holds.*

The proof of Lemma 3.2 is given in Appendix A.2. Lemma 3.2 implies that under Assumptions 1, 2*, 4* and condition (3.1), $\widehat{\beta}_a$ continues to be consistent even if $U \not\perp\!\!\!\perp G$. As previously mentioned, MR GENIUS may be viewed as a special case of G-estimation (Robins, 1997). In fact, under Assumption 4a and the additional assumption of no unobserved confounding given G, that is, if either $U \perp\!\!\!\perp A|G$ or $U \perp\!\!\!\perp Y|A, G$, the G-estimator $\widetilde{\beta}_a$ which solves an estimating equation of the form:

$$0 = \mathbb{P}_n[h(G)\{A - \widehat{E}(A|G)\}\{Y - \widetilde{\beta}_a A\}],$$

is consistent and asymptotically normal for any user-specified function $h(\cdot)$ (up to regularity conditions).

It is straightforward to verify that the MR GENIUS estimator (3.2) solves the estimating equation:

$$(3.5) \quad 0 = \mathbb{P}_n[\{G - \mathbb{P}_n(G)\}\{A - \widehat{E}(A|G)\}\{Y - \widehat{\beta}_a A\}],$$

therefore formally establishing an equivalence between MR GENIUS and G-estimation for the choice $h(G) = G - E(G)$. Remarkably, as we have established above, this specific choice of h renders G-estimation robust to unmeasured confounding under certain no-additive interactions conditions with unmeasured factors used in selecting exposure levels, therefore motivating the choice of acronym for the proposed approach.

4. GENERALIZATIONS

4.1 Incorporating Covariates

One may wish in an MR analysis to adjust for covariates, either to account for observed confounding of the exposure effect on the outcome, or to account for confounding of the effects of the genetic markers primarily by ancestry (known as population stratification) or simply to improve efficiency. In order to account for covariates C, we propose to solve:

$$(4.1) \quad \begin{aligned} 0 = \mathbb{P}_n[h(C)\{G - \widehat{E}(G|C)\} \\ \times \{A - \widehat{E}(A|G, C)\}\{Y - \widehat{\beta}_a A\}], \end{aligned}$$

for user-specified choice of h, where $\widehat{E}(G|C)$ and $\widehat{E}(A|G, C)$ are consistent estimators of $E(A|G, C)$ and $\widehat{E}(G|C)$ obtained say by fitting appropriate generalized linear models. For example, as G is binary, one may specify $\text{logit} \Pr(G = 1|C) = \omega_0 + \omega' C$ to obtain $\widehat{E}(G|C)$ by standard maximum likelihood estimation of a logistic regression, and likewise when A is binary, one may obtain $\widehat{E}(A|G, C)$ by fitting a similar logistic regression, and when A is continuous, an analogous linear regression could be used instead. Identification results established in previous Sections continue to apply by further conditioning on C.

4.2 Incorporating Multiple IVs

MR designs with multiple candidate genetic IVs may be used to strengthen identification and improve efficiency. Multiple candidate IVs can be incorporated by adopting a standard generalized method of moments approach. Specifically, under models (3.3) and (3.4) with G now a vector, we propose to obtain $\widehat{\beta}_a$ by solving:

$$(4.2) \quad \widehat{\beta}_a = \arg\min_{\beta_a} \mathbb{P}_n[\widehat{U}'(\beta_a)] W \mathbb{P}_n[\widehat{U}(\beta_a)],$$

where

$$\begin{aligned} \widehat{U}(\beta_a) = \{h(G, C) - \widehat{E}(h(G, C)|C)\} \\ \times \{A - \widehat{E}(A|G, C)\}\{Y - \beta_a A\} \end{aligned}$$

for a user-specified function $h(G, C)$ of dimension $K \geq 1$, and W is user-specified weight matrix. In practice, it may be convenient to set $h(G, C) = G$ and $W = I_{K \times K}$ the K dimensional identity matrix. Let $\overline{\beta}_a$ denote the corresponding estimator. A more efficient estimator $\widehat{\beta}_a$ can then be obtained by solving (4.2) with weight $W_{\text{opt}} = \mathbb{P}_n[\widehat{U}(\overline{\beta}_a)\widehat{U}(\overline{\beta}_a)']^-$ where T^- denotes the generalized inverse of matrix T. Identification of GMM is guaranteed (at least locally) provided that the second derivative with respect to β_a of the GMM objective function $\mathbb{P}_n[\widehat{U}'(\beta_a)]W\mathbb{P}_n[\widehat{U}(\beta_a)]$ is nonsingular at the truth, which is a generalization of condition (3.1). The asymptotic distribution of $\widehat{\beta}_a$ which solves (4.2) is described in Appendix B.2.

4.3 Multiplicative Causal Effects

In this section, we consider making inferences about the multiplicative causal effect of exposure A, under the model

$$(4.3) \qquad \frac{E(Y|A = a, G, U)}{E(Y|A = 0, G, U)} = \exp(\beta_a a),$$

where for simplicity, we assume no baseline covariates, binary A and scalar G. Therefore, If Y is binary, $\beta_a a$ encodes the conditional log risk ratio

$$\log\{\Pr(Y = 1|A = a, G, U)/\Pr(Y = 1|A = 0, G, U)\},$$

which is assumed to be independent of U and G, that is, there is no multiplicative interaction between A and (G, U). In order to state our identification result with an invalid IV, consider the following assumption.

ASSUMPTION 4a†. Equations (4.3) and

$$E(Y|A = 0, G, U) = \beta_g(U, G) + \beta_u(U),$$

$$E(A|G, U) = \alpha_g(U, G) + \alpha_u(U)$$

hold.

LEMMA 4.1. *Suppose that Assumptions 1, 2*, 4a† and 4b* hold, then β_a is the unique solution to equation*:

$$0 = U_{\text{mul}}(Y, A, G; \beta_a)$$

$$(4.4) \qquad \equiv E[\{G - E(G)\}\{A - E(A|G)\}Y$$

$$\times \exp\{-\beta_a A\}],$$

provided that $\partial U_{\text{mul}}(Y, A, G; \beta_a)/\partial \beta_a \neq 0$ at the truth.

The results follows upon noting that $E[Y \exp\{-\beta_a A\}| A, G, U] = E[Y|A = 0, G, U]$. The proof then proceeds as in Lemma 3.2.

According to Lemma 4.1, a consistent estimator of β_a can be obtained by solving an empirical version of equation (4.4) in a similar manner as in previous Sections. The unbiasedness property given by equation (4.4) continues to hold for continuous A under the conditions given in Lemma 4.1, and generalizations to allow for covariates

and multiple IVs can easily be deduced from previous Sections.

Interestingly, equation (4.4) continues to hold under case-control sampling with respect to the outcome Y, however note that $E(G)$ and $E(A|G)$ must be evaluated wrt the underlying distribution for the target population which will in general not match the corresponding distributions in the case-control sample. To use the result in practice, one would either need to obtain these quantities from an external source or one could alternatively approximate them with the corresponding data distribution in the controls (i.e., units with $Y = 0$) provided the outcome is sufficiently rare. In the event sampling fractions for cases and controls are available, one could in principle implement inverse-probability of sampling weights to consistently estimate $E(G)$ and $E(A|G)$. Unbiasedness under case-control sampling follows from noting that $f(A, G, U|Y = 1) \propto \Pr(Y = 1|A, G, U)f(A, G, U)$, and therefore

$$E[\{G - E(G)\}\{A - E(A|G)\}\exp\{-\beta_a A\}|Y = 1]$$

$$\propto E[\{G - E(G)\}\{A - E(A|G)\}$$

$$\times \exp\{-\beta_a A\}E(Y|A, G, U)]$$

$$= E[\{G - E(G)\}\{A - E(A|G)\}\exp\{-\beta_a A\}Y]$$

$$= 0,$$

where the last equality follows from Lemma 4.1. The approach therefore extends that proposed by Bowden and Vansteelandt (2011) who give a detailed study of IV inferences using G-estimation under case-control sampling, in order to account for potentially invalid IVs.

4.4 Multiplicative Exposure Model

A multiplicative exposure model may also be used for count or binary exposure under the following assumption:

ASSUMPTION 4†.

(4.a†) There is no additive $A - (U, G)$ interaction in model for $E(Y|A, G, U)$

$$(4.5) \quad E(Y|A = a, G, U) - E(Y|A = 0, G, U) = \beta_a a$$

and no additive $G - U$ interaction in model for $E(Y|A, G, U)$

$$(4.6) \quad \begin{aligned} &E(Y|A = 0, G = g, U) - E(Y|A = 0, G = 0, U) \\ &= \beta_g(g) \end{aligned}$$

for an unknown function $\beta_g(\cdot)$ that satisfies $\beta_g(0) = 0$

(4.b†) There is no multiplicative $G - U$ interaction in model for $E(A|G, U)$

$$(4.7) \qquad \log \frac{E(A|G = g, U)}{E(A|G = 0, U)} = \alpha_g(g)$$

for an unknown function $\alpha_g(\cdot)$ that satisfies $\alpha_g(0) = 0$.

MR GENIUS can be adapted to this setting according to the following result. Let

$$\log \frac{E(A|G=g)}{E(A|G=0)} = \varpi_g(g), \quad (4.8)$$

and $U_a = E(A|G=0, U)$.

LEMMA 4.2. *Suppose that Assumptions* 1, 2 *and* 4[†] *hold, then*

$$\beta_a = \frac{E[\{G - E(G)\}\{A\exp(-\varpi_g(G)) - E(A\exp(-\varpi_g(G)))\}Y]}{E[\{G - E(G)\}\{A\exp(-\varpi_g(G)) - E(A\exp(-\varpi_g(G)))\}A]}$$

provided that $\mathrm{var}(A|g)/\mathrm{var}(A|g=0) \neq \exp(\varpi_g(g))$ *for at least one value of* g.

The proof for Lemma 4.2 can be found in Appendix A.3. A consistent estimator of β_a is therefore obtained as in the previous section, by substituting in consistent estimators of unknown parameters and sample averages for expectations. To ground ideas, suppose that $\varpi_g(g) = \varpi_g^T g$ for vector ϖ_g, then a consistent estimator $\widehat{\varpi}_g$ of ϖ_g is given by the solution to the estimating equation:

$$\mathbb{P}_n\big[A\exp(-\widehat{\varpi}_g^T G)(G - \mathbb{P}_n G)\big] = 0.$$

Note that if A is a rare binary exposure then $\mathrm{var}(A|g)/\mathrm{var}(A|g=0) \approx \exp(\varpi_g(g))$ for all g, therefore violating the identification condition. In such instance, we recommend using the additive model described in the previous section. For count data, the result rules out using a Poisson model for exposure, however other models that accommodate over-dispersion such as the negative binomial distribution may be used. Finally, it is straightforward to verify that Lemma 4.2 continues to hold if Assumption 2 is dropped to allow for unmeasured confounding of the effects of G provided that the conditional covariance between the residual $(U_a/E(U_a|G) - 1)$ and U_y given G does not depend on G. Note that in this latter case $E(A|G=g) = \exp(\varpi_g(g)) = \exp(\alpha_g(g))E(U_a|G=g)$.

4.5 Odds Ratio Exposure Model

We briefly consider how MR GENIUS might be applied in a setting where Assumption 2 is replaced by the following weaker conditional independence assumption:

ASSUMPTION 2[††]. IV conditional independence: $G \perp\!\!\!\perp U | A = 0$.

A key implication of this assumption is that the causal effect of G on Y, is now identified conditional on A, because the assumption implies no unmeasured confounding of the effects of G on Y. Note however that G and U are not marginally independent. Suppose one wishes to encode the IV-exposure association on the odds ratio scale, under the following homogeneity assumption:

ASSUMPTION 4[††].

(4a[††]) Equations (4.3) and $E(Y|A=0, G, U) = \beta_g(G) + \beta_u(U)$ hold.

(4b[††]) There is no odds ratio $G - U$ interaction in model for $E(A|G, U)$

$$\mathrm{logit}\,\mathrm{Pr}(A=1|G=g, U) - \mathrm{logit}\,\mathrm{Pr}(A=1|G=0, U)$$
$$= \chi_g(g)$$

for an unknown function $\chi_g(\cdot)$ that satisfies $\chi_g(0) = 0$.

