Original Article

The GENIUS Approach to Robust Mendelian Randomization Inference

Eric J. Tchetgen Tchetgen*, BaoLuo Sun†, Stefan Walter**

*Department of Biostatistics,Harvard University

**Department of Epidemiology and Biostatistics, University of California

†Computational and Systems Biology, Genome Institute of Singapore

Abstract

Mendelian randomization (MR) is a popular instrumental variable (IV) approach, in which one or several genetic markers serve as IVs that can be leveraged to recover under certain conditions, 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 has no direct effect on the outcome that is not mediated by the exposure in view. In MR studies, such an assumption requires an unrealistic level of knowledge and understanding of the mechanism by which the genetic markers causally affect the outcome, particularly when a large number of genetic variants are considered as IVs. As a result, possible violation of the exclusion restriction can seldom be ruled out in such MR studies, and if present, such violation can invalidate IV-based inferences even if unbeknownst to the analyst, confounding is either negligible or absent. To address this concern, we introduce a new class of IV estimators which are robust to violation of the exclusion restriction under a large collection of data generating mechanisms consistent with parametric models commonly assumed in the MR literature. Our approach which we have named ”MR G-Estimation under No Interaction with Unmeasured Selection” (MR GENIUS) may in fact be viewed as a modification to Robins’ G-estimation approach that is robust to both additive unmeasured confounding and violation of the exclusion restriction assumption. We also establish that estimation with MR GENIUS may also be viewed as a robust generalization of the well-known Lewbel estimator for a triangular system of structural equations with endogeneity. Specifically, we show that unlike Lewbel estimation, MR GENIUS is under fairly weak conditions also robust to unmeasured confounding of the effects of the genetic IVs on both the exposure and the outcome, another possible violation of a key IV Identification condition. Furthermore, while Lewbel estimation involves specification of linear models both for the outcome and the exposure, MR GENIUS generally does not require specification of a structural model for the direct effect of invalid IVs on the outcome, therefore allowing the latter model to be unrestricted. Finally, unlike Lewbel estimation, MR GENIUS is shown to equally apply for binary, discrete or continuous exposure and outcome variables and can be used under prospective sampling, or retrospective sampling such as in a case-control study, as well as for right censored time-to-event outcomes under an additive hazards model.

KEY WORDS: Instrumental variable, exclusion restriction, additive model, g-estimation, confounding, 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, 2008). In order to be valid IVs, the genetic markers must satisfy the following key conditions:

(a) They must be associated with the exposure.

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

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

The last assumption (c) 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 pleitropic 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 violation 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 formal statistical methods to detect and account for violation of IV assumptions (a)-(c), primarily in a multiple-IV setting in which standard linear models for outcome and exposure are assumed. 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; some recent papers on this topic include Staiger and Stock (1997), Stock and Wright (2000), Stock and Yogo (2002), Chao and Swanson (2005). Methodology to address violations of (b) or (c) is far less developed, and constitutes the central focus of this paper. Three strands of work stand out in recent literature concerning violation of either of these assumptions. In the first strand, Kang et al (2016) developed a penalized regression approach that can under certain conditions recover valid inferences about the causal effect of interest provided fewer than fifty percent of genetic markers are invalid IVs; also see Windmeijer et al (2016) for improvements on the penalized approach of Kang et al (2016), including a proposal for standard error estimation which was not provided 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 fewer than fifty percent of IVs are invalid 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 (2017) proposed two stage hard thresholding (TSHT) with voting, which is able to recover a consistent causal effect estimator under linear models for the outcome and exposure, and a certain plurality condition which can be considerably weaker than the fifty percent rule (also known as majority rule). The plurality condition 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 above regression coefficients. 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 condition. 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, i.e. either (b) or (c) fails. Both the penalized approach and the median estimator may be inconsistent if 50% or more candidate IVs turnout to be invalid, while TSHT may be inconsistent if the 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 order to remedy this difficulty, in a third strand of work, Kolesár et al (2011) 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 et al, 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 pleitropic effects of the genetic markers often with little to no biological basis. A notable feature of aforementioned methods is that they are primarely 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 (2015) and Guo et al (2017) rely on a consistent model selection procedure, such confidence intervals fail to be uniformly valid over the entire model space (Guo et al, 2017, Leeb and Pötscher, 2008).

Because in practice, it is not possible to ensure that fewer than fifty percent of candidate IVs are invalid or that the plurality condition holds, nor is it practically possible to enforce the orthogonality condition of Kolesár et al (2011), an important goal of MR research aims to develop alternative methods of estimation and inference that are fully robust to possible violation of IV assumptions without relying on majority, plurality or orthogonality conditions. In this paper, a class of estimators fulfilling this desideratum is proposed, which unlike the aforementioned robust methods equally applies whether one has observed a single or many candidate IVs.

In Section 2, we introduce notation used throughout. We also provide a formal definition of the IV model for which we describe previously proposed sufficient conditions in the canonical case of binary exposure and IV, for nonparametric identification of the exposure average causal effect in terms of the so-called Wald estimand (Wang and Tchetgen Tchetgen, 2017). In Section 3, we present our first result which provides an alternative identification formula for the average causal effect in the IV context, which unlike the Wald estimand, is robust to violation of the exclusion restriction (c) under a large collection of possible data generating mechanisms that assume both (i) no additive interaction between the exposure, an unmeasured confounder and the candidate IV in a mean model for the outcome; and (ii) no additive interaction between the candidate IV and an unmeasured confounder in a mean model for the exposure. Conditions similar to assumptions (i) and (ii) are fairly common in MR and other IV literature. For instance, Kolesár et al (2011), Kang et al (2016) and Bowden et al (2015) rely on analogous assumptions. In Section 3, we establish that the proposed approach readily accounts for continuous exposure. We establish that our approach which we call "MR G-Estimation under No Interaction with Unmeasured Selection" (MR GENIUS) may in fact be viewed as a modification to Robins' G-estimation approach (Robins, 1997) which we have made robust to both additive unmeasured confounding and violation of the exclusion restriction assumption. Identification with MR GENIUS relies primarily on an assumption that the conditional variance of the exposure given the candidate IVs is heteroscedastic with respect to the candidate IVs, an assumption which generally holds for binary or discrete exposure except at certain exceptional data generating mechanisms. In case of continuous exposure and outcome, this assumption is closely related to Lewbel's recent proposal to leverage heteroscedasticity for identification and estimation in endogenous regression models (Lewbel, 2012). In this case, MR GENIUS and Lewbel estimation are quite similar, although unlike estimation with the Lewbel approach, estimation with MR GENIUS avoids specification of a model for the direct effect of invalid IVs with the outcome, therefore allowing the latter to remain unrestricted. In Section 4, we describe conditions under which MR GENIUS is also robust to unmeasured confounding of the effects of the genetic IVs on both the exposure and the outcome, a violation of assumption (b) which is also not appropriately accounted for by Lewbel regression which assumes that candidate IVs are independent of unmeasured confounders. As we further establish in Section 5, MR GENIUS can easily incorporate multiple IVs in a generalized methods of moments (GMM) approach. 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 et al (2015) who require uncorrelated IVs and Kang et al (2016) who likewise require IV correlation structure to be somewhat restricted (Windmeijer et al, 2015). Section 5 also extends the proposed approach to target a multiplicative average causal effect, and establishes that in the case of binary outcome, the approach is equally valid under either prospective or retrospective sampling designs. Therefore, MR GENIUS can also be viewed as further generalizing Lewbel's estimator to these important settings. In Section 5, we also briefly extend MR GENIUS to the context of a right censored time-to-event endpoint under a structural additive hazards model, therefore further robustifying the recent semiparametric IV estimator of Martinussen et al (2017) against possible violation of the exclusion restriction assumption. In Section 6, we evaluate the proposed methods and compare them to a number of previous MR methods in extensive simulation studies. In Section 7 we illustrate the methods in an MR analysis of the effect of diabetes on memory in the Health and Retirement Study. Section 8 offers some concluding remarks.

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:

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 1 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, i.e. the causal diagram in Figure 1, 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. Consider the following condition:

Assumption 4.

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

$$E(Y|A=a,G,U)-E(Y|A=0,G,U)=\beta_a a \tag{1}$$

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

$$E(Y|A=0,G=g,U)-E(Y|A=0,G=0,U)=\beta_g(g) \tag{2}$$

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

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

$$E(A|G=g,U)-E(A|G=0,U)=\alpha_g(g) \tag{3}$$

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

Clearly the condition does not require G to be a valid IV. Equation (1) implies that the average causal effect of A on Y conditional on U and G does not depend on U and G on the additive scale, i.e. the additive causal effect of A on Y is not modified by either U or G. Likewise, equation (2) additionally states that the additive average effect of G on Y is not modified by U. These restrictions imply the following additive models:

$$E(Y|A,G,U)=\beta_a A+\beta_g(G)+U_y,$$
$$E(A|G,U)=\alpha_g(G)+U_a,$$

where $U_y=\beta_u(U)$ and $U_a=\alpha_u(U)$ for functions $\beta_u(\cdot)$ and $\alpha_u(\cdot)$ only restricted by natural features of the model, e.g. such that the outcome and exposure means are bounded between zero and one in the binary case. If G is a valid IV, then $E(Y|A,G,U)=E(Y|A,U)$ does not depend on G by the exclusion restriction implying that G neither interacts with U nor with A in the model for $E(Y|A,G,U)$, so that assumption 4.a. reduces to the assumption of no $U-A$

interaction in the model for $E\left(Y|A,G,U\right)$. In case of a valid binary IV and binary exposure, Wang and Tchetgen Tchetgen (2017) recently established that the average causal effect β_a is nonparametrically identified by the so-called Wald estimand

$$\beta_a = \delta \equiv \frac{E(Y|G=1) - E(Y|G=0)}{E(A|G=1) - E(A|G=0)}, \tag{4}$$

if either of 4.a. or 4.b holds but not necessarily both conditions hold. Note that the models for $E(Y|A,G,U)$ and $E(A|G,U)$ considered by Bowden et al (2015) satisfy assumptions 4.a. and 4.b. with $\beta_g\left(\cdot\right)$ and $\alpha_g\left(\cdot\right)$ linear functions, while Kang et al (2016) specified models implied by these two restrictions. Below, unless stated otherwise, assume A and G are both binary.

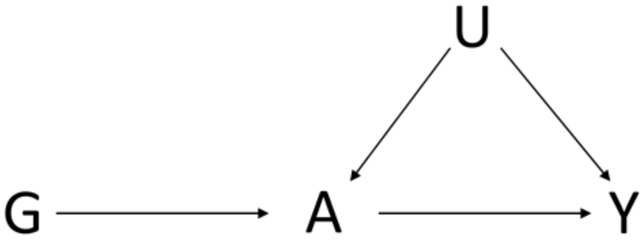

Figure 1: Directed acyclic graph depicting a valid instrument G which satisfies assumptions 1-3.

3 Identification under violation of exclusion restriction

Next, suppose that as encoded in the diagram given in Figure 2, the exclusion restriction assumption 3 does not necessarily hold, then the Wald estimand $\delta \neq \beta_a$ will generally fail to equal the average causal effect of A on Y, even if assumptions 1, 2, and 4 hold. The following result provides an alternative identifying formula which may be used instead of the Wald estimand to identify the causal effect under these conditions.

Lemma 1 *Suppose that Assumptions 1, 2 and 4 hold, then* $\beta_a = \mu$, *where*

$$\begin{aligned}\mu &= \frac{E\left[\{G-E(G)\}\{A-E(A|G)\}Y\right]}{E\left[\{G-E(G)\}\{A-E(A|G)\}A\right]} \\ &= \frac{E\left[\{G-E(G)\}\{A-E(A|G)\}Y\right]}{var(G)\left\{var(A|G=1)-var(A|G=0)\right\}}\end{aligned}$$

provided that

$$var\left(A|G=1\right)-var\left(A|G=0\right)\neq 0 \tag{5}$$

Proof. Below we make use of the fact that under our assumptions $E\left\{A-E(A|G)|G,U\right\} = \alpha_u(U)-E(\alpha_u(U))$. Consider

$$\begin{aligned}&E\left[\{G-E(G)\}\{A-E(A|G)\}Y\right] \\ &= E\left[\{G-E(G)\}\{A-E(A|G)\}E\left(Y|A,G,U\right)\right] \\ &= E\left[\{G-E(G)\}\{A-E(A|G)\}\{\beta_a A+\beta_g G+\beta_u\left(U\right)\}\right] \\ &= E\left[\{G-E(G)\}\{A-E(A|G)\}A\right]\beta_a \\ &+ \beta_g E\left[\{G-E(G)\}G\underbrace{E\left[\{A-E(A|G)\}|G\right]}_{=0}\right] \\ &+ \underbrace{E\left\{G-E(G)\right\}}_{=0}cov\left\{\alpha_u(U),\beta_u\left(U\right)\right\}\end{aligned}$$

Therefore,

$$\begin{aligned}&\frac{E\left[\{G-E(G)\}\{A-E(A|G)\}Y\right]}{E\left[\{G-E(G)\}\{A-E(A|G)\}A\right]} \\ &= \beta_a\end{aligned}$$

provided that $E\left[\{G-E(G)\}\{A-E(A|G)\}A\right]\neq 0$, which holds under (5). ■

Lemma 1 provides an explicit identifying expression for the average causal effect β_a 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,i.e. that $var(A|G)$ must depend on G. It is key to note that this assumption is empirically testable, and will typically hold for binary A, except at certain exceptional laws. To illustrate, let $\pi(g) = \Pr(A = 1|G = g)$ and suppose that assumptions 1, 2 and 4 hold, however $\pi\left(1\right) = 1 - \pi\left(0\right)$, in which case (5) fails because $var(A|G = g) = \pi(g)\left(1 - \pi(g)\right) = \pi(1)\left(1 - \pi(1)\right) = \pi(0)\left(1 - \pi(0)\right)$ 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, i.e. $\pi\left(1\right) \neq \pi\left(0\right)$. Below, we extend Lemma 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 2, 4 and equation (5) irrespective of whether or not assumption 3 also holds:

$$\widehat{\beta}_a = \frac{\mathbb{P}_n\left[\left\{G - \mathbb{P}_n(G)\right\}\left\{A - \widehat{E}(A|G)\right\}Y\right]}{\mathbb{P}_n\left[\left\{G - \mathbb{P}_n(G)\right\}\left\{A - \widehat{E}(A|G)\right\}A\right]}, \tag{6}$$

where $\mathbb{P}_n = n^{-1}\sum_{i=1}^{n}[\cdot]_i$ and $\widehat{E}(A|G = g) = \mathbb{P}_n\left[A_i 1\left(G_i = g\right)\right]/\mathbb{P}_n\left[1\left(G_i = g\right)\right]$. This estimator is the simplest instance of MR GENIUS estimation. The asymptotic distribution of the estimator is described in Appendix A2.