We then have the following identification result for the multiplicative causal effect β_a of model (4.3)

LEMMA 4.3. *Under Assumptions* 1, 2[††] *and* 4[††], *we have that* $\beta_a = \theta$, *where* θ *is the unique solution to equation*:

$$0 = E[\{G - E(G|A=0)\}\{A - E(A|G=0)\}Y$$
$$\times \exp\{-(\varphi_g(G) + \theta)A\}],$$

where

$$\varphi_g(g) = \mathrm{logit}\,\mathrm{Pr}(A=1|G=g) - \mathrm{logit}\,\mathrm{Pr}(A=1|G=0),$$

provided that $\gamma_{ag}(g) \neq 0$ *for some value of* g, *with*:

$$\gamma_{ag}(g) = E(Y|A=1, G=g, u)$$
$$- E(Y|A=1, G=0, u)$$
$$- E(Y|A=0, G=g, u)$$
$$+ E(Y|A=0, G=0, u) \neq 0.$$

The proof for Lemma 4.3 can be found in Appendix A.4. Assumption 2[††] in fact implies that $\varphi_g(\cdot) = \chi_g(\cdot)$ (Ma, Xie and Geng, 2006). Lemma 4.3 establishes that under Assumptions 1, 2[††] and 4[††], the multiplicative causal effect of A is identified, provided that $\gamma_{ag}(g) \neq 0$. In the proof of the lemma we establish that under our assumptions $\gamma_{ag}(g) = (\exp(\beta_a) - 1)\beta_g(g)$, and therefore the causal effect is not identified by the lemma if all IVs satisfy the exclusion restriction assumption, such that $\beta_g(g) = 0$ for all g. Note that the latter assumption is empirically testable because the direct effect of G on Y is unconfounded. If $\beta_g(g) \neq 0$ for some g, a valid test for the causal null hypothesis can be performed by testing whether the estimating equation given in the Lemma holds at $\theta = 0$. An estimator of β_a based on the lemma is easily deduced from previous sections.

4.6 MR GENIUS for Censored Failure Time Under a Multiplicative Survival Model

Censored time-to-event endpoints are common in MR studies and IV methods to address such data are increasingly of interest; recent contributions to this literature include Nie, Cheng and Small (2011), Tchetgen Tchetgen et al. (2015), Li, Fine and Brookhart (2015) and

Martinussen et al. (2017). While these methods have been shown to produce a consistent causal effect estimator encoded either on the scale of survival probabilities, or as a hazards ratio or hazards difference, leveraging a valid IV which satisfies Assumptions 1–3, they are not robust to violation of any of these assumptions. In this section, we briefly extend MR GENIUS to survival analysis under an additive hazards model. Thus, suppose now that Y is a time-to-event outcome which satisfies the following additive hazards model

$$(4.9) \quad h(y|A, U, G) = \beta_0(y) + \beta_a(y)A + \beta_g(y)G + \beta_u(y, U),$$

where $h(y|A, U, G)$ is the hazard function of Y evaluated at y, conditional on A, U and G, and the functions $(\beta_0(\cdot), \beta_a(\cdot), \beta_g(\cdot), \beta_u(\cdot, \cdot))$ are unrestricted. The model states that conditional on U, the effect of A on Y encoded on the additive hazards scale is linear in A for each y, although, the effect size $\beta_a(y)$ may vary with y. The model is quite flexible in the unobserved confounder association with the outcome $\beta_u(\cdot, \cdot)$, which is allowed to remain unrestricted at each time point y and across time points. This is the model considered by Tchetgen Tchetgen et al. (2015) who further assumed that $\beta_g(y) = 0$ for all y by the exclusion restriction Assumption 3. Here we do not make this assumption. As usually the case in survival analysis, Y is subject to right-censoring due to drop-out, and therefore instead of observing Y for all subjects, one observes $Y^* = \min(Y, X)$ and $\Delta = I(\min(Y, X) = Y)$, where X is an independent censoring time (i.e., independent of Y, A, G, U). Let $R(y) = I(Y^* \geq y)$ denote the at-risk process and $N(y) = I(Y^* \leq y, \Delta = 1)$ the counting process associated with failure time. As discussed in Martinussen et al. (2017), the additive hazards model (4.9) is particularly attractive because it implies a multiplicative survival model for the joint causal effect of A and G on Y:

$$\frac{\Pr(Y > y|A = a, G = g, U)}{\Pr(Y > y|A = 0, G = 0, U)}$$
$$= \exp\{-\mathbf{B}_a(y)a - \mathbf{B}_g(y)g\},$$

where $\mathbf{B}_a(y) = \int_0^y \beta_a(v)\, dv$, $\mathbf{B}_g(y) = \int_0^y \beta_g(v)\, dv$. Our objective is therefore to identify and estimate $\mathbf{B}_a(y)$. We have the following result which extends the result of Martinussen et al. (2017) in order to accommodate possible violation of the exclusion restriction assumption:

LEMMA 4.4. *Under Assumptions 1, 2 and equations (4.9) and (2.4), we have that for each y*

$$(4.10) \quad 0 = E\{W(y, \mathbf{B}_a(y), \mathbf{B}_g(y))\},$$

where

$$W(y, \mathbf{B}_a(y), \mathbf{B}_g(y))$$
$$= [dN(y) - d\mathbf{B}_a(y)A - d\mathbf{B}_g(y)G]$$

$$\times \exp\{\mathbf{B}_a(y)A + \mathbf{B}_g(y)G\}R(y)h(G, A),$$
$$h(G, A) = \begin{pmatrix} (G - E(G)) \\ (G - E(G))(A - E(A|G)) \end{pmatrix}.$$

The proof for Lemma 4.4 is given in Appendix A.5. As in Martinussen et al. (2017), the unbiasedness of equation $W(y, \mathbf{B}_a(y), \mathbf{B}_g(y))$ suggests a way of estimating the increments $(d\mathbf{B}_a(y), d\mathbf{B}_g(y))$ by solving an empirical version of equation (4.10) for each y with population expectations replaced by sample analogs, giving the following recursive estimator

$$(\widehat{\mathbf{B}}_a(y), \widehat{\mathbf{B}}_g(y))$$
$$= \int_0^y \mathbb{P}_n[\widehat{h}(A, G)' \exp\{\widehat{\mathbf{B}}_a(s^-)A + \widehat{\mathbf{B}}_g(s^-)G\}\, dN(s)]\widehat{\mathbb{M}}^{-1}(s),$$

where $\widehat{\mathbf{B}}_a(s^-)$ is the value of $\widehat{\mathbf{B}}_a$ right prior to s, and likewise for $\widehat{\mathbf{B}}_g(s^-)$, and

$$\widehat{h}(A, G) = \begin{pmatrix} (G - \widehat{E}(G)) \\ (G - \widehat{E}(G))(A - \widehat{E}(A|G)) \end{pmatrix},$$

$$\widehat{\mathbb{M}}(s) = \mathbb{P}_n\left[\begin{pmatrix} A \\ G \end{pmatrix} \widehat{h}(A, G)' R(s) \times \exp\{\widehat{\mathbf{B}}_a(s^-)A + \widehat{\mathbf{B}}_g(s^-)G\}\right].$$

Because of its recursive structure, this estimator can be solved forward in time starting with $(d\mathbf{B}_a(0), d\mathbf{B}_g(0)) = (0, 0)$. The resulting estimator is a counting process integral, therefore only changing values at observed event time. The estimator is only defined provided $\widehat{\mathbb{M}}(y)$ is invertible at each such jump time, which is essentially a necessary condition for identification. The large sample behavior of the resulting estimator follows from results derived in Martinussen et al. (2017) and is therefore omitted. Note that the result relies on Assumption 2 therefore ruling out confounding of the effect of the IV on the outcome.

4.7 More Efficient MR GENIUS

Similar to standard g-estimation, MR GENIUS can be made more efficient by incorporating information about the association between G and Y. This can be achieved by the following steps:

1. Obtain the MR GENIUS estimator $\widehat{\beta}_a$ either on the additive or multiplicative scale.
2. Define a treatment-free outcome $\widehat{Y}_0(\widehat{\beta}_a) = Y - \widehat{\beta}_a A$ or $\widehat{Y}_0(\widehat{\beta}_a) = Y \exp\{-\widehat{\beta}_a A\}$.
3. Regress $\widehat{Y}_0(\widehat{\beta}_a)$ on G using a generalized linear model with appropriate link function, and define $\widehat{\mu}(G)$ a person's corresponding fitted (predicted) value.

4. Define $\widehat{\beta}_a^{\mathrm{opt}}$ as the solution to

$$0 = \mathbb{P}_n[\{G - \mathbb{P}_n(G)\}\{A - \widehat{E}(A|G)\}\{\widehat{Y}_0(\widehat{\beta}_a^{\mathrm{opt}}) - \widehat{\mu}(G)\}]$$

with $\widehat{Y}_0(\widehat{\beta}_a^{\mathrm{opt}}) = Y - \widehat{\beta}_a^{\mathrm{opt}} A$ or $\widehat{Y}_0(\widehat{\beta}_a^{\mathrm{opt}}) = Y \exp\{-\widehat{\beta}_a^{\mathrm{opt}} A\}$.

If all regression models are correctly specified (including the glm for $E(Y_0(\beta_a)|G)$ required in Step 3 of the above procedure), a standard argument of semiparametric theory implies that the asymptotic variance of $\widehat{\beta}_a^{\mathrm{opt}}$ is guaranteed to be no larger than that of $\widehat{\beta}_a$ (Robins, 1997). Interestingly, MR GENIUS and its more efficient version coincide (up to asymptotic equivalence) whenever nonparametric methods are used to estimate all nuisance parameters, that is, to estimate $E(G)$, $E(A|G)$ and $\mu(G) = E(Y - \beta_a A|G)$. For instance, in the case of binary G, such that regression models $E(A|G)$ and $\mu(G) = E(Y - \beta_a A|G)$ are saturated, the two estimators are exactly equal and yield identical inferences. Both approaches also coincide if all IVs are valid, however the above modification will tend to be more efficient with increasing number of invalid IVs. Note that $\mu(G)$ does not necessarily have a causal interpretation as the effect of G on Y may be confounded by U. Also note that misspecification of a model for $\mu(G)$ does not affect consistency and asymptotic normality of the MR GENIUS estimator of β_a provided that as we have assumed throughout, the model for $E(A|G)$ is correct.

In the case of multiplicative outcome model, it is straightforward to extend the robustness properties of the efficient MR GENIUS estimator described above under an assumption of no multiplicative interaction (rather than no additive interaction) between G and U. This would simply entail replacing $\{\widehat{Y}_0(\widehat{\beta}_a^{\mathrm{opt}}) - \widehat{\mu}(G)\}$ in Step 4 with $\{\widehat{Y}_0(\widehat{\beta}_a^{\mathrm{opt}})\widehat{\mu}(0; \widehat{\beta}_a^{\mathrm{opt}})/\widehat{\mu}(G; \widehat{\beta}_a^{\mathrm{opt}})\}$, where $\widehat{\mu}(G; \widehat{\beta}_a^{\mathrm{opt}})$ is the regression of $Y \exp\{-\widehat{\beta}_a^{\mathrm{opt}} A\}$ on G under an appropriate GLM and solving the estimating equation in Step 4 for $\widehat{\beta}_a^{\mathrm{opt}}$. One can show using the same method of proof used throughout, that the resulting estimator is consistent for the causal effect of interest under violation of both Assumptions 2 and 3, under an assumption analogous to Assumption 2*. Note however that $\widehat{\mu}(g)/\widehat{\mu}(0)$ would now need to be consistent for $E(Y|A = 0, G = g)/E(Y|A = 0, G = 0)$. It is likewise possible to modify the above procedure to accommodate a multiplicative exposure model by substituting in $\{A \exp(-\widehat{\varpi}_g(G)) - \mathbb{P}_n(A \exp(-\widehat{\varpi}_g(G)))\}$ for $\{A - \widehat{E}(A|G)\}$ in Step 4.