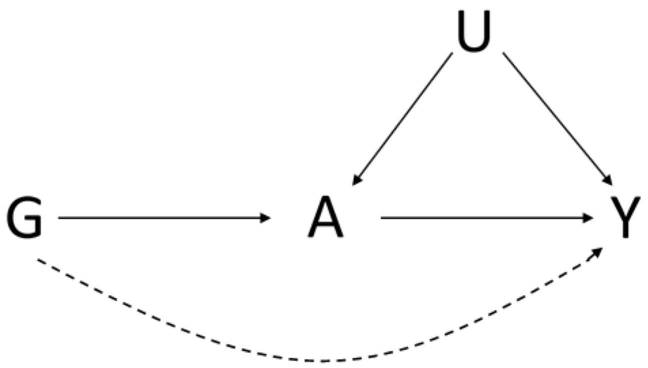

Figure 2: Directed acyclic graph depicting the situation in which exclusion restriction (assumption 3) does not necessarily hold. The dashed line indicates possible direct effect of G on outcome Y.

Continuous exposure

Suppose now that A is continuous, then, it is straightforward to verify that Lemma 1 continues to hold as its proof does not depend on A being binary. Note that for continuous A, Assumption (1) resticts the effect of A on Y to be linear and condition (5) implies that the conditional density of

$\varepsilon_A = A - E(A|G)$ must be heteroscedastic, i.e. $var(A|G) = E(\varepsilon_A^2|G)$ depends on G. As mentioned in the introduction, Lewbel (2012) obtains a closely related identification result to Lemma 1 under a triangular system of linear structural equation models; see Theorem 1 on Page 70 of Lewbel (2012). In addition to establishing the result for binary A in Lemma 1 without specification of a triangular system of linear equations, below we generalize this identification result in several important directions particularly relevant to MR studies.

We note that while $var(A|G)$ will generally depend on G for binary or discrete A (except perhaps at exceptional data generating mechanisms such as the one described in the previous Section), 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

$$A = \alpha_g^*\left(G, \varepsilon_g\right) + U_a + \varepsilon_a^*$$

$$E\left(\varepsilon_a^*\right) = 0$$

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, e.g. $\alpha_g^*(G, \epsilon_g) = \alpha_g^* G + \epsilon_g G$. Then, one can verify that the model in the above display implies that $A = \alpha_g\left(G\right) + \varepsilon_a$ where $\alpha_g\left(G\right) = E\left(\alpha_g^*\left(G, \varepsilon_g\right)|G\right) + E\left(U_a\right)$ and $var\left(\varepsilon_a|G\right) = var\left(\left\{U_a - E\left(U_a\right) + \varepsilon_a^* + \alpha_g^*\left(G, \varepsilon_g\right) - E\left(\alpha_g^*\left(G, \varepsilon_g\right)|G\right)\right\}\right.$ 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 (5). 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).

4 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 3. Therefore, we will consider replacing Assumption 2 with the following weaker condition:

Assumption 2*. Homoscedastic confounding: $cov\,(U_y, U_a|G) = \rho$ does not depend on G.

To illustrate Assumption 2* it is instructive to consider the following submodels of (1) and (2): $U_y = \beta_0 + \beta_u U$ and $U_a = \alpha_0 + \alpha_a U$, such that $E(Y|A, U, G)$ and $E(A|G, U)$ are both linear in U; then Assumption 2* implies $var(U|G) = \rho/\left(\beta_u \alpha_a\right)$, i.e. 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 2 *Suppose that Assumptions 1, 2*,4 hold, then $\beta_a = \mu$ provided that condition* (5) *holds.*

Proof. Proceeding as in the proof of Lemma 1,

$$
\begin{aligned}
& E\left[\{G - E(G)\}\{A - E(A|G)\}\,Y\right] \\
& = E\left[\{G - E(G)\}\{A - E(A|G)\}\,A\right]\beta_a \\
& + E\left[\{G - E(G)\}\{\alpha_u(U) - E\left(\alpha_u(U)|G\right)\}\,\beta_u\left(U\right)\right] \\
& = E\left[\{G - E(G)\}\{A - E(A|G)\}\,A\right]\beta_a \\
& + E\left[\{G - E(G)\}\,cov\left(\alpha_u(U), \beta_u\left(U\right)|G\right)\right] \\
& = E\left[\{G - E(G)\}\{A - E(A|G)\}\,A\right]\beta_a \\
& + E\left\{G - E(G)\right\}\rho \\
& = E\left[\{G - E(G)\}\{A - E(A|G)\}\,A\right]\beta_a,
\end{aligned}
$$

proving the result. ■

Lemma 2 implies that under Assumptions 1, 2*, 4 and condition (5) , $\widehat{\beta}_a$ 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 4.a.1 and the additional assumption of no unobserved confounding given G, i.e. 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:

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

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 (6) solves the estimating equation:

$$0 = \mathbb{P}_n \left[\{G - \mathbb{P}_n(G)\} \left\{ A - \widehat{E}(A|G) \right\} \left\{ Y - \widehat{\beta}_a A \right\} \right], \tag{7}$$

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.

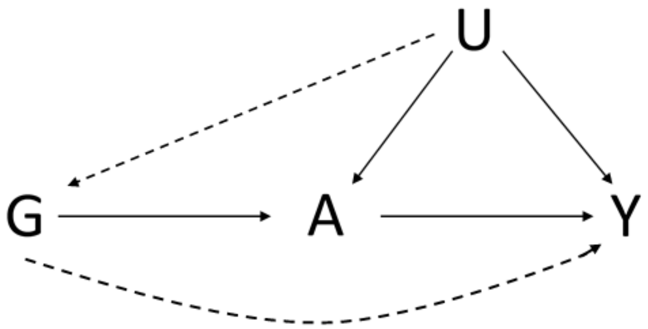

Figure 3: Directed acyclic graph depicting the situation in which IV independence (assumption 2) and exclusion restriction (assumption 3) do not necessarily hold. The dashed lines indicate possible direct effects of U on G, and of G on Y.

5 Generalizations

5.1 Multiplicative exposure model

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

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

$$\log \frac{E(A|G=g,U)}{E(A|G=0,U)} = \alpha_g(g) \tag{8}$$

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

$$\log \frac{E(A|G=g)}{E(A|G=0)} = \varpi_g(g), \tag{9}$$

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

Lemma 3 *Suppose that Assumptions 1, 2, 4.a and 4.b* hold, then*

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

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

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

$$\begin{aligned}&\exp(\varpi_g(G))\,E(A\exp(-\varpi_g(G)))\\&= E(A|G)\\&= \exp(\alpha_g(G))\,E(U_a),\end{aligned}$$

and

$$\begin{aligned}
&E\left(A|G,U\right)-\exp\left(\varpi_g\left(G\right)\right)E(A\exp\left(-\varpi_g\left(G\right)\right))\\
&=\left[U_a-E\left(U_a\right)\right]\exp\left(\alpha_g\left(G\right)\right).
\end{aligned}$$

Therefore

$$\begin{aligned}
&E\left[\left\{G-E(G)\right\}\left\{A\exp(-\varpi_g\left(G\right))-E(A\exp\left(-\varpi_g\left(G\right)\right))\right\}Y\right]\\
&=E\left[\left\{G-E(G)\right\}\left\{A\exp(-\varpi_g\left(G\right))-E(A\exp\left(-\varpi_g\left(G\right)\right))\right\}\beta_aA\right]\\
&+E\left[\left\{G-E(G)\right\}\left\{A\exp(-\varpi_g\left(G\right))-E(A\exp\left(-\varpi_g\left(G\right)\right))\right\}\beta_u\left(U\right)\right]\\
&+E\left[\left\{G-E(G)\right\}\left\{A\exp(-\varpi_g\left(G\right))-E(A\exp\left(-\varpi_g\left(G\right)\right))\right\}\beta_g(G)\right]\\
&=\beta_aE\left[\left\{G-E(G)\right\}\left\{A\exp(-\varpi_g\left(G\right))-E(A\exp\left(-\varpi_g\left(G\right)\right))\right\}A\right]\\
&+\underbrace{E\left[\left\{G-E(G)\right\}\left[U_a-E\left(U_a\right)\right]\beta_u\left(U\right)\right]}_{=0}\\
&+\underbrace{E\left[\left\{G-E(G)\right\}\left\{U_a-E\left(U_a\right)\right\}\beta_g(G)\right]}_{=0}\\
&=\beta_aE\left[\left\{G-E(G)\right\}\left\{A-E\left(A|G\right)\right\}A\exp(-\varpi_g\left(G\right))\right]\\
&=\beta_aE\left[\left\{G-E(G)\right\}var(A|G)\exp(-\varpi_g\left(G\right))\right],
\end{aligned}$$

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

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 $\varpi_g\left(g\right)=\varpi_g g$ for vector ϖ_g, then a consistent estimator $\widehat{\varpi}_g$ of ϖ_g is given by the solution to the estimating equation:

$$\mathbb{P}_n\left[A\exp\left(-\widehat{\varpi}_gG\right)\left(G-\mathbb{P}_nG\right)\right]=0$$

Note that if A is a rare binary exposure then $var\left(A|g\right)/var\left(A|g=0\right)\approx\exp\left(\varpi_g\left(g\right)\right)$ 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 the Lemma 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 $E\left(A|G=g\right)=\exp\left(\varpi_g\left(g\right)\right)=\exp\left(\alpha_g\left(g\right)\right)E\left(U_a|G=g\right)$.

5.2 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:

$$0=\mathbb{P}_n\left[h(C)\left\{G-\widehat{E}(G|C)\right\}\left\{A-\widehat{E}(A|G,C)\right\}\left\{Y-\widehat{\beta}_aA\right\},\right] \tag{10}$$

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 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 in Assumptions 1,2,2*,3,4, as well as on the left hand-side of equation (5) . Note that effect modification can be incorporated upon conditioning on C in equation (1) , by modeling the conditional causal effect of A on Y given U and C as a function of C, for example

$$E\left(Y|A=a,G,C=c,U\right)-E\left(Y|A=0,G,C=c,U\right)=\beta_aa+\beta'_{ac}ac, \tag{11}$$

in which case β_{ac} captures effect modification by C. In contrast, effect modification by C in equation (2) can remain unrestricted, i.e.

$$E\left(Y|A=0,G=g,C=c,U\right)-E\left(Y|A=0,G=0,C=c,U\right)=\beta_{g|c}\left(g,c\right),$$

where $\beta_{g|c}\left(0,c\right)=0$ for all c but is otherwise unrestricted. Estimation of (β_a,β_{ac}) requires modifying equation (10) as followed :

$$0=\mathbb{P}_n\left[h(C)\left\{G-\widehat{E}(G|C)\right\}\left\{A-\widehat{E}(A|G,C)\right\}\left\{Y-\widehat{\beta}_aA-\widehat{\beta}'_{ac}AC\right\}\right],$$

where $h(C)$ is of the same dimension as $(A,C')'$, e.g. $h(C)=(1,C')'$.

5.3 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, suppose that G is a vector of genetic variants, then, assuming for simplicity that there is no effect modification of A by C in the outcome model, i.e. $\beta'_{ac}=0$, we propose to obtain $\widehat{\beta}_a$ by solving:

$$\widehat{\beta}_a=\arg\min_{\beta_a}\mathbb{P}_n\left[\widehat{U}'\left(\beta_a\right)\right]W\mathbb{P}_n\left[\widehat{U}\left(\beta_a\right)\right] \tag{12}$$

where

$$\widehat{U}\left(\beta_a\right)=\left\{h\left(G,C\right)-\widehat{E}(h\left(G,C\right)|C)\right\}\left\{A-\widehat{E}(A|G,C)\right\}\left\{Y-\beta_aA\right\}$$

for a user-specified function $h\left(G,C\right)$ of dimension $K\geq 1$, and W is user-specified weight matrix. In practice, it may be convenient to set $h\left(G,C\right)=G$ and $W=I_{KxK}$ 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 (12) with weight $W_{opt}=\mathbb{P}_n\left[\widehat{U}\left(\overline{\beta}_a\right)\widehat{U}\left(\overline{\beta}_a\right)'\right]^-$ where T^- denotes the generalized inverse of matrix T. Identification of GMM is guaranteed (at least locally) provided that the second derivative wrt β_a of the GMM objective function $\mathbb{P}_n\left[\widehat{U}'\left(\beta_a\right)\right]W\mathbb{P}_n\left[\widehat{U}\left(\beta_a\right)\right]$ is nonsingular at the truth, which is a generalization of condition (5). The asymptotic distribution of $\widehat{\beta}_a$ which solves (12) is described in Appendix A3.

5.4 Multiplicative causal effects

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

$$\frac{E\left(Y|A=a,G,U\right)}{E\left(Y|A=0,G,U\right)}=\exp\left(\beta_a a\right), \tag{13}$$

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 $\log\left\{\Pr\left(Y=1|A=a,G,U\right)/\Pr\left(Y=1|A=0,G,U\right)\right\}$ which is assumed to be independent of U and G,i.e. there is no multiplicative interaction between A and $\left(G,U\right)$. In order to state our identification result with an invalid IV, consider the following assumption.

Assumption 5. Equations $\left(2\right),\left(3\right)$, and $\left(13\right)$ hold.

Lemma 4 *Suppose that Assumptions 1,2*,5 hold, then β_a is the unique solution to equation:*

$$0=E\left[\left\{G-E(G)\right\}\left\{A-E(A|G)\right\}Y\exp\left\{-\beta_a A\right\}\right], \tag{14}$$

provided that condition (5) *holds.*

Proof. The results follows upon noting that $E\left[Y\exp\left\{-\beta_a A\right\}|A,G,U\right]=E\left[Y|A=0,G,U\right]$. The proof then proceeds as in Lemma 1. ■

According to the Lemma, a consistent estimator of β_a can be obtained by solving an empirical version of equation (14) in a similar manner as in previous Sections. The unbiasedness property given by equation (14) continues to hold for continuous A under the conditions given in the Lemma, and generalizations to allow for covariates and multiple IVs can easily be deduced from previous Sections.

Interestingly, equation (14) continues to hold under case-control sampling wrt 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

$$\begin{aligned}
&E\left[\{G - E(G)\}\{A - E(A|G)\}\exp\{-\beta_a A\}|Y = 1\right]\\
&\propto E\left[\{G - E(G)\}\{A - E(A|G)\}\exp\{-\beta_a A\}E(Y|A, G, U)\right]\\
&= E\left[\{G - E(G)\}\{A - E(A|G)\}\exp\{-\beta_a A\}Y\right],\\
&= 0
\end{aligned}$$

where the last equality follows from Lemma 3.

5.5 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 $\widehat{Y}_0\left(\widehat{\beta}_a\right) = Y - \widehat{\beta}_a A$ under (1) and $\widehat{Y}_0\left(\widehat{\beta}_a\right) = Y\exp\left\{-\widehat{\beta}_a A\right\}$ under (13).
3. Regress $\widehat{Y}_0\left(\widehat{\beta}_a\right)$ 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 $\widehat{\beta}_a^{opt}$ as the solution to

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

with $\widehat{Y}_0\left(\widehat{\beta}_a^{opt}\right) = Y - \widehat{\beta}_a^{opt}A$ under (1) and $\widehat{Y}_0\left(\widehat{\beta}_a^{opt}\right) = Y\exp\left\{-\widehat{\beta}_a^{opt}A\right\}$ under (13).