5. SIMULATION STUDY

5.1 Single IV

We investigate the finite-sample properties of MR GENIUS proposed above and compare them with existing estimators under a variety of settings. For a single binary

IV G, we generate independent and identically distributed (G_i, U_i, A_i, Y_i), $i = 1, 2, \ldots, n$ as follows:

$$G_i \sim \mathrm{Bernoulli}(p = 0.5),$$
$$Y_i \sim \mathrm{N}(\alpha G_i + \beta A_i + U_i, 1^2),$$

where for binary exposure A,

$$\epsilon_i \sim \text{truncated N}(a = 0.2, b = 0.5, \mu = 0.35, \sigma^2 = 1^2),$$
$$U_i = \phi_b G_i + \epsilon_i,$$
$$A_i \sim \mathrm{Bernoulli}\Big(p_i = \frac{\exp(\gamma_b G_i)}{1 + \exp(\gamma_b G_i)}$$
$$+ U_i - E(U_i|G_i)\Big),$$

where ϵ_i is appropriately bounded to ensure that p falls in the unit interval, and for continuous A,

$$U_i = \phi_c G_i + N(0, 1^2),$$
$$A_i \sim \mathrm{N}(\gamma_c G_i + U_i, |\lambda_0 + \lambda_1 G_i|^2).$$

The data generating mechanism satisfies Assumptions 2* and 4*. We set $\gamma_b = -0.5$ or -1 (binary A), and $\gamma_c = -1$, $\lambda_0 = 1$, $\lambda_1 = 1$ or 5 (continuous A) which satisfy both Assumption 1 and condition (5). Assumptions 2 and 3 are violated when we set $\phi_b = -0.2$, $\phi_c = -2$ and $\alpha = -0.5$, respectively. The causal parameter is set equal to $\beta = 0.5$ throughout this simulation. The IV strength is tuned by varying the values of γ_b and λ_1, for binary and continuous A, respectively.

MR GENIUS is implemented as given in (3.2), with $\widehat{E}(A|G)$ estimated with linear or logistic regression when A is continuous or binary, respectively. In this single-IV setting, we also implement the two-stage least squares (TSLS) estimator, which is the most common approach used in practice. The simulation results based on 1000 replicates at sample sizes $n = 500$ and $n = 1000$ are summarized in Tables 1 and 2, for continuous and binary exposure, respectively. When Assumptions 2 and 3 both hold, TSLS and MR GENIUS have small bias regardless of sample size. When the IV is invalid, TSLS is biased while in accordance with theory MR GENIUS continues to have small bias.

5.2 Multiple IVs

Here we generate i.i.d. $L_i = (G_i, U_i, A_i, Y_i)$, $i = 1, 2, \ldots, n$, with $p_G = 10$ IVs from:

$$G_{ij} \sim \mathrm{Bernoulli}(p = 0.5), \quad j = 1, 2, \ldots, p_G,$$
$$Y_i \sim \mathrm{N}(\alpha^T G_i + \beta A_i + U_i, 1^2),$$

where $G_i = (G_{i1}, G_{i2}, \ldots, G_{ip_G})^T$. For binary exposure,

$$\epsilon_i \sim \text{Truncated N}(a = 0.2, b = 0.5, \mu = 0.35, \sigma^2 = 1^2),$$
$$U_i = \phi_b^T G_i + \epsilon_i,$$

TABLE 1
Monte Carlo results of MR GENIUS and TSLS estimation of $\beta_0 = 0.5$ with continuous exposure and single IV at two different strengths ($\lambda_1 = 1, 5$). The first and second rows' results for each estimator correspond to sample sizes $n = 500, 1000$, respectively

	$|\lambda_1| = 1$			$|\lambda_1| = 5$		
	TTT[†]	TTF	TFF	TTT	TTF	TFF
Median absolute value of bias						
MR GENIUS	0.001	0.001	0.001	0.001	0.001	0.001
	0.001	0.001	0.001	0.000	0.000	0.000
TSLS	0.001	0.501	0.835	0.000	0.507	0.837
	0.002	0.499	0.834	0.002	0.498	0.832
Monte Carlo standard error[‡]						
MR GENIUS	0.084	0.084	0.084	0.017	0.017	0.017
	0.060	0.060	0.060	0.011	0.011	0.011
TSLS	0.122	0.133	0.052	0.126	0.248	0.114
	0.093	0.085	0.036	0.094	0.152	0.080
Monte Carlo standard error/median estimated standard error						
MR GENIUS	1.003	1.003	1.003	1.037	1.037	1.037
	1.008	1.008	1.008	0.990	0.990	0.990
TSLS	0.951	1.068	1.058	0.925	1.121	1.019
	1.042	0.984	1.021	1.042	1.009	1.027
Coverage proportion[*]						
MR GENIUS	0.953	0.953	0.953	0.948	0.948	0.948
	0.948	0.948	0.948	0.952	0.952	0.952
TSLS	0.953	0.027	0.000	0.990	0.369	0.000
	0.959	0.010	0.000	0.977	0.058	0.000

TABLE 2
Monte Carlo results of MR GENIUS and TSLS estimation of $\beta_0 = 0.5$ with binary exposure and single IV at two different strengths ($\gamma_b = -0.5, -1$). The first and second rows' results for each estimator correspond to sample sizes $n = 500, 1000$ respectively

	$|\gamma_b| = 0.5$			$|\gamma_b| = 1$		
	TTT[†]	TTF	TFF	TTT	TTF	TFF
Median absolute value of bias						
MR GENIUS	0.006	0.006	0.009	0.030	0.030	0.014
	0.009	0.009	0.002	0.008	0.008	0.002
TSLS	0.002	4.050	2.162	0.001	2.171	1.613
	0.021	4.074	2.176	0.013	2.178	1.614
Monte Carlo standard error[‡]						
MR GENIUS	0.540	0.540	0.316	0.349	0.349	0.341
	0.389	0.389	0.224	0.245	0.245	0.238
TSLS	0.786	1.372	0.369	0.408	0.519	0.273
	0.550	1.065	0.278	0.293	0.374	0.182
Monte Carlo standard error/median estimated standard error						
MR GENIUS	1.002	1.002	0.970	0.968	0.968	1.003
	1.054	1.054	0.984	0.971	0.971	1.000
TSLS	1.024	0.984	0.981	1.034	0.986	1.097
	1.045	1.097	1.044	1.059	1.011	1.022
Coverage proportion[*]						
MR GENIUS	0.987	0.987	0.960	0.952	0.952	0.958
	0.969	0.969	0.964	0.959	0.959	0.949
TSLS	0.986	0.097	0.000	0.952	0.000	0.000
	0.964	0.003	0.000	0.946	0.000	0.000

[†]TTT: IV Assumptions 1–3 hold; TTF: IV Assumption 3 does not hold; TFF: both IV Assumptions 2 and 3 do not hold.
[‡]Robust normal-consistent estimate obtained from dividing the interquartile range of causal effect estimates by 1.349.
[*]Nominal level = 95%.

Note: See the footnote of Table 1.

$$A_i \sim \text{Bernoulli}\left(p_i = \frac{\exp(\gamma_b{}^T G_i)}{1 + \exp(\gamma_b{}^T G_i)} + [U_i - E(U|G_i)] \right),$$

where ϵ_i is appropriately bounded to ensure that p_i falls in the unit interval, and for continuous exposure,

$$U_i = \phi_c{}^T G_i + \text{N}(0, 1^2),$$
$$A_i \sim \text{N}(\gamma_c{}^T G_i + U_i, |\lambda_0 + \lambda_1^T G_i|^2).$$

For binary exposure, $\gamma_b \sim \text{Uniform}(-0.15, -0.05)$ so that IV strength is variable, while in the continuous exposure case $\gamma_c \sim \text{Uniform}(-3, -2)$ and (λ_0, λ_1) is set to $(1, 0.5)$. We first generate an ideal scenario in which all 10 IVs are valid and satisfy Assumptions 1–3, next we consider scenarios where the first three, six or all of the IVs are invalid. With three invalid IVs, $\alpha^T = -0.5 \cdot (1, 1, 1, 0, \ldots, 0)$ and $\phi_c^T = -0.25 \cdot (1, 1, 1, 0, \ldots, 0)$, $\phi_b^T = -0.05 \cdot (1, 1, 1, 0, \ldots, 0)$ when Assumption 3 or 2 is violated, respectively; with six invalid IVs, $\alpha^T = -0.25(1, 1, 2, 2, 4, 4, 0, \ldots, 0)$ and

$\phi_c^T = -0.25 \cdot (0.5, 0.5, 1, 1, 2, 2, 0, \ldots, 0)$, $\phi_b^T = -0.01 \cdot (1, 1, 3, 3, 5, 5, 0, \ldots, 0)$ accordingly. When all IVs are invalid, $\alpha \sim \text{Uniform}(-2, -0.5)$, $\phi_c \sim \text{Uniform}(-2, -0.5)$ and $\phi_b \sim \text{Uniform}(-0.02, -0.01)$. The setting with three invalid IVs investigates the condition in which fewer than 50% of the IVs are invalid (Kang et al., 2016, Windmeijer et al., 2019); in the setting with six invalid IVs this condition is violated, but the set of valid IVs form the largest group according to the plurality rule (Guo et al., 2018).

MR GENIUS is implemented as the solution to (4.2) with optimal weight; a more efficient version of MR GENIUS as described in Section 4.7 is also implemented. MR-Egger, TSLS and sisVIVE are implemented using the R packages MendelianRandomization, AER and sisVIVE (Yavorska and Staley, 2019, Kleiber and Zeileis, 2008, Kang, 2017) respectively, under default settings. The adaptive Lasso and TSHT estimation methods are implemented as described in Windmeijer et al. (2019) and Guo et al. (2018), respectively. We also implement post-adaptive Lasso which uses adaptive Lasso for the purpose of selecting valid IVs but not in the process of estimating the causal effect. We also implement the oracle TSLS which assumes the set of valid IVs to be known a priori.

TABLE 3

Simulation results for estimation of $\beta_0 = 0.5$ with continuous exposure and $p_G = 10$ IVs. The two rows of results for each estimator correspond to sample sizes of $n = 1000$ and $n = 2000$ respectively

		TTF†			TFF				TTF			TFF		
#invalid IV	0	3	6	10	3	6	10	0	3	6	10	3	6	10
	Median absolute value of bias							Monte Carlo standard error‡						
MR GENIUS	0.006	0.010	0.017	0.046	0.016	0.032	0.115	0.033	0.034	0.037	0.056	0.036	0.046	0.083
	0.001	0.003	0.007	0.022	0.006	0.015	0.063	0.023	0.024	0.025	0.043	0.026	0.034	0.072
Efficient MR GENIUS	0.006	0.006	0.008	0.013	0.008	0.011	0.030	0.035	0.035	0.034	0.032	0.035	0.036	0.040
	0.002	0.002	0.003	0.006	0.003	0.004	0.014	0.023	0.024	0.024	0.024	0.024	0.024	0.026
TSLS	0.000	0.057	0.133	0.476	0.121	0.264	0.656	0.011	0.012	0.012	0.055	0.013	0.015	0.041
	0.001	0.057	0.133	0.480	0.121	0.264	0.652	0.008	0.008	0.009	0.057	0.009	0.010	0.038
Oracle TSLS	–	0.000	0.000	–	0.000	0.000	–	–	0.013	0.019	–	0.013	0.019	–
	–	0.001	0.000	–	0.001	0.000	–	–	0.010	0.012	–	0.010	0.012	–
sisVIVE	0.001	0.030	0.106	0.476	0.035	0.223	0.654	0.011	0.014	0.021	0.082	0.013	0.027	0.052
	0.001	0.022	0.109	0.475	0.025	0.215	0.652	0.008	0.010	0.019	0.089	0.010	0.021	0.051
ALasso	0.001	0.024	0.094	0.477	0.026	0.207	0.655	0.011	0.014	0.029	0.093	0.013	0.042	0.059
	0.001	0.017	0.103	0.480	0.018	0.199	0.653	0.008	0.010	0.032	0.098	0.010	0.026	0.057
post-ALasso	0.001	0.003	0.101	0.478	0.001	0.203	0.654	0.011	0.016	0.049	0.104	0.013	0.048	0.067
	0.001	0.001	0.111	0.480	0.001	0.196	0.653	0.008	0.010	0.037	0.105	0.010	0.022	0.066
TSHT	0.001	0.009	0.062	0.477	0.000	0.107	0.653	0.011	0.023	0.031	0.107	0.013	0.071	0.068
	0.001	0.001	0.060	0.491	0.001	0.042	0.662	0.008	0.010	0.028	0.124	0.010	0.070	0.071
MR-Egger	0.017	0.008	0.017	0.113	0.406	0.822	0.545	0.063	0.139	0.206	0.338	0.196	0.213	0.250
	0.009	0.025	0.022	0.085	0.463	0.927	0.543	0.046	0.116	0.174	0.364	0.148	0.179	0.288
	Monte Carlo SE/median estimated SE							Coverage proportion*						
MR GENIUS	1.005	1.017	1.014	0.922	0.978	0.970	0.858	0.959	0.947	0.924	0.858	0.924	0.869	0.733
	0.997	0.992	0.955	0.952	0.967	0.962	0.940	0.962	0.963	0.946	0.901	0.951	0.917	0.822
Efficient MR GENIUS	1.066	1.070	1.058	0.973	1.076	1.105	1.163	0.942	0.938	0.934	0.918	0.932	0.920	0.811
	1.023	1.041	1.030	1.035	1.050	1.037	1.109	0.957	0.957	0.954	0.941	0.952	0.945	0.863
TSLS	1.018	1.054	1.069	3.230	1.143	1.195	2.877	0.942	0.000	0.000	0.000	0.000	0.000	0.000
	1.018	1.060	1.037	4.717	1.084	1.075	3.749	0.954	0.000	0.000	0.000	0.000	0.000	0.000
Oracle TSLS	–	0.998	1.090	–	0.998	1.090	–	–	0.940	0.948	–	0.940	0.948	–
	–	1.048	1.000	–	1.048	1.000	–	–	0.950	0.946	–	0.950	0.946	–
post-ALasso	1.025	1.213	2.871	4.583	1.026	2.179	3.678	0.938	0.898	0.051	0.000	0.935	0.000	0.000
	1.026	1.083	2.578	5.734	1.068	1.219	4.592	0.951	0.943	0.008	0.000	0.950	0.000	0.000
TSHT	1.015	1.835	2.125	5.592	0.984	2.687	4.420	0.944	0.717	0.159	0.000	0.940	0.294	0.000
	1.016	1.060	1.665	6.962	1.059	6.053	5.237	0.956	0.945	0.126	0.000	0.949	0.382	0.000
MR-Egger	0.456	0.758	0.718	1.113	0.678	0.622	0.956	1.000	0.985	0.981	0.887	0.758	0.283	0.473
	0.424	0.580	0.543	1.083	0.470	0.501	1.039	1.000	1.000	0.999	0.889	0.787	0.159	0.505

†For the invalid IVs, TTF: IV Assumption 3 does not hold; TFF: both IV Assumptions 2 and 3 do not hold.