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 $\widehat{\beta}_a^{opt}$ 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, i..e. 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 $\left\{\widehat{Y}_0\left(\widehat{\beta}_a^{opt}\right) - \widehat{\mu}(G)\right\}$ in step 4 with $\left\{\widehat{Y}_0\left(\widehat{\beta}_a^{opt}\right)\widehat{\mu}\left(0;\widehat{\beta}_a^{opt}\right)/\widehat{\mu}\left(G;\widehat{\beta}_a^{opt}\right)\right\}$, where $\widehat{\mu}\left(G;\widehat{\beta}_a^{opt}\right)$ is the regression of $Y\exp\left\{-\widehat{\beta}_a^{opt}A\right\}$ on G under an appropriate GLM and solving the estimating equation in Step 4 for $\widehat{\beta}_a^{opt}$. 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 $\left\{A - \widehat{E}(A|G)\right\}$ in Step 4.

5.6 Odds ratio exposure model

In this Section, we briefly consider how MR GENUIS might be applied in a setting where assumption 2 is replaced by the following weaker conditional independence assumption:

Assumption 2†. IV conditional independence: $G \perp\perp U|A$;

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 also that instead of assumption 4.b, one wishes to encode the IV-exposure association on the odds ratio scale, under the following homogeneity assumption:

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

$$\text{logit } Pr\left(A=1|G=g,U\right)-\text{logit}\Pr\left(A=1|G=0,U\right)=\chi_{g}\left(g\right) \tag{15}$$

for an unknown function $\chi_{g}\left(\cdot\right)$ that satisfies $\chi_{g}\left(0\right)=0$.

We then have the following identification result for the multiplicative causal effect β_{a} of model $\left(13\right)$.

Lemma 5 *Under Assumptions 1.2†,4.b† and equations* $\left(2\right)$ *and* $\left(13\right)$, *we have that* $\beta_{a}=\theta$, *where* θ *is the unique solution to equation:*

$$0=E\left[\left\{G-E(G|A=0)\right\}\left\{A-E(A|G=0)\right\}Y\exp\left\{-\left(\varphi_{g}\left(G\right)+\theta\right)A\right\}\right]$$

where

$$\varphi_{g}\left(g\right)=logit\ Pr\left(A=1|G=g\right)-logit\Pr\left(A=1|G=0\right),$$

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

$$\begin{aligned}\gamma_{ag}(g)=&E\left(Y|A=1,G=g,u\right)-E\left(Y|A=1,G=0,u\right)\\&-E\left(Y|A=0,G=g,u\right)+E\left(Y|A=0,G=0,u\right)\neq 0.\end{aligned}$$

Assumption $2^{\dagger}$ in fact implies that $\varphi_g(.) = \chi_g(.)$ (Ma et al 2006). The Lemma establishes that under Assumptions 1, $2^{\dagger}$,4.b† and equations (2) and (13) , the multiplicative causal effect of A is identified, provided that $\gamma_{ag}(g) \neq 0$. In the proof of the Lemma given in Appendix A1, 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.

5.7 MR GENUIS 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 et al (2011), Tchetgen Tchetgen et al (2015), Li et al (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

$$h(y|A,U,G) = \beta_0(y) + \beta_a(y)\, A + \beta_g(y)G + \beta_u(y,U) \tag{16}$$

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 (16) is particularly attractive because it implies a multiplicative survival model for the joint causal effect of A and G on Y :

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

where $\mathbf{B}_a\left(y\right) = \int_0^y \beta_a(v)dv, \mathbf{B}_g\left(y\right) = \int_0^y \beta_g(v)dv$. Our objective is therefore to identify and estimate $\mathbf{B}_a\left(y\right)$. 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 6 *Under assumptions 1,2,4.b and equation* (16), *we have that for each* y

$$0 = E\left\{W\left(y, \mathbf{B}_a\left(y\right), \mathbf{B}_g\left(y\right)\right)\right\}, \tag{17}$$

where

$$W\left(y, \mathbf{B}_a\left(y\right), \mathbf{B}_g\left(y\right)\right) = \left(dN(y) - d\mathbf{B}_a\left(y\right)A - d\mathbf{B}_g\left(y\right)G\right)\exp\left\{\mathbf{B}_a\left(y\right)A + \mathbf{B}_g\left(y\right)G\right\}R(y)h(G, A),$$

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

Proof. We note that by assumption

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

and

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

Therefore

$$\begin{aligned}&E\left\{W\left(y,\mathbf{B}_a\left(y\right),\mathbf{B}_g\left(y\right)\right)\right\}\\&=E\left\{\left(d\mathbf{B}_0\left(y\right)+d\mathbf{B}_u\left(y,U\right)\right)\exp\left\{-\mathbf{B}_0\left(y\right)-\mathbf{B}_u\left(y,U\right)\right\}\begin{pmatrix}(G-E(G))\\(G-E(G))\left(A-E(A|G)\right)\end{pmatrix}\right\}\\&=E\left\{\begin{pmatrix}0\\\left(d\mathbf{B}_0\left(y\right)+d\mathbf{B}_u\left(y,U\right)\right)\exp\left\{-\mathbf{B}_0\left(y\right)-\mathbf{B}_u\left(y,U\right)\right\}\left(U-E(U)\right)\left(G-E(G)\right)\end{pmatrix}\right\}\\&=0.\end{aligned}$$

■

As in Martinussen et al (2017), the unbiasedness of equation $W\left(y,\mathbf{B}_a\left(y\right),\mathbf{B}_g\left(y\right)\right)$ suggests a way of estimating the increments $\left(d\mathbf{B}_a\left(y\right),d\mathbf{B}_g\left(y\right)\right)$ by solving an empirical version of equation (17) for each y with population expectations replaced by sample analogs, giving the following recursive estimator

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

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

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

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.

6 Simulation Study

6.1 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), $i = 1, 2, ..., n$ as follows:

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

where for binary exposure A,

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

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

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

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 using (6), with $\hat{E}(A|G)$ 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. Coverage of the Wald-type 95% confidence interval (CI) for the causal parameter is also close to nominal level. Efficiency of the estimators increases with IV strength. When the IV is invalid, TSLS is biased and its 95% CI undercovers, while in accordance with theory MR GENIUS continues to have small bias and correct coverage.

Multiple IVs

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

$$\begin{aligned} G_{ij} &\sim \text{Bernoulli}(p = 0.5), j = 1, 2, ..., p_G \\ Y_i &\sim \text{N}(\alpha^T G_i + \beta A_i + U_i, 1^2), \end{aligned}$$

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

$$\begin{aligned} &\epsilon_i \sim \text{truncated N}(a = 0.2, b = 0.5, \mu = 0.35, \sigma^2 = 1^2), \\ &U_i = {\phi_b}^T G_i + \epsilon_i, \\ &A_i \sim \text{Bernoulli}\left(p_i = \frac{\exp\left({\gamma_b}^T G_i\right)}{1 + \exp\left({\gamma_b}^T G_i\right)} + [U_i - E(U|G_i)]\right), \end{aligned}$$

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

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

For binary exposure, IV strength is set to -0.15 for each entry of γ_b, while in the continuous exposure case each entry of γ_c and λ_1 is set identically to -2 and 0.5 respectively. 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, $\alpha^T = -0.5 \cdot (1, 1, 1, 0, ..., 0)$ and $\phi_c^T = -0.25 \cdot (1, 1, 1, 0, ..., 0), \phi_b^T = -0.05 \cdot (1, 1, 1, 0, ..., 0)$ when Assumption 3 or 2 is violated, respectively; with six invalid IVs, $\alpha^T = -0.25(1, 1, 2, 2, 4, 4, 0, ..., 0)$ and $\phi_c^T = -0.25 \cdot (0.5, 0.5, 1, 1, 2, 2, 0, ..., 0), \phi_b^T = -0.01 \cdot (1, 1, 3, 3, 5, 5, 0, ..., 0)$ accordingly. When all IVs are invalid, $\alpha^T = -0.5 \cdot (1, 1, ..., 1)$ and $\phi_c^T = -0.25 \cdot (1, 1, ..., 1), \phi_b^T = -0.02 \cdot (1, 1, ..., 1)$. 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, 2016); 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, 2017).