‡Robust normal-consistent estimate obtained from dividing the interquartile range of causal effect estimates by 1.349.

*Nominal level = 95%.

Simulation results based on 1000 replications for sample sizes of $n = 1000$ and 2000 with continuous exposure are presented in Table 3. When there are zero or three invalid IVs (majority rule holds), the sisVIVE, adaptive, post-adaptive Lasso and TSHT estimators exhibit small bias which becomes negligible at sample size of $n = 2000$. Adaptive Lasso and TSHT on average correctly identifies invalid IVs, while sisVIVE on average selects four IVs as invalid when there are three in truth (see Table 4 for results on IV selection). The naive TSLS estimator performs well in terms of bias only when all IVs are valid; as expected, it is biased in all other settings with at least one invalid IV. Post-adaptive Lasso is generally less biased in finite sample than adaptive Lasso. Post-adaptive Lasso and oracle TSLS perform similarly in terms of bias and efficiency when the majority rule holds, in agreement with theory since they are asymptotically equivalent under these settings (Windmeijer et al., 2019). MR GENIUS also has small bias at all sample sizes and its bias becomes negligible at $n = 2000$. When six IVs are invalid and the majority rule is violated, sisVIVE and adaptive/post-adaptive Lasso are significantly biased, with no improvement as sample size increases. On average, sisVIVE and adaptive Lasso select 7 to 8 IVs as invalid when only six are actually invalid, and fails to select all the IVs as invalid when all in fact are. TSHT is also biased when all IVs are

TABLE 4
Average number of IVs selected as invalid by adaptive Lasso and sisVIVE, and average number of IVs selected as relevant ($\hat{S}$) and relevant but invalid ($\hat{I}$) by TSHT. The two rows of results for each estimator correspond to sample sizes of $n = 1000$ and $n = 2000$ respectively

#invalid IV	0	TTF†			TFF			0	TTF†			TFF		
		3	6	10	3	6	10		3	6	10	3	6	10
			Continuous exposure							Binary exposure				
ALasso	0.0	3.1	5.5	4.8	3.0	7.3	4.4	0.1	3.2	5.0	1.0	3.1	5.1	0.9
	0.0	3.0	6.8	5.9	3.0	7.8	5.5	0.1	3.1	5.0	1.0	3.1	5.1	0.9
sisVIVE	0.0	3.8	7.1	6.2	4.2	7.9	5.6	0.0	3.8	5.0	1.1	3.8	5.0	1.1
	0.0	3.9	7.8	7.3	4.2	8.3	6.8	0.0	3.7	5.1	1.3	3.7	5.2	1.4
TSHT ($\hat{I}$)	0.0	2.4	4.5	3.4	3.0	6.8	2.7	0.0	0.7	0.4	0.0	0.7	0.5	0.1
	0.0	3.0	7.1	5.7	3.0	5.6	5.1	0.0	0.6	0.5	0.1	0.6	0.5	0.1
TSHT ($\hat{S}$)	10.0	10.0	10.0	10.0	10.0	10.0	10.0	4.7	4.7	4.7	4.7	4.7	4.7	4.5
	10.0	10.0	10.0	10.0	10.0	10.0	10.0	3.2	3.2	3.2	3.2	3.2	3.2	3.4

†For the invalid IVs, TTF: IV Assumption 3 does not hold; TFF: both IV Assumptions 2 and 3 do not hold.

invalid (with about 5 of the IVs selected as invalid on average in this case); however when six IVs are invalid, the plurality rule holds and its bias diminishes at $n = 2000$. The efficiency of all estimators generally decreases with increasing number of invalid IVs.

The bias of MR GENIUS improves with increasing sample size when six or all IVs are invalid. The efficient MR GENIUS is generally less biased and more efficient compared to MR GENIUS, especially when more IVs are invalid. MR-Egger shows little bias when any or all of the IVs are invalid, provided only Assumption 3 (exclusion restriction) is violated, in agreement with theory. However, MR-Egger is generally more biased when the invalid IVs violate both Assumptions 2 and 3, which corresponds to a violation of the InSIDE assumption (Bowden, Davey Smith and Burgess, 2015).

Simulation results with a binary exposure are reported in Table 5; the conclusions are mostly qualitatively similar to those in the continuous exposure setting. However, when there are six invalid IVs, TSHT is biased with no improvement as sample size increases. While the exposure is generated under a logit model (upon marginalizing over U), TSHT assumes a linear model which is misspecified in this simulation study. In addition, because the exposure is binary, most if not all IVs are weakly associated with A on the additive scale. Weak IVs may not be selected as valid IVs in the first thresholding step of TSHT (the number of IVs selected as relevant is 3.2 on average at $n = 2000$); even if they are included, their inclusion may lead to incorrect inference in the subsequent estimation step (the number of IVs selected as relevant but invalid is close to 0.5 on average at sample size of $n = 2000$, when in fact 6 are invalid). MR-Egger also appears to exhibit more bias, since the exposure model is misspecified as the linear probability model. Additional simulation results under the null ($\beta_0 = 0$) are reported in the Supplementary material (Tchetgen Tchetgen, Sun and Walter, 2021).

6. DATA APPLICATION

The prevalence of type 2 diabetes mellitus is increasing across all age groups in the United States possibly as a consequence of the obesity epidemic. Many epidemiological studies have suggested that individuals with type 2 diabetes mellitus (T2D) are at higher risk of various memory impairments which are highly associated with dementia and Alzheimer's Disease. However, such observational studies are well known to be vulnerable to confounding bias. Therefore, obtaining an unbiased estimate of the association between diabetes status and cognitive functioning is key to predicting the future health burden in the population and to evaluating the effectiveness of possible public health interventions.

In order to illustrate the proposed MR approach, we used data from the Health and Retirement Study, a cohort initiated in 1992 with repeated assessments every 2 years. We used externally validated genetic predictors of type 2 diabetes as IVs to estimate effects on memory functioning among HRS participants. The Health and Retirement Study is a well-documented nationally representative sample of persons aged 50 years or older and their spouses (Juster and Suzman, 1995). Genotype data were collected on a subset of respondents in 2006 and 2008. Genotyping was completed on the Illumina Omni-2.5 chip platform and imputed using the 1000G phase 1 reference panel and filed with the Database for Genotypes and Phenotypes (dbGaP, study accession number: phs000428.v1.p1) in April 2012. Exact information on the process performed for quality control is available via Health and Retirement Study and dbGaP21 (Mailman et al., 2007). From the 12,123 participants for whom genotype data was available, we restricted the sample to 7738 non-hispanic white persons with valid self-reported diabetes status at baseline and memory assessment score two years later. Self-reported diabetes in the Health and

TABLE 5

Simulation results for estimation of $\beta_0 = 0.5$ with binary exposure and $p_G = 10$ IVs. The two rows of results for each estimator correspond to sample sizes of $n = 1000$ and $n = 2000$ respectively

#invalid IV	0	TTF† 3	6	10	TFF 3	6	10	0	TTF 3	6	10	TFF 3	6	10
				Median absolute value of bias							Monte Carlo standard error‡			
MR GENIUS	0.068	0.075	0.233	0.631	0.076	0.250	0.689	0.910	0.996	1.238	2.231	1.012	1.258	2.207
	0.000	0.084	0.205	0.661	0.089	0.211	0.658	0.847	0.933	1.135	2.218	0.945	1.167	2.102
Efficient MR GENIUS	0.020	0.034	0.053	0.119	0.037	0.054	0.086	0.897	0.912	0.930	1.013	0.914	0.931	0.992
	0.021	0.010	0.0001	0.069	0.010	0.005	0.037	0.824	0.823	0.821	0.860	0.823	0.823	0.811
TSLS	0.041	1.704	5.880	20.241	1.866	6.199	20.381	0.513	1.822	2.939	7.351	1.979	3.093	7.242
	0.015	2.306	8.317	28.526	2.530	8.73	28.220	0.437	1.640	2.786	7.525	1.832	2.948	7.607
Oracle TSLS	–	0.051	0.089	–	0.051	0.089	–	–	0.593	0.831	–	0.593	0.831	–
	–	0.031	0.038	–	0.031	0.038	–	–	0.509	0.657	–	0.509	0.657	–
sisVIVE	0.041	0.533	3.802	20.122	0.523	3.932	20.256	0.509	0.775	3.438	7.309	0.773	3.651	7.283
	0.015	0.452	5.700	28.215	0.450	5.950	28.080	0.437	0.590	4.129	7.764	0.590	4.381	7.932
ALasso	0.041	0.368	2.041	20.086	0.364	2.039	20.289	0.510	0.617	2.939	7.354	0.597	3.129	7.173
	0.016	0.308	3.457	28.269	0.304	3.588	27.999	0.425	0.461	4.081	7.483	0.455	4.344	7.853
post-ALasso	0.041	0.098	1.600	20.019	0.092	1.575	19.952	0.501	0.617	2.676	7.501	0.603	2.819	7.119
	0.015	0.020	2.786	28.154	0.022	2.776	27.715	0.428	0.504	3.772	7.567	0.503	3.913	8.055
TSHT	0.026	0.270	3.681	17.799	0.245	3.759	17.496	0.646	1.462	5.575	8.885	1.348	5.931	7.962
	0.004	0.065	3.352	22.412	0.058	3.565	22.045	0.558	0.699	7.210	9.625	0.702	7.570	9.443
MR-Egger	0.012	1.213	4.805	15.051	1.341	5.061	16.086	0.929	4.382	7.644	15.343	4.803	8.080	16.189
	0.040	0.316	5.695	16.037	0.349	6.008	16.059	0.821	5.065	9.235	16.762	5.611	9.722	19.136
				Monte Carlo SE/median estimated SE							Coverage proportion*			
MR GENIUS	0.860	0.863	0.914	0.851	0.862	0.912	0.845	0.990	0.990	0.982	0.977	0.990	0.983	0.982
	0.950	0.965	0.967	0.975	0.964	0.973	0.933	0.980	0.977	0.976	0.965	0.977	0.975	0.967
Efficient MR GENIUS	0.846	0.858	0.878	0.919	0.860	0.879	0.913	0.993	0.992	0.993	0.993	0.992	0.993	0.992
	0.923	0.921	0.919	0.949	0.920	0.921	0.913	0.981	0.981	0.980	0.980	0.981	0.980	0.974
TSLS	0.961	2.499	1.940	1.494	2.598	1.954	1.511	0.983	0.353	0.083	0.019	0.330	0.082	0.022
	0.970	2.434	1.573	1.250	2.562	1.586	1.299	0.974	0.201	0.013	0.000	0.183	0.012	0.002
Oracle TSLS	–	0.956	0.928	–	0.956	0.928	–	–	0.985	0.995	–	0.985	0.995	–
	–	0.998	0.924	–	0.998	0.924	–	–	0.976	0.988	–	0.976	0.988	–
post-ALasso	0.937	0.978	2.366	1.512	0.959	2.469	1.465	0.985	0.976	0.632	0.018	0.979	0.631	0.023
	0.948	0.987	3.203	1.255	0.985	3.263	1.373	0.976	0.970	0.485	0.004	0.971	0.491	0.004
TSHT	0.942	1.751	4.088	1.561	1.619	4.230	1.372	0.990	0.728	0.382	0.013	0.734	0.370	0.012
	0.908	1.037	6.211	1.511	1.040	6.219	1.506	0.985	0.883	0.446	0.000	0.885	0.446	0.002
MR-Egger	0.886	1.044	1.034	0.947	1.042	1.039	1.005	0.977	0.919	0.872	0.846	0.920	0.871	0.829
	0.891	0.998	1.076	0.998	1.004	1.078	1.138	0.960	0.927	0.866	0.833	0.923	0.866	0.836

Note: See the footnote of Table 3.