MR GENIUS is implemented as the solution to (12) with optimal weight; the more efficient version of MR GENIUS as described in section 5.4 is also implemented. MR-Egger regression estimation, TSLS (which assumes all IVs are valid) and sisVIVE are implemented using the R packages `MendelianRandomization`, `AER` and `sisVIVE` (Yavorska and Burgess, 2017; 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 (2016) and Guo et al (2017) 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.

Simulation results based on 1000 replications for sample sizes of $n = 500$, 2000 and $10,000$ with continuous exposure are presented in Tables 3 and 4. 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 $n = 10,000$. Empirical coverage of CIs is close to nominal level once $n \geq 2000$ for adaptive/post-adaptive Lasso and TSHT. 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 7 for results on IV selection). The naive TSLS estimator performs well in terms of bias and coverage only when all IVs are valid; as expected, it is biased and its 95% CI severely undercovers 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 once $n \geq 2000$ (when the majority rule holds), in agreement with theory since they are asymptotically equivalent under these settings. MR GENIUS also has small bias at all sample sizes and its bias becomes negligible at $n = 10000$, with adequate 95% CI empirical coverage at $n \geq 500$. MR GENIUS is generally less efficient than the other estimators when the majority rule holds, except for MR-Egger. MR-Egger exhibits some bias, but its 95% CI coverage is adequate when there are no invalid IVs, with slight undercoverage when there are three invalid IVs. Since MR-Egger assumes a two-sample analysis whereby association coefficients relating IVs and exposure/outcome are uncorrelated, the observed bias may be a reflection of the single sample simulation setting.

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. There is also increasing undercoverage of 95% CI as sample size increases for post-adaptive Lasso. On average, sisVIVE and adaptive Lasso select eight IVs as invalid when only six are actually invalid, and fails to select any IV as invalid when all are. TSHT is also biased and its 95% CI undercovers when all IVs are invalid (with none 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 $n = 10,000$. Adequate 95% CI coverage is also achieved at $n = 10,000$, with the right number of IVs selected as invalid on average. The efficiency of estimators generally decreases with increasing number of invalid IVs, except when all the IVs are invalid. The bias of MR GENIUS improves with increasing sample size when six or all IVs are invalid; adequate 95% CI coverage is achieved at $n \geq 2000$ with six invalid IVs, and at $n = 10,000$ with all IVs invalid. Efficiency comparisons MR GENIUS is again generally less efficient than TSHT when six IVs are invalid, however it outperforms MR-Egger even when the InSIDE assumption holds (Bowden et al, 2015). However, MR-Egger is generally more biased, with severe 95% CI undercoverage, when the invalid IVs violate both Assumptions 2 and 3, which corresponds to a violation of the InSIDE assumption. The efficient MR GENIUS is generally less biased and more efficient compared to MR GENIUS, especially when more IVs are invalid. The estimated asymptotic relative efficiency of efficient MR GENIUS to MR GENIUS is approximately 0.5 with 10 invalid IVs (which violate both assumptions 2 and 3); the 95% CIs based on efficient MR GENIUS also attain correct coverage across all the scenarios at $n \geq 2000$.

Simulation results with a binary exposure are summarized in Tables 5 and 6; the conclusions are mostly qualitatively similar to those in the continuous exposure setting. However, when there are six invalid IVs, TSHT is biased and its 95% CI undercovers, 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 nine on average at $n = 10,000$); 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 three on average, when six are valid at $n = 10,000$). MR-Egger appears to be less biased compared to the continuous exposure setting; this may be due to smaller values of ϕ_b used for binary exposure setting, so that violation of the InSIDE assumption is less severe.

7 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

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	0.00	0.00	0.00	0.00	0.00	0.00
	0.00	0.00	0.00	0.00	0.00	0.00
TSLS	0.00	0.50	0.83	0.00	0.51	0.84
	0.00	0.50	0.83	0.00	0.50	0.83
Median estimated SD						
MR GENIUS	0.08	0.08	0.08	0.02	0.02	0.02
	0.06	0.06	0.06	0.01	0.01	0.01
TSLS	0.13	0.12	0.05	0.14	0.22	0.11
	0.09	0.09	0.03	0.09	0.15	0.08
Monte Carlo SD[‡]						
MR GENIUS	0.08	0.08	0.08	0.02	0.02	0.02
	0.06	0.06	0.06	0.01	0.01	0.01
TSLS	0.12	0.13	0.05	0.13	0.25	0.11
	0.09	0.09	0.04	0.09	0.15	0.08
95% Wald-type CI coverage						
MR GENIUS	95.3	95.3	95.3	94.8	94.8	94.8
	94.8	94.8	94.8	95.2	95.2	95.2
TSLS	95.3	2.7	0.0	99.0	36.9	0.0
	95.9	0.1	0.0	97.7	5.8	0.0

[†]: `TTT`: IV assumptions (1), (2) and (3) hold; `TTF`: IV assumption (3) (exclusion restriction) does not hold; `TFF`: both IV assumptions (2) and (3) (IV independence) do not hold.

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

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.

		$\lvert\gamma_b\rvert = 0.5$			$\lvert\gamma_b\rvert = 1$		
		TTT[†]	TTF	TFF	TTT	TTF	TFF
Median absolute value of bias							
	MR GENIUS	0.01	0.01	0.01	0.03	0.03	0.01
		0.01	0.01	0.00	0.01	0.01	0.00
	TSLS	0.00	4.05	2.16	0.00	2.17	1.61
		0.02	4.07	2.18	0.01	2.18	1.61
Median estimated SD							
	MR GENIUS	0.54	0.54	0.33	0.36	0.36	0.34
		0.37	0.37	0.23	0.25	0.25	0.24
	TSLS	0.77	1.39	0.38	0.40	0.53	0.25
		0.53	0.97	0.27	0.28	0.37	0.18
Monte Carlo SD[‡]							
	MR GENIUS	0.54	0.54	0.32	0.35	0.35	0.34
		0.39	0.39	0.22	0.25	0.25	0.24
	TSLS	0.79	1.37	0.37	0.41	0.52	0.27
		0.55	1.06	0.28	0.29	0.37	0.18
95% Wald-type CI coverage							
	MR GENIUS	98.7	98.7	96.0	95.2	95.2	95.8
		96.9	96.9	96.4	95.9	95.9	94.9
	TSLS	98.6	9.7	0.0	95.2	0.0	0.0
		96.4	0.3	0.0	94.6	0.0	0.0

[†]: TTT: IV assumptions (1), (2) and (3) hold; TTF: IV assumption (3) (exclusion restriction) does not hold; TFF: both IV assumptions (2) and (3) (IV independence) do not hold.

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

Table 3: Median absolute value of bias and Monte Carlo standard error in estimation of $\beta_0 = 0.5$ with continuous exposure and $p_G = 10$ IVs. All entries are original values multiplied by 100. The three rows of results for each estimator correspond to sample sizes of $n = 500$, $n = 2000$ and $n = 10,000$ respectively.

		TTF†			TFF		
#invalid IV	0	3	6	10	3	6	10
Median absolute value of bias							
MR GENIUS	0.9	1.5	2.5	3.4	2.0	3.8	5.1
	0.1	0.2	0.6	0.8	0.3	0.9	1.2
	0.1	0.1	0.2	0.3	0.1	0.2	0.3
Efficient MR GENIUS	1.0	1.2	1.4	1.6	1.3	1.7	2.0
	0.1	0.2	0.2	0.3	0.2	0.3	0.3
	0.1	0.1	0.1	0.1	0.1	0.1	0.2
TSLS	0.2	7.6	17.3	24.6	11.7	25.8	32.9
	0.1	7.5	17.4	24.9	11.7	26.1	33.2
	0.0	7.5	17.5	25.0	11.7	26.1	33.3
Oracle TSLS		0.2	0.4		0.2	0.4	
	-	0.1	0.1	-	0.1	0.1	-
		0.0	0.0		0.0	0.0	
sisVIVE	0.2	6.0	13.0	24.6	6.4	19.6	32.9
	0.1	2.9	12.5	24.9	3.2	19.0	33.2
	0.0	1.2	12.5	25.0	1.4	18.2	33.3
ALasso	0.2	5.2	11.6	24.6	5.0	15.8	32.8
	0.1	2.2	10.4	24.9	2.4	16.3	33.2
	0.0	0.9	12.2	25.0	1.0	17.6	33.3
post-ALasso	0.3	2.2	9.8	24.5	0.5	15.1	32.8
	0.1	0.2	11.5	24.9	0.1	17.5	33.2
	0.0	0.0	12.5	25.0	0.0	17.7	33.3
TSHT	0.2	5.7	9.8	24.6	1.9	14.1	32.8
	0.1	0.1	6.7	24.9	0.1	4.6	33.2
	0.0	0.0	0.0	25.0	0.0	0.0	33.3
MR-Egger	4.9	7.9	10.3	14.2	34.0	71.8	18.9
	4.5	8.5	14.1	13.9	85.9	151.9	18.3
	4.8	4.8	11.1	14.2	196.8	252.3	18.9
Monte Carlo SD‡							
MR GENIUS	4.4	4.4	4.7	4.6	4.7	5.1	5.0
	2.3	2.4	2.5	2.6	2.5	3.0	3.1
	1.0	1.1	1.2	1.2	1.1	1.3	1.4
Efficient MR GENIUS	4.6	4.8	4.8	4.8	4.7	4.8	4.8
	2.3	2.3	2.4	2.4	2.3	2.4	2.3
	1.0	1.0	1.0	1.0	1.0	1.0	1.0
TSLS	1.9	2.0	2.2	2.2	2.1	2.4	2.2
	1.0	1.1	1.2	1.1	1.1	1.2	1.1
	0.5	0.5	0.5	0.5	0.5	0.6	0.5
Oracle TSLS		2.4	3.1		2.4	3.1	
	-	1.2	1.6	-	1.2	1.6	-
		0.5	0.7		0.5	0.7	
sisVIVE	1.9	2.3	2.8	2.2	2.3	3.9	2.2
	1.0	1.3	2.3	1.1	1.3	2.4	1.1
	0.5	0.6	1.0	0.5	0.6	1.0	0.5
ALasso	1.9	2.5	3.2	2.2	2.3	4.8	2.2
	1.0	1.2	3.7	1.1	1.2	3.5	1.1
	0.5	0.6	0.9	0.5	0.6	1.0	0.5
post-ALasso	2.0	4.0	4.7	2.2	2.6	6.9	2.2
	1.0	1.3	4.5	1.1	1.2	2.5	1.1
	0.5	0.5	1.0	0.5	0.5	0.9	0.5
TSHT	1.9	3.0	3.1	2.2	4.8	4.1	2.2
	1.0	1.2	3.0	1.1	1.2	3.6	1.1
	0.5	0.5	0.8	0.5	0.5	0.7	0.5
MR-Egger	11.0	25.3	38.2	11.5	28.8	36.1	11.5
	9.9	45.5	74.5	10.1	34.4	35.9	10.9
	11.0	97.3	158.4	11.0	36.2	32.2	10.7

†: For the invalid IVs, TTF: IV assumption (3) (exclusion restriction) does not hold; TFF: both IV assumptions (2) and (3) (IV independence) do not hold.

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

Table 4: Ratio of estimated to Monte Carlo standard error and empirical 95% Wald-type CI coverage in estimation of $\beta_0 = 0.5$ with continuous exposure and $p_G = 10$ IVs. The three rows of results for each estimator correspond to sample sizes of $n = 500$, $n = 2000$ and $n = 10,000$ respectively. Only point estimation is implemented for sisVIVE and adaptive Lasso, hence their results are not available.

			TTF†			TFF	
#invalid IV	0	3	6	10	3	6	10
Median estimated standard error / Monte Carlo SD							
MR GENIUS	1.1	1.1	1.1	1.1	1.1	1.1	1.1
	1.0	1.0	1.1	1.1	1.0	1.0	1.1
	1.0	1.0	1.0	1.0	1.1	1.1	1.1
Efficient MR GENIUS	1.0	1.0	1.0	1.0	1.0	1.0	1.0
	1.0	1.0	1.0	1.0	1.0	1.0	1.0
	1.0	1.0	1.0	1.0	1.0	1.0	1.0
TSLS	1.0	1.0	1.0	1.0	1.0	0.9	1.0
	1.0	0.9	0.9	1.0	0.9	0.9	1.0
	1.0	1.0	0.9	1.0	1.0	0.9	1.0
Oracle TSLS	-	1.0	1.0	-	1.0	1.0	-
		1.0	1.0		1.0	1.0	
		1.0	1.0		1.0	1.0	
post-ALasso	1.0	0.6	0.5	1.0	0.9	0.4	1.0
	1.0	0.9	0.4	1.0	1.0	0.9	1.0
	1.0	1.0	1.0	1.0	1.0	1.0	1.0
TSHT	1.0	0.7	0.7	1.0	0.5	0.6	1.0
	1.0	1.0	0.6	1.0	1.0	0.5	1.0
	1.0	1.0	0.9	1.0	1.0	1.0	1.0
MR-Egger	2.0	1.0	0.9	2.2	1.0	1.1	2.4
	2.2	0.9	0.9	2.5	1.3	1.3	2.5
	2.0	0.9	0.9	2.3	1.3	1.1	2.5
Empirical 95% Wald-type CI coverage							
MR GENIUS	96.1	94.5	92.2	88.5	93.9	90.1	82.6
	96.4	96.2	94.6	93.2	95.9	93.6	91.2
	95.2	95.6	95.2	95.0	95.4	94.4	94.4
Efficient MR GENIUS	94.4	93.8	92.6	91.9	92.8	92.2	90.7
	96.0	95.8	95.6	95.5	95.7	95.3	95.0
	95.1	95.1	95.2	95.2	95.2	95.1	95.0
TSLS	94.4	4.5	0.0	0.0	0.0	0.0	0.0
	95.3	0.0	0.0	0.0	0.0	0.0	0.0
	94.1	0.0	0.0	0.0	0.0	0.0	0.0
Oracle TSLS	-	94.6	94.6	-	94.6	94.6	-
		94.9	94.6		94.9	94.6	
		95.0	95.8		95.0	95.8	
post-ALasso	94.2	65.9	13.4	0.0	89.9	4.8	0.0
	95.3	93.9	1.1	0.0	94.8	0.0	0.0
	93.9	94.8	0.0	0.0	94.8	0.0	0.0
TSHT	94.6	28.9	4.3	0.0	63.2	3.2	0.0
	95.4	94.4	15.4	0.0	94.8	49.2	0.0
	94.4	95.2	95.2	0.0	95.2	94.8	0.0
MR-Egger	100.0	94.9	88.4	99.9	73.2	53.7	99.7
	100.0	91.8	90.4	99.8	49.5	15.8	99.7
	100.0	91.2	92.1	99.8	4.3	0.0	99.8

†: For the invalid IVs, TTF: IV assumption (3) (exclusion restriction) does not hold; TFF: both IV assumptions (2) and (3) (IV independence) do not hold.