Retirement Study has been shown to have 87% sensitivity and 97% specificity for Hemoglobin A1c defined diabetes among non-Hispanic white HRS participants (White et al., 2014). Memory was assessed by immediate and delayed recall of a 10-word list plus the proxy assessments for severely impaired individuals. The validity and reliability of these measures have been documented elsewhere (Ofstedal, Fisher and Herzog, 2005, Wu et al., 2012).

Standard MR relies on the assumption that all 39 SNPs affect a person's memory score at follow-up only through baseline diabetes status which is unlikely, even if all 39 SNPs only affect memory through diabetes. This is because there is likely to be a nonnegligible direct effect from one of the SNPs to diabetes incidence among persons who are diabetes-free at baseline. This would constitute a violation of the exclusion restriction and therefore would invalidate a standard MR analysis for assessing effects of baseline diabetes on memory score at followup. Nonetheless, although possibly positively biased under the alternative hypothesis, the two-stage regression estimator could still be interpreted as a valid test of the null hypothesis of no association between diabetes disease (whether baseline or time-updated) and memory score. It may also be true that unknown pleiotropic effects of at least one of the SNPs exists through a pathway not involving diabetes, which would constitute an even more serious violation, as it would also invalidate our MR analysis as a valid test of a causal association between diabetes and

memory functioning. In light of these possible limitations, a more robust MR analysis is naturally of interest.

We used GENIUS to estimate the relationship between diabetes status (coded 1 for diabetic and 0 otherwise) and memory score. As genetic instruments, we used 39 independent single nucleotide polymorphisms previously established to be significantly associated with diabetes (Morris et al., 2012).

We first performed an observational analysis, which entailed fitting a linear model with memory score as outcome, diabetes status as exposure, adjusting for age at cognitive assessment and sex. Next, we implemented an MR analysis of the effects of diabetes status on cognitive score incorporating all 39 SNPs as candidate IV using TSLS, sisVIVE, adaptive LASSO, TSHT, MR Egger, and the proposed GENIUS approaches.

Participants were, on average, 68.1 years old (standard deviation [SD] = 10.1 years old) at baseline and 1282 of them self-reported that they had diabetes (16.7%). The 39 SNPs jointly included in a first-stage logistic regression model to predict diabetes status explained 3.5% (Nagelkerke R^2) of the variation in diabetes in the study sample. The estimated coefficients from first-stage logistic regression of type 2 diabetes status on the 39 SNPs, as well as memory score-SNP linear regression coefficient estimates, are displayed as a scatter plot in Figure 2.

Table 6 shows results from both observational and IV analyses. In the observational analysis, being diabetic was associated with an average decrease of 0.04 points (s.e. = 0.02) in memory score. MR GENIUS suggests a notably larger diabetes-associated decrease in average memory score equal to 0.18 points (s.e. = 0.14). The efficient MR GENIUS produced a similar decrease of 0.16 points (s.e. = 0.14). MR-Egger produced an estimate suggesting a protective effect of diabetes (beta = 0.25, s.e. = 0.35) and so did TSLS (beta = 0.48, s.e. = 0.22), sisVIVE (beta = 0.48) and adaptive lasso (beta = 0.48, s.e. = 0.22) which gave the same point estimate, while TSHT (beta =
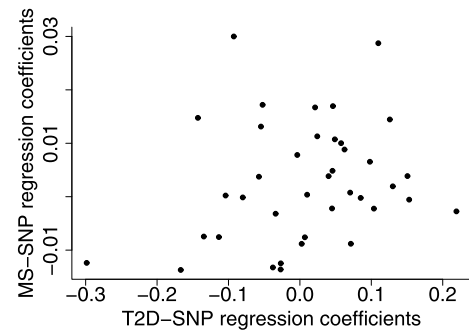
FIG. 2. *Scatter plot of memory score-SNP (MS-SNP) joint linear regression coefficient estimates versus type 2 diabetes status-SNP (T2D-SNP) joint logistic regression coefficient estimates for the 39 SNPs.*

0.45, s.e. = 0.28) gave a slightly smaller but still protective estimate. TSLS, sisVIVE and adaptive lasso inferences coincide exactly in this application because all 39 candidate SNPs ended up being selected as "valid" by sisVIVE and adaptive lasso. In contrast, TSHT selected six candidate IVs only as both valid and relevant which were therefore used to estimate the causal effect. In conclusion, both the observational analysis and MR GENIUS found some evidence of a harmful effect of diabetes on memory score, which supports the prevailing hypothesis in the diabetes literature. In contrast, all other (robust and non-robust) MR methods suggest a protective effect of diabetes on memory, a hypothesis with little if any scientific basis in the diabetes literature. However, the results from MR GENIUS, MR-Egger and TSHT are not statistically significant at the 95% confidence level.

We note that while TSLS is asymptotically efficient with a valid IV under linear models and the conditional homokesdasticity condition that $\text{cov}\{(A, Y)^T | G\}$ is constant in G (Wooldridge, 2010, Vansteelandt and Didelez, 2018), the latter is typically violated with binary exposure. Furthermore, the simulation study results in Table 5 suggest that MR GENIUS generally has smaller variance

TABLE 6
Estimation of $\beta_{t2d\text{-}ms}$, the association between type 2 diabetes and memory score

	$\hat{\beta}_{t2d\text{-}ms}$	SE	95% CI	# of instruments selected as invalid
Observational analysis				
	−0.04	0.02	(−0.08, 0.001)	−
IV analyses				
MR GENIUS	−0.18	0.14	(−0.45, 0.08)	−
Efficient MR GENIUS	−0.16	0.14	(−0.43, 0.11)	−
MR-Egger	0.25	0.35	(−0.43, 0.93)	−
sisVIVE	0.48	−	−	0
TSLS	0.48	0.22	(0.05, 0.90)	−
Adaptive Lasso	0.48	−	−	0
Post-adaptive Lasso	0.48	0.22	(0.05, 0.90)	0
TSHT	0.45	0.28	(−0.10, 1.00)	0 (out of 6 selected as relevant)

than TSLS when exposure is binary and at least some of the potential IVs in G are invalid.

7. DISCUSSION

As MR gains popularity as a promising strategy to address confounding bias in observational studies, there clearly also is a growing need for robust MR methodology that relax the standard IV assumptions. Although a variety of methods have recently been proposed, we have argued that MR GENIUS stands out as an effective approach with clear advantages over other existing methods. Whereas existing methods are technically only consistent either as the number of candidate IVs goes to infinity (MR-Egger), or as a majority (adaptive lasso) or a plurality (TSTH) of IVs are valid, MR GENIUS is guaranteed to be consistent without even one valid IV. An R package which implements MR GENIUS is available at https://github.com/bluosun/MR-GENIUS.

In closing, we acknowledge certain limitations of MR GENIUS. It is important to check how strongly $\mathrm{var}(A|G)$ varies with G in practice, for example by performing the Breusch–Pagan test (Breusch and Pagan, 1979) when exposure is continuous, because identification via heteroscedasticity may yield noisier estimates under certain settings compared to standard first-order moment exclusion restrictions. This may be less of an issue if one has available a moderate to large number of candidate IVs for the exposure of interest, such as in MR studies, to increase precision. The 39 SNPs considered in the application were strongly associated as a set with the binary endogenous variable (Likelihood ratio test Chi-square statistic = 162 with 39 degrees of freedom, which corresponds to a significance value <0.001). This provides fairly compelling evidence that the IVs are not only jointly relevant but also satisfy the first stage heteroscedasticity condition required by MR GENIUS. Secondly, the approach may be vulnerable to weak instrument bias which may occur if $\mathrm{var}(A|G_j)$ is weakly dependent on each candidate IV G_j, a possibility that was largely ruled out in this paper. Estimation of causal effects with many weak invalid IVs is the subject of a companion manuscript currently in preparation. Finally, MR GENIUS is also currently not designed to handle high dimensional IVs (where the number of IVs may exceed sample size). We plan to further develop MR GENIUS to address all of these remaining challenges in future work.

APPENDIX A: PROOFS OF LEMMAS

A.1 Proof of Lemma 3.1

PROOF. Under Assumption 4a and taking iterated expectation with respect to (A, G, U) followed by (G, U),

$$E[\{G - E(G)\}\{A - E(A|G)\}Y]$$

$$= E[\{G - E(G)\}\{A - E(A|G)\}E(Y|A, G, U)]$$

$$= E[\{G - E(G)\}\{A - E(A|G)\}\{\beta_a(U)A$$
$$+ \beta_g(U, G) + \beta_u(U)\}]$$

$$= E[\{G - E(G)\}\{A - E(A|G)\}\beta_a(U)A]$$
$$+ E[\{G - E(G)\}\{A - E(A|G)\}\beta_g(U, G)]$$
$$+ E[\{G - E(G)\}\{A - E(A|G)\}\beta_u(U)]$$

$$= E[\{G - E(G)\}\{1 - E(A|G)\}\beta_a(U)$$
$$\times \{\alpha_g(U, G) + \alpha_u(U)\}]$$
$$+ E[\{G - E(G)\}\{\alpha_g(U, G) + \alpha_u(U) - E(A|G)\}$$
$$\times \beta_g(U, G)]$$
$$+ E[\{G - E(G)\}\{\alpha_g(U, G) + \alpha_u(U) - E(A|G)\}$$
$$\times \beta_u(U)]$$

$$= E[\{G - E(G)\}\{A - E(A|G)\}A E\{\beta_a(U)|G\}]$$
$$+ E[\{G - E(G)\}\{1 - E(A|G)\}$$
$$\times \mathrm{cov}\{\beta_a(U), \alpha_u(U)|G\}]$$
$$+ E[\{G - E(G)\}\{1 - E(A|G)\}$$
$$\times \mathrm{cov}\{\beta_a(U), \alpha_g(U, G)|G\}]$$
$$+ E[\{G - E(G)\} \mathrm{cov}\{\alpha_g(U, G), \beta_g(U, G)|G\}]$$
$$+ E[\{G - E(G)\} \mathrm{cov}\{\alpha_u(U), \beta_g(U, G)|G\}]$$
$$+ E[\{G - E(G)\} \mathrm{cov}\{\alpha_g(U, G), \beta_u(U)|G\}]$$
$$+ E[\{G - E(G)\} \mathrm{cov}\{\alpha_u(U), \beta_u(U)|G\}].$$

Under Assumption 2, $E\{\beta_a(U)|G\} = E\{\beta_a(U)\}$ and $\mathrm{cov}\{\alpha_u(U), \beta_u(U)|G\} = \mathrm{cov}\{\alpha_u(U), \beta_u(U)\}$, so that $E[\{G - E(G)\} \mathrm{cov}\{\alpha_u(U), \beta_u(U)|G\}] = 0$. Therefore, under Assumption 4b,

$$\frac{E[\{G - E(G)\}\{A - E(A|G)\}Y]}{E[\{G - E(G)\}\{A - E(A|G)\}A]} = E\{\beta_a(U)\}$$

provided that $E[\{G - E(G)\}\{A - E(A|G)\}A] = E[\{G - E(G)\} \mathrm{var}(A|G)] \neq 0$. □

A.2 Proof of Lemma 3.2

PROOF. Under Assumption 4* and taking iterated expectation with respect to (A, G, U) followed by (G, U),

$$E[\{G - E(G)\}\{A - E(A|G)\}Y]$$

$$= E[\{G - E(G)\}\{A - E(A|G)\}E(Y|A, G, U)]$$

$$= E[\{G - E(G)\}\{A - E(A|G)\}$$
$$\times \{\beta_a A + \beta_g(U, G) + \beta_u(U)\}]$$

$$= E[\{G - E(G)\}\{A - E(A|G)\}A]\beta_a$$
$$+ E[\{G - E(G)\} \mathrm{cov}\{\alpha_g(U, G), \beta_g(U, G)|G\}]$$
$$+ E[\{G - E(G)\} \mathrm{cov}\{\alpha_u(U), \beta_g(U, G)|G\}]$$

$$+ E\big[\{G - E(G)\} \operatorname{cov}\{\alpha_g(U, G), \beta_u(U)|G\}\big]$$
$$+ E\big[\{G - E(G)\} \operatorname{cov}\{\alpha_u(U), \beta_u(U)|G\}\big].$$

The proof for Lemma 3.2 follows from the observation that instead of requiring Assumption 2, we just need $\operatorname{cov}\{\alpha_u(U), \beta_u(U)|G\} = \rho$ so that

$$E\big[\{G - E(G)\} \operatorname{cov}\{\alpha_u(U), \beta_u(U)|G\}\big] = 0,$$

and hence

$$\frac{E[\{G - E(G)\}\{A - E(A|G)\}Y]}{E[\{G - E(G)\}\{A - E(A|G)\}A]} = \beta_a. \qquad \square$$

A.3 Proof of Lemma 4.2

PROOF. The proof follows upon noting that under our assumptions,

$$\exp(\varpi_g(G)) E\big(A \exp(-\varpi_g(G))\big)$$
$$= E(A|G)$$
$$= \exp(\alpha_g(G)) E(U_a),$$

and

$$E(A|G, U) - \exp(\varpi_g(G)) E\big(A \exp(-\varpi_g(G))\big)$$
$$= [U_a - E(U_a)] \exp(\alpha_g(G)).$$

Therefore

$$
\begin{aligned}
&E\big[\{G - E(G)\} \\
&\quad \times \{A \exp(-\varpi_g(G)) - E(A \exp(-\varpi_g(G)))\}Y\big] \\
&= E\big[\{G - E(G)\} \\
&\quad \times \{A \exp(-\varpi_g(G)) - E(A \exp(-\varpi_g(G)))\}\beta_a A\big] \\
&\quad + E\big[\{G - E(G)\} \\
&\quad \times \{A \exp(-\varpi_g(G)) - E(A \exp(-\varpi_g(G)))\}\beta_u(U)\big] \\
&\quad + E\big[\{G - E(G)\} \\
&\quad \times \{A \exp(-\varpi_g(G)) - E(A \exp(-\varpi_g(G)))\}\beta_g(G)\big] \\
&= \beta_a E\big[\{G - E(G)\} \\
&\quad \times \{A \exp(-\varpi_g(G)) - E(A \exp(-\varpi_g(G)))\}A\big] \\
&\quad + \underbrace{E\big[\{G - E(G)\}[U_a - E(U_a)]\beta_u(U)\big]}_{=0} \\
&\quad + \underbrace{E\big[\{G - E(G)\}\{U_a - E(U_a)\}\beta_g(G)\big]}_{=0} \\
&= \beta_a E\big[\{G - E(G)\}\{A - E(A|G)\}A \exp(-\varpi_g(G))\big] \\
&= \beta_a E\big[\{G - E(G)\} \operatorname{var}(A|G) \exp(-\varpi_g(G))\big],
\end{aligned}
$$

where we used the fact that under Assumption 2, $\varpi_g(g) = \alpha_g(g)$, therefore proving identification provided that $\operatorname{var}(A|G) \exp(-\varpi_g(G))$ is a function of G, which holds as long as $\operatorname{var}(A|g)/\operatorname{var}(A|g = 0) \neq \exp(\varpi_g(g))$. $\square$

A.4 Proof of Lemma 4.3

PROOF. We first note that for any additive function $t(A, G) = t_1(A) + t_2(G)$,

$$E\big(t(A, G)\{G - E(G|A = 0)\}\{A - E(A|G = 0)\}$$
$$\times \exp\{-\varphi_g(G)A\}\big) = 0$$

because

$$
\begin{aligned}
&E\big(t(A, G)\{G - E(G|A = 0)\}\{A - E(A|G = 0)\} \\
&\quad \times \exp\{-\varphi_g(G)A\}\big) \\
&= \sum_{a,g} f(a, g) t(a, g)\{g - E(G|A = 0)\} \\
&\quad \times \{a - E(A|G = 0)\} \exp\{-\varphi_g(g)a\} \\
&\propto \sum_{a,g} \{f(a|g = 0) f(g|a = 0) \exp\{\varphi_g(g)a\} t(a, g) \\
&\quad \times \{g - E(G|A = 0)\}\{a - E(A|G = 0)\} \\
&\quad \times \exp\{-\varphi_g(g)a\}\} \\
&= \sum_{a,g} f(a|g = 0) f(g|a = 0) t(a, g) \\
&\quad \times \{g - E(G|A = 0)\}\{a - E(A|G = 0)\} \\
&= 0,
\end{aligned}
$$

where we used the fact that

$$f(a, g) \propto (a|g = 0) f(g|a = 0) \exp\{\varphi_g(g)a\},$$

see, for example, Tchetgen Tchetgen, Robins and Rotnitzky (2010). It is straightforward to verify that the

$$
\begin{aligned}
\theta &= -\ln\big(1 - E[\{G - E(G|A = 0)\}\{A - E(A|G = 0)\}Y \\
&\quad \times \exp\{-\varphi_g(G)A\}](E[\{G - E(G|A = 0)\} \\
&\quad \times \{A - E(A|G = 0)\}AY \exp\{-\varphi_g(G)A\}])^{-1}\big).
\end{aligned}
$$

Next,

$$
\begin{aligned}
&E[\{G - E(G|A = 0)\} \\
&\quad \times \{A - E(A|G = 0)\}Y \exp\{-\varphi_g(G)A\}] \\
&= E[\{G - E(G|A = 0)\}\{A - E(A|G = 0)\} \\
&\quad \times \exp(\beta_a A) E(Y|A = 0, G, U) \exp\{-\varphi_g(G)A\}] \\
&= E[\{G - E(G|A = 0)\}\{A - E(A|G = 0)\}(\exp(\beta_a) - 1) \\
&\quad \times A E(Y|A = 0, G, U) \exp\{-\varphi_g(G)A\}] \\
&\quad + E[\{G - E(G|A = 0)\}\{A - E(A|G = 0)\} E(Y|A = 0, G, U) \exp\{-\varphi_g(G)A\}] \\
&= (\exp(\beta_a) - 1) E[\{G - E(G|A = 0)\}\{A - E(A|G = 0)\} \\
&\quad \times A E(Y|A = 0, G, U) \exp\{-\varphi_g(G)A\}] \\
&\quad + E[\{G - E(G|A = 0)\}\{A - E(A|G = 0)\}(E(Y|A = 0, G, U) - E(Y|A = 0, G = 0, U)) \\
&\quad \times \exp\{-\varphi_g(G)A\}] \\
&\quad + E[\{G - E(G|A = 0)\}\{A - E(A|G = 0)\}(E(Y|A = 0, G = 0, U)) \exp\{-\varphi_g(G)A\}] \\
&= (\exp(\beta_a) - 1) E[\{G - E(G|A = 0)\}\{A - E(A|G = 0)\} \\
&\quad \times A E(Y|A = 0, G, U) \exp\{-\varphi_g(G)A\}] \\
&\quad + \underbrace{E[\{G - E(G|A = 0)\}\{A - E(A|G = 0)\}\beta_g(G) \exp\{-\varphi_g(G)A\}]}_{=0}
\end{aligned}
$$

$$+ E\big[\{G - E(G|A=0)\}\{A - E(A|G=0)\}\underbrace{E\big[E(Y|A=0,G=0,U)|A\big]\exp\{-\varphi_g(G)A\}}_{=0}\big].$$

Likewise

$$E\big[\{G - E(G|A=0)\}\{A - E(A|G=0)\}$$
$$\times AY\exp\{-\varphi_g(G)A\}\big]$$
$$= E\big[\{G - E(G|A=0)\}\{A - E(A|G=0)\}$$
$$\times \exp(\beta_a)E(Y|A=0,G,U)A\exp\{-\varphi_g(G)A\}\big].$$

Therefore,

$$E\big[\{G - E(G|A=0)\}\{A - E(A|G=0)\}Y$$
$$\times \exp\{-\varphi_g(G)A\}\big]\big(E\big[\{G - E(G|A=0)\}$$
$$\times \{A - E(A|G=0)\}AY\exp\{-\varphi_g(G)A\}\big]\big)^{-1}$$
$$= (\exp(\beta_a) - 1)E\big[\{G - E(G|A=0)\}$$
$$\times \{A - E(A|G=0)\}AE(Y|A=0,G,U)$$
$$\times \exp\{-\varphi_g(G)A\}\big]$$
$$\times \big(\exp(\beta_a)E\big[\{G - E(G|A=0)\}$$
$$\times \{A - E(A|G=0)\}$$
$$\times E(Y|A=0,G,U)A\exp\{-\varphi_g(G)A\}\big]\big)^{-1}$$
$$= \frac{(\exp(\beta_a) - 1)}{\exp(\beta_a)},$$
$$\theta = -\ln\bigg(1 - \frac{(\exp(\beta_a) - 1)}{\exp(\beta_a)}\bigg)$$
$$= -\ln\exp(-\beta_a)$$
$$= \beta_a$$

provided that

$$E\big[\{G - E(G|A=0)\}\{A - E(A|G=0)\}$$
$$\times E(Y|A=0,G,U)A\exp\{-\varphi_g(G)A\}\big]$$
$$= E\big[\{G - E(G|A=0)\}\{A - E(A|G=0)\}$$
$$\times \beta_g(G)A\exp\{-\varphi_g(G)A\}\big]$$
$$\neq 0$$

which holds by assumption because $\gamma_{ag}(g) = (\exp(\beta_a) - 1)\beta_g(g)$. □

A.5 Proof of Lemma 4.4

PROOF. We note that by assumption

$$E(dN(y) - d\mathbf{B}_a(y)A - d\mathbf{B}_g(y)G|R(y)=1,A,G,U)$$
$$= d\mathbf{B}_0(y) + d\mathbf{B}_u(y,U),$$

and

$$E(\exp\{\mathbf{B}_a(y)A + \mathbf{B}_g(y)G\}R(y)|A,G,U)$$
$$= \exp\{-\mathbf{B}_0(y) - \mathbf{B}_u(y,U)\}.$$

Therefore

$$E\{W(y,\mathbf{B}_a(y),\mathbf{B}_g(y))\}$$
$$= E\bigg\{(d\mathbf{B}_0(y) + d\mathbf{B}_u(y,U))\exp\{-\mathbf{B}_0(y) - \mathbf{B}_u(y,U)\}\binom{(G - E(G))}{(G - E(G))(A - E(A|G))}\bigg\}$$
$$= E\bigg\{\binom{0}{(d\mathbf{B}_0(y) + d\mathbf{B}_u(y,U))\exp\{-\mathbf{B}_0(y) - \mathbf{B}_u(y,U)\}(U - E(U))(G - E(G))}\bigg\}$$
$$= 0.$$ □

APPENDIX B: VARIANCE ESTIMATION

B.1 Single IV

The estimating equation in (3.2) involves the estimated nuisance parameters $\hat{\mu} = \mathbb{P}_n(G)$ and $\hat{\psi}$ of the model $E(A|G;\psi)$. To account for the effect of nuisance parameter estimation on the subsequent estimation of β_a, the empirical moment conditions are stacked to form

$$m_\theta(\theta) = \mathbb{P}_n\begin{bmatrix} G - \mu \\ (1,G)'[A - E(A|G;\psi)] \\ (G - \mu)[A - E(A|G;\psi)](Y - \beta_a A) \end{bmatrix}$$

where $\theta = (\mu, \psi, \beta_a)$.

The estimation procedure satisfies the joint conditions $m_\theta(\hat{\theta}) = 0$. Without loss of generality, we specify $[A - E(A|G;\psi_0)]$ as a main effects model with intercept. Assume standard regularity conditions and expand $\hat{\theta}$ around the true parameter value θ_0 yields

$$\sqrt{n}(\hat{\theta} - \theta_0) = -\bigg[\frac{\partial m_\theta(\theta)}{\partial\theta}\bigg|_{\theta*}\bigg]^{-}\sqrt{n}m_\theta(\theta_0),$$

where θ^* is intermediate in value between $\hat{\theta}$ and θ_0. It follows that

$$\sqrt{n}m_\theta(\theta_0)$$
$$= \sqrt{n}\mathbb{P}_n\begin{bmatrix} G - \mu_0 \\ (1,G)'[A - E(A|G;\psi_0)] \\ (G - \mu_0)[A - E(A|G;\psi_0)](Y - \beta_{a0}A) \end{bmatrix}$$
$$= \sqrt{n}\mathbb{P}_n\{\tilde{m}(\theta_0)\} \xrightarrow{d} N(0, E[\tilde{m}(\theta_0)\tilde{m}(\theta_0)']),$$

while for the "bread" matrix

$$\frac{\partial m_\theta(\theta)}{\partial\theta}\bigg|_{\theta*}$$
$$= B^*(\theta^*)$$
$$= \mathbb{P}_n\begin{bmatrix} -1 & 0_{1\times 2} & 0 \\ 0_{2\times 1} & -\big\{(1,G)'\frac{\partial}{\partial\psi}E(A|G;\psi)|_{\psi*}\big\} & 0_{2\times 1} \\ & \frac{\partial\widehat{U}}{\partial\theta}|_{\theta*} & \end{bmatrix},$$

where

$$\frac{\partial}{\partial\psi}E(A|G;\psi)$$

$$
= \begin{cases} (1, G), \text{ for continuous } A \\ \dfrac{\exp(1, G)\psi}{1 + \exp(1, G)\psi} \left(1 - \dfrac{\exp(1, G)\psi}{1 + \exp(1, G')\psi}\right)(1, G), \\ \qquad \text{for binary A (logit model),} \end{cases}
$$

and

$$
\frac{\partial \widehat{U}}{\partial \mu} = -(A - E(A|G; \psi))(Y - \beta_a A)
$$

$$
\frac{\partial \widehat{U}}{\partial \psi} = -(G - \mu)(Y - \beta_a A)\frac{\partial}{\partial \psi} E(A|G; \psi)
$$

$$
\frac{\partial \widehat{U}}{\partial \beta_a} = -(G - \mu)(A - E(A|G; \psi))A.
$$

Assume that the matrix $B(\theta_0)$ is nonsingular, where the entries in $B(\theta_0)$ are the expected values of the sample averages in $B^*(\theta^*)$, evaluated at θ_0. Then $B^*(\theta^*) \xrightarrow{p} B(\theta_0)$, and

$$
\text{(S1)} \quad \begin{aligned} &\sqrt{n}(\hat{\theta} - \theta_0) \\ &\xrightarrow{d} N\big(0, B(\theta_0)^- E[\tilde{m}(\theta_0)\tilde{m}(\theta_0)']B(\theta_0)^{-'}\big). \end{aligned}
$$

Replacing the expected values in (S1) with sample averages evaluated at $\hat{\theta}$ yields a consistent estimator of the asymptotic covariance matrix. For inference about β_a, one may report its Wald-type 95% confidence interval constructed with the corresponding component of the estimated covariance matrix for $\hat{\theta}$.

B.2 Multiple IVs

Let $\widehat{\beta}_a$ be the solution to (4.2) with optimal weight $\widehat{W}_{opt} = \mathbb{P}_n[\widehat{U}(\beta_a)\widehat{U}(\beta_a)']^-$ where T^- denotes the generalized inverse of matrix T. The empirical moment conditions $\widehat{U}(\beta_a)$ in (4.2) involves the first stage estimates $\hat{\mu} = \mathbb{P}_n G$ as well as $\hat{\psi}$ of the model $E(A|G; \psi)$, which effects need to be accounted for in the subsequent estimation of β_a. Without loss of generality, we specify $[A - E(A|G; \psi_0)]$ as a main effects model with intercept. If there are k IVs, let

$$
m_\mu(\mu) = \mathbb{P}_n(G - \mu),
$$

$$
m_\psi(\psi) = \mathbb{P}_n(1, G')'[A - E(A|G; \psi)]
$$

be the k and $(k+1)$ empirical moment conditions of obtaining $\hat{\mu}$ and $\hat{\psi}$ respectively. For iterated or continuously updated GMM procedures in which β_a is estimated simultaneously with the optimal weight, the first-order condition of (4.2) is

$$
m_{\beta_a}(\beta_a) = \left\{\mathbb{P}_n\left[\frac{\partial \widehat{U}(\beta_a)}{\partial \beta_a}\right]\right\}' \widehat{W}_{opt}(\beta_a)\mathbb{P}_n[\widehat{U}(\beta_a)]
$$

$$
+ o_p(n^{-1/2}).
$$

Let $m_\theta(\theta) = (m'_\mu(\mu), m'_\psi(\psi), m_{\beta_a}(\beta_a))'$ where $\theta = (\mu', \psi', \beta_a)'$. The two-stage procedure solution satisfies

the joint moment condition $m_\theta(\hat{\theta}) = 0$. Assume standard regularity conditions and expand $\hat{\theta}$ around the true parameter value θ_0 yields

$$
\sqrt{n}(\hat{\theta} - \theta_0) = -\left[\frac{\partial m_\theta(\theta)}{\partial \theta}\Big|_{\theta^*}\right]^- \sqrt{n}m_\theta(\theta_0),
$$

where θ^* is intermediate in value between $\hat{\theta}$ and θ_0. Consider

$$
\sqrt{n}m_\theta(\theta_0)
$$

$$
= \begin{bmatrix} I_{(2k+1)\times(2k+1)} & 0_{(2k+1)\times k} \\ 0_{1\times(2k+1)} & \left\{\mathbb{P}_n\left[\frac{\partial \widehat{U}(\beta_a)}{\partial \beta_a}\Big|_{\beta_{a0}}\right]\right\}' \widehat{W}_{opt}(\beta_{a0}) \end{bmatrix}
$$

$$
\times \sqrt{n}\mathbb{P}_n\begin{bmatrix} G - \mu_0 \\ (1, G')'[A - E(A|G; \psi_0)] \\ U(\beta_{a0}) \end{bmatrix} + o_p(1).
$$

Let

$$
\Lambda = E\left(\frac{\partial U(\beta_a)}{\partial \beta_a}\Big|_{\beta_{a0}}\right), \quad \Omega = E[U(\beta_{a0})U(\beta_{a0})'],
$$

so that

$$
\left\{\mathbb{P}_n\left[\frac{\partial \widehat{U}(\beta_a)}{\partial \beta_a}\Big|_{\beta_{a0}}\right]\right\}' \xrightarrow{p} \Lambda', \quad \widehat{W}_{opt}(\beta_{a0}) \xrightarrow{p} \Omega^-.
$$

Then

$$
\sqrt{n}\mathbb{P}_n\begin{bmatrix} G - \mu_0 \\ (1, G')'[A - E(A|G; \psi_0)] \\ U(\beta_{a0}) \end{bmatrix} = \sqrt{n}\mathbb{P}_n\{\tilde{m}(\theta_0)\}
$$

$$
\xrightarrow{d} N(0, E[\tilde{m}(\theta_0)\tilde{m}(\theta_0)']),
$$

and by Slutsky's theorem,

$$
\sqrt{n}m_\theta(\theta_0) \xrightarrow{d} \begin{bmatrix} I_{(2k+1)\times(2k+1)} & 0_{(2k+1)\times k} \\ 0_{1\times(2k+1)} & \Lambda'\Omega^- \end{bmatrix}
$$

$$
\times N(0, E[\tilde{m}(\theta_0)\tilde{m}(\theta_0)'])
$$

$$
= M(\theta_0)N(0, E[\tilde{m}(\theta_0)\tilde{m}(\theta_0)']).
$$

Next consider the "bread" matrix

$$
\frac{\partial m_\theta(\theta)}{\partial \theta}\Big|_{\theta^*} = B^*(\theta^*) =
$$

$$
\begin{bmatrix} -I_{k\times k} & 0_{k\times(k+1)} & 0_{k\times 1} \\ 0_{(k+1)\times k} & -\mathbb{P}_n\left\{(1, G')'\frac{\partial}{\partial \psi} E(A|G; \psi)\Big|_{\psi^*}\right\} & 0_{(k+1)\times 1} \\ \left\{\mathbb{P}_n\left[\frac{\partial \widehat{U}(\beta_a)}{\partial \beta_a}\Big|_{\beta_a^*}\right]\right\}' \widehat{W}_{opt}(\beta_a^*)\mathbb{P}_n\left\{\frac{\partial \widehat{U}}{\partial \mu}\Big|_{\mu^*}, \frac{\partial \widehat{U}}{\partial \psi}\Big|_{\psi^*}, \frac{\partial \widehat{U}}{\partial \beta_a}\Big|_{\beta_a^*}\right\} + o_p(1) \end{bmatrix}
$$

where

$$
\frac{\partial}{\partial \psi} E(A|G; \psi)
$$

$$
= \begin{cases} (1, G'), \text{ for continuous } A \\ \dfrac{\exp(1, G')\psi}{1 + \exp(1, G')\psi} \left(1 - \dfrac{\exp(1, G')\psi}{1 + \exp(1, G')\psi}\right)(1, G'), \\ \qquad \text{for binary A (logit model),} \end{cases}
$$

and

$$\frac{\partial \widehat{U}}{\partial \mu} = -I_{k \times k}\big(A - E(A|G; \psi)\big)(Y - \beta_a A),$$

$$\frac{\partial \widehat{U}}{\partial \psi} = -(G - \mu)(Y - \beta_a A)\frac{\partial}{\partial \psi} E(A|G; \psi),$$

$$\frac{\partial \widehat{U}}{\partial \beta_a} = -(G - \mu)\big(A - E(A|G; \psi)\big)A.$$

Assume that the matrix $B(\theta_0)$ is nonsingular, where the entries in $B(\theta_0)$ are the expected values of the sample averages in $B^*(\theta^*)$, evaluated at θ_0. Then $B^*(\theta^*) \xrightarrow{p} B(\theta_0)$, and

$$\sqrt{n}(\hat{\theta} - \theta_0)$$

(S4)
$$\xrightarrow{d} N\big(0, B(\theta_0)^- M(\theta_0) E\big[\tilde{m}(\theta_0)\tilde{m}(\theta_0)'\big]$$
$$\times M(\theta_0)' B(\theta_0)^{-'}\big).$$

In practice, replacing the expected values in (S4) with sample averages evaluated at $\hat{\theta}$ yields a consistent estimator of the asymptotic covariance matrix. In addition, centering the IV moment conditions $\widehat{U}(\beta_a)$ when estimating the covariance matrix $E\big[\tilde{m}(\theta_0)\tilde{m}(\theta_0)'\big]$ may improve finite sample inference. For inference about β_a, one may report its Wald-type 95% confidence interval constructed with the corresponding component of the estimated covariance matrix for $\hat{\theta}$. The above variance estimation framework can accommodate baseline covariates C by stacking the moment conditions for $\hat{E}(G|C)$ and $\hat{E}(A|G, C)$ instead, as described in estimating equation (4.1).

ACKNOWLEDGMENTS

FUNDING

SUPPLEMENTARY MATERIAL

Supplement to "The GENIUS Approach to Robust Mendelian Randomization." (DOI: 10.1214/20-STS802SUPP; .pdf). Supplementary information.

REFERENCES

ALBERT, J. M. (2008). Mediation analysis *via* potential outcomes models. *Stat. Med.* **27** 1282–1304. MR2420158 https://doi.org/10.1002/sim.3016

ANGRIST, J. D., IMBENS, G. W. and RUBIN, D. B. (1996). Identification of causal effects using instrumental variables. *J. Amer. Statist. Assoc.* **91** 444–455.

BOWDEN, J., DAVEY SMITH, G. and BURGESS, S. (2015). Mendelian randomization with invalid instruments: Effect estimation and bias detection through egger regression. *Int. J. Epidemiol.* **44** 512–525.

BOWDEN, J. and VANSTEELANDT, S. (2011). Mendelian randomization analysis of case-control data using structural mean models. *Stat. Med.* **30** 678–694. MR2767465 https://doi.org/10.1002/sim.4138

BOWDEN, J., DAVEY SMITH, G., HAYCOCK, P. C. and BURGESS, S. (2016). Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. *Genet. Epidemiol.* **40** 304–314.

BREUSCH, T. S. and PAGAN, A. R. (1979). A simple test for heteroscedasticity and random coefficient variation. *Econometrica* **47** 1287–1294. MR0545960 https://doi.org/10.2307/1911963

CHAO, J. C. and SWANSON, N. R. (2005). Consistent estimation with a large number of weak instruments. *Econometrica* **73** 1673–1692. MR2156676 https://doi.org/10.1111/j.1468-0262.2005.00632.x

CONLEY, T. G., HANSEN, C. B. and ROSSI, P. E. (2012). Plausibly exogenous. *Rev. Econ. Stat.* **94** 260–272.

DAVEY SMITH, G. and EBRAHIM, S. (2003). 'Mendelian randomization': Can genetic epidemiology contribute to understanding environmental determinants of disease? *Int. J. Epidemiol.* **32** 1–22.

DAVEY SMITH, G. and EBRAHIM, S. (2004). Mendelian randomization: Prospects, potentials, and limitations. *Int. J. Epidemiol.* **33** 30–42.

DIDELEZ, V., MENG, S. and SHEEHAN, N. A. (2010). Assumptions of IV methods for observational epidemiology. *Statist. Sci.* **25** 22–40. MR2741813 https://doi.org/10.1214/09-STS316

EMSLEY, R., DUNN, G. and WHITE, I. R. (2010). Mediation and moderation of treatment effects in randomised controlled trials of complex interventions. *Stat. Methods Med. Res.* **19** 237–270. MR2757115 https://doi.org/10.1177/0962280209105014

FLORES, C. A. and FLORES-LAGUNES, A. (2013). Partial identification of local average treatment effects with an invalid instrument. *J. Bus. Econom. Statist.* **31** 534–545. MR3173699 https://doi.org/10.1080/07350015.2013.822760

GUO, Z., KANG, H., CAI, T. T. and SMALL, D. S. (2018). Confidence intervals for causal effects with invalid instruments by using two-stage hard thresholding with voting. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 793–815. MR3849344 https://doi.org/10.1111/rssb.12275

HAN, C. (2008). Detecting invalid instruments using L_1-GMM. *Econom. Lett.* **101** 285–287. MR2477476 https://doi.org/10.1016/j.econlet.2008.09.004

HECKMAN, J. and VYTLACIL, E. (1998). Instrumental variables methods for the correlated random coefficient model: Estimating the average rate of return to schooling when the return is correlated with schooling. *J. Hum. Resour.* 974–987.

HIRANO, K., IMBENS, G. W., RUBIN, D. B. and ZHOU, X.-H. (2000). Assessing the effect of an influenza vaccine in an encouragement design. *Biostatistics* **1** 69–88.

HUBER, M. (2014). Sensitivity checks for the local average treatment effect. *Econom. Lett.* **123** 220–223.

JUSTER, F. T. and SUZMAN, R. (1995). An overview of the health and retirement study. *J. Hum. Resour.* **30** S7–S56.

KANG, H. (2017). sisVIVE: Some Invalid Some Valid Instrumental Variables Estimator. R package version 1.4.

KANG, H., ZHANG, A., CAI, T. T. and SMALL, D. S. (2016). Instrumental variables estimation with some invalid instruments and its application to Mendelian randomization. *J. Amer. Statist. Assoc.* **111** 132–144. MR3494648 https://doi.org/10.1080/01621459.2014.994705

KLEIBER, C. and ZEILEIS, A. (2008). *Applied Econometrics with R*. Springer, New York.

KOLESÁR, M., CHETTY, R., FRIEDMAN, J., GLAESER, E. and IMBENS, G. W. (2015). Identification and inference with many invalid instruments. *J. Bus. Econom. Statist.* **33** 474–484. MR3416595 https://doi.org/10.1080/07350015.2014.978175

LAWLOR, D. A., HARBORD, R. M., STERNE, J. A. C., TIMPSON, N. and SMITH, G. D. (2008). Mendelian randomization: Using genes as instruments for making causal inferences in epidemiology. *Stat. Med.* **27** 1133–1163. MR2420151 https://doi.org/10.1002/sim.3034

LEEB, H. and PÖTSCHER, B. M. (2008). Sparse estimators and the oracle property, or the return of Hodges' estimator. *J. Econometrics* **142** 201–211. MR2394290 https://doi.org/10.1016/j.jeconom.2007.05.017

LEWBEL, A. (2012). Using heteroscedasticity to identify and estimate mismeasured and endogenous regressor models. *J. Bus. Econom. Statist.* **30** 67–80. MR2899185 https://doi.org/10.1080/07350015.2012.643126

LEWBEL, A. (2018). Identification and estimation using heteroscedasticity without instruments: The binary endogenous regressor case. *Econom. Lett.* **165** 10–12. MR3768154 https://doi.org/10.1016/j.econlet.2018.01.003

LI, J., FINE, J. and BROOKHART, A. (2015). Instrumental variable additive hazards models. *Biometrics* **71** 122–130. MR3335356 https://doi.org/10.1111/biom.12244

LITTLE, J. and KHOURY, M. J. (2003). Mendelian randomisation: A new spin or real progress? *Lancet* **362** 930–931. https://doi.org/10.1016/S0140-6736(03)14396-6

MA, Z., XIE, X. and GENG, Z. (2006). Collapsibility of distribution dependence. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **68** 127–133. MR2212578 https://doi.org/10.1111/j.1467-9868.2005.00536.x

MAILMAN, M. D., FEOLO, M., JIN, Y., KIMURA, M., TRYKA, K., BAGOUTDINOV, R., HAO, L., KIANG, A., PASCHALL, J. et al. (2007). The NCBI dbGaP database of genotypes and phenotypes. *Nat. Genet.* **39** 1181.

MANOLIO, T. A., COLLINS, F. S., COX, N. J., GOLDSTEIN, D. B., HINDORFF, L. A., HUNTER, D. J., MCCARTHY, M. I., RAMOS, E. M., CARDON, L. R. et al. (2009). Finding the missing heritability of complex diseases. *Nature* **461** 747.

MANSKI, C. F. and PEPPER, J. V. (2000). Monotone instrumental variables: With an application to the returns to schooling. *Econometrica* **68** 997–1010. MR1771587 https://doi.org/10.1111/1468-0262.00144

MARTINUSSEN, T., VANSTEELANDT, S., TCHETGEN TCHETGEN, E. J. and ZUCKER, D. M. (2017). Instrumental variables estimation of exposure effects on a time-to-event endpoint using structural cumulative survival models. *Biometrics* **73** 1140–1149. MR3744528 https://doi.org/10.1111/biom.12699

MEALLI, F. and PACINI, B. (2013). Using secondary outcomes to sharpen inference in randomized experiments with noncompliance. *J. Amer. Statist. Assoc.* **108** 1120–1131. MR3174688 https://doi.org/10.1080/01621459.2013.802238

MORRIS, A. P., VOIGHT, B. F., TESLOVICH, T. M., FERREIRA, T., SEGRE, A. V., STEINTHORSDOTTIR, V., STRAWBRIDGE, R. J., KHAN, H., GRALLERT, H. et al. (2012). Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat. Genet.* **44** 981.

NEVO, A. and ROSEN, A. M. (2012). Identification with imperfect instruments. *Rev. Econ. Stat.* **94** 659–671.

NEWEY, W. K. and WINDMEIJER, F. (2009). Generalized method of moments with many weak moment conditions. *Econometrica* **77** 687–719. MR2531359 https://doi.org/10.3982/ECTA6224

NIE, H., CHENG, J. and SMALL, D. S. (2011). Inference for the effect of treatment on survival probability in randomized trials with noncompliance and administrative censoring. *Biometrics* **67** 1397–1405. MR2872390 https://doi.org/10.1111/j.1541-0420.2011.01575.x

OFSTEDAL, M. B., FISHER, G. G. and HERZOG, A. R. (2005). Documentation of cognitive functioning measures in the health and retirement study.

ROBINS, J. M. (1989). The analysis of randomized and non-randomized AIDS treatment trials using a new approach to causal inference in longitudinal studies. In: *Health Service Research Methodology*: *A Focus on AIDS* 113–159.

ROBINS, J. M. (1994). Correcting for non-compliance in randomized trials using structural nested mean models. *Comm. Statist. Theory Methods* **23** 2379–2412. MR1293185 https://doi.org/10.1080/03610929408831393

ROBINS, J. M. (1997). Causal inference from complex longitudinal data. In *Latent Variable Modeling and Applications to Causality* (*Los Angeles*, *CA*, 1994). *Lect. Notes Stat.* **120** 69–117. Springer, New York. MR1601279 https://doi.org/10.1007/978-1-4612-1842-5_4

SHARDELL, M. and FERRUCCI, L. (2016). Instrumental variable analysis of multiplicative models with potentially invalid instruments. *Stat. Med.* **35** 5430–5447. MR3573068 https://doi.org/10.1002/sim.7069

SMALL, D. S. (2012). Mediation analysis without sequential ignorability: Using baseline covariates interacted with random assignment as instrumental variables. *J. Statist. Res.* **46** 91–103. MR3114108

SPILLER, W., SLICHTER, D., BOWDEN, J. and DAVEY SMITH, G. (2018). Detecting and correcting for bias in Mendelian randomization analyses using gene-by-environment interactions. https://doi.org/10.1093/ije/dyy204

STAIGER, D. and STOCK, J. H. (1997). Instrumental variables regression with weak instruments. *Econometrica* **65** 557–586. MR1445622 https://doi.org/10.2307/2171753

STOCK, J. H. and WRIGHT, J. H. (2000). GMM with weak identification. *Econometrica* **68** 1055–1096. MR1779144 https://doi.org/10.1111/1468-0262.00151

STOCK, J. H., WRIGHT, J. H. and YOGO, M. (2002). A survey of weak instruments and weak identification in generalized method of moments. *J. Bus. Econom. Statist.* **20** 518–529. MR1973801 https://doi.org/10.1198/073500102288618658

TCHETGEN TCHETGEN, E. J., ROBINS, J. M. and ROTNITZKY, A. (2010). On doubly robust estimation in a semiparametric odds ratio model. *Biometrika* **97** 171–180. MR2594425 https://doi.org/10.1093/biomet/asp062

TCHETGEN TCHETGEN, E., SUN, B. and WALTER, S. (2021). Supplement to "The GENIUS Approach to Robust Mendelian Randomization Inference." https://doi.org/10.1214/20-STS802SUPP

TCHETGEN TCHETGEN, E. J., WALTER, S., VANSTEELANDT, S., MARTINUSSEN, T. and GLYMOUR, M. (2015). Instrumental variable estimation in a survival context. *Epidemiology* **26** 402.

TEN HAVE, T. R., JOFFE, M. M., LYNCH, K. G., BROWN, G. K., MAISTO, S. A. and BECK, A. T. (2007). Causal mediation analyses with rank preserving models. *Biometrics* **63** 926–934. MR2395813 https://doi.org/10.1111/j.1541-0420.2007.00766.x

VANSTEELANDT, S. and DIDELEZ, V. (2018). Improving the robustness and efficiency of covariate-adjusted linear instrumental variable estimators. *Scand. J. Stat.* **45** 941–961. MR3884895 https://doi.org/10.1111/sjos.12329

WANG, L. and TCHETGEN TCHETGEN, E. (2018). Bounded, efficient and multiply robust estimation of average treatment effects using instrumental variables. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 531–550. MR3798877 https://doi.org/10.1111/rssb.12262

WHITE, K., MONDESIR, F. L., BATES, L. M. and GLYMOUR, M. M. (2014). Diabetes risk, diagnosis, and control: Do psychosocial factors predict hemoglobin A1C defined outcomes or accuracy of self-reports? *Ethn. Dis.* **21** 19–27.

WINDMEIJER, F., FARBMACHER, H., DAVIES, N. and SMITH, G. D. (2019). On the use of the Lasso for instrumental variables estimation with some invalid instruments. *J. Amer. Statist. Assoc.* **114** 1339–1350. MR4011783 https://doi.org/10.1080/01621459.2018.1498346

WOOLDRIDGE, J. M. (2003). Further results on instrumental variables estimation of average treatment effects in the correlated random coefficient model. *Econom. Lett.* **79** 185–191. MR1968966 https://doi.org/10.1016/S0165-1765(02)00318-X

WOOLDRIDGE, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*, 2nd ed. MIT Press, Cambridge, MA. MR2768559

WU, Q., TCHETGEN TCHETGEN, E., OSYPUK, T., WHITE, K., MUJAHID, M. and GLYMOUR, M. (2012). Combining direct and proxy assessments to reduce attrition bias in a longitudinal study. *Alzheimer Dis. Assoc. Disord.* **27** 207–212.

YAVORSKA, O. and STALEY, J. (2019). MendelianRandomization: Mendelian Randomization Package. R package version 0.4.1.

ZHENG, C. and ZHOU, X.-H. (2015). Causal mediation analysis in the multilevel intervention and multicomponent mediator case. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **77** 581–615. MR3351447 https://doi.org/10.1111/rssb.12082