Table 5: Median absolute value of bias and Monte Carlo standard error in estimation of $\beta_0 = 0.5$ with binary exposure and $p_G = 10$ IVs. All entries are original values multiplied by 100. The three rows of results for each estimator correspond to sample sizes of $n = 500$, $n = 2000$ and $n = 10,000$ respectively.

		TTF[†]			TFF		
#invalid IV	0	3	6	10	3	6	10
Median absolute value of bias							
MR GENIUS	0.7	14.6	28.2	35.8	16.3	29.6	36.6
	0.4	6.3	12.4	15.1	7.0	13.3	16.4
	0.4	0.9	2.7	3.3	1.0	2.8	3.7
Efficient MR GENIUS	2.3	1.6	1.0	0.1	1.2	0.9	0.2
	1.0	0.9	0.0	0.8	0.8	0.2	0.9
	0.3	0.4	0.5	0.7	0.4	0.6	0.7
TSLS	1.0	177.9	433.4	616.8	197.0	455.3	642.0
	0.1	328.3	788.9	1,112.9	361	828.2	1,156.9
	0.6	431.6	1,006.4	1,436.1	474.1	1,057.7	1,493.4
Oracle TSLS		3.7	0.6		3.7	0.6	
	-	0.4	1.2	-	0.4	1.2	-
		0.4	0.7		0.4	0.7	
sisVIVE	1.0	82.9	302.1	616.2	81.9	311.0	642.0
	0.2	75.5	568.6	1,111.1	74.5	590.2	1,156.7
	0.6	46.9	716.3	1,435.6	46.8	747.8	1,492.7
ALasso	0.9	56.0	217.1	616.2	54.8	221.0	641.8
	0.1	45.7	468.4	1,112.2	44.7	482.7	1,156.7
	0.7	29.2	585.1	1,434.5	29.1	611.0	1,492.0
post-ALasso	0.8	13.1	180.0	615.8	11.8	181.7	641.3
	1.0	0.8	395.0	1,111.1	0.4	408.8	1,156.0
	0.7	0.2	558.7	1,432.7	0.2	579.8	1,490.7
TSHT	1.6	97.0	322.2	514.1	97.6	340.3	531.3
	2.4	15.1	270.0	874.3	14.8	273.2	909.9
	0.4	0.2	408.3	1,401.2	0.2	424.8	1,456.9
MR-Egger	8.8	165.1	339.3	447.4	182.5	356.8	463.8
	2.4	133.6	333.8	171.7	148.0	353.4	175.6
	0.4	10.8	45.2	17.9	15.1	47.7	17.7
Monte Carlo SD[‡]							
MR GENIUS	69.3	76.1	91.4	88.5	76.9	93.4	89.7
	48.2	54.5	65.9	69.1	55.3	66.6	71.0
	26.3	28.7	35.0	35.9	29.5	35.5	36.4
Efficient MR GENIUS	67.9	68.4	67.8	67.4	68.4	67.7	67.1
	48.4	48.2	48.2	48.7	48.2	48.3	48.8
	26.1	26.1	26.3	26.1	26.1	26.3	26.1
TSLS	52.9	141.8	240.0	186.0	151.7	253.0	192.9
	36.0	131.0	204.0	185.1	143.9	212.8	191.3
	18.1	75.8	129.8	119.1	82.9	136.3	123.3
Oracle TSLS		65.0	89.0		65.0	89.0	
	-	46.8	60.7	-	46.8	60.7	-
		22.3	29.2		22.3	29.2	
sisVIVE	52.8	101.9	231.9	187.2	101.9	244.9	195.3
	36.0	59.7	266.6	187.6	59.7	277.1	192.5
	18.1	24.5	172.2	121.7	24.6	182.4	126.2
ALasso	52.8	70.6	210.1	186.1	65.3	219.7	193.5
	36.0	38.9	261.4	184.5	38.1	285.4	191.6
	18.0	22.0	170.0	121.4	21.9	186.4	125.9
post-ALasso	52.3	75.6	206.5	185.5	70.0	214.2	191.7
	35.3	47.8	240.4	187.5	47.9	259.6	193.2
	18.0	22.3	186.2	123.4	22.3	204.7	128.7
TSHT	64.2	221.6	391.0	193.2	239.5	414.7	201.7
	48.4	69.6	463.9	125.2	68.5	498.4	128.8
	18.1	22.5	222.5	115.8	22.5	236.4	119.8
MR-Egger	101.2	330.2	559.7	409.1	363.6	590.7	427.1
	78.4	508.6	898.7	590.6	558.5	947.6	614.5
	73.8	1,055.8	1,608.0	94.8	1,161.4	1,693.6	95.6

[†]: For the invalid IVs, TTF: IV assumption (3) (exclusion restriction) does not hold; TFF: both IV assumptions (2) and (3) (IV independence) do not hold.

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

Table 6: Ratio of estimated to Monte Carlo standard error and empirical 95% Wald-type CI coverage in estimation of $\beta_0 = 0.5$ with binary exposure and $p_G = 10$ IVs. The three rows of results for each estimator correspond to sample sizes of $n = 500$, $n = 2000$ and $n = 10,000$ respectively. Only point estimation is implemented for sisVIVE and adaptive Lasso, hence their results are not available.

		TTF[†]			TFF		
#invalid IV	0	3	6	10	3	6	10
Median estimated standard error / Monte Carlo SD							
MR GENIUS	1.2	1.2	1.2	1.2	1.2	1.2	1.2
	1.1	1.0	1.0	1.0	1.0	1.0	1.0
	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Efficient MR GENIUS	1.2	1.2	1.2	1.2	1.2	1.2	1.2
	1.1	1.1	1.1	1.1	1.1	1.1	1.1
	1.0	1.0	1.0	1.0	1.0	1.0	1.0
TSLS	1.1	0.5	0.5	0.9	0.5	0.5	0.9
	1.0	0.5	0.7	1.0	0.5	0.7	1.0
	1.1	0.6	0.7	1.1	0.5	0.7	1.1
Oracle TSLS		1.1	1.1		1.1	1.1	
	-	1.0	1.0	-	1.0	1.0	-
		1.0	1.0		1.0	1.0	
post-ALasso	1.1	1.0	0.5	0.9	1.0	0.5	0.9
	1.1	1.0	0.4	1.0	1.0	0.4	1.0
	1.1	1.0	0.4	1.0	1.0	0.4	1.0
TSHT	1.1	0.4	0.3	1.0	0.4	0.3	1.0
	1.0	0.8	0.2	1.6	0.8	0.2	1.6
	1.1	1.0	0.2	1.1	1.0	0.2	1.1
MR-Egger	1.1	1.0	1.0	1.1	1.0	1.0	1.1
	1.1	0.9	0.9	0.3	0.9	0.9	0.3
	1.1	0.9	0.9	1.1	0.9	0.9	1.1
Empirical 95% Wald-type CI coverage							
MR GENIUS	99.5	99.6	98.9	97.5	99.5	98.9	97.6
	97.5	97.6	96.6	96.9	97.7	96.6	97.1
	95.9	95.9	95.4	94.3	95.5	95.6	94.5
Efficient MR GENIUS	99.8	99.8	99.8	99.9	99.8	99.8	99.9
	97.5	97.4	97.6	97.7	97.4	97.6	97.7
	96.2	96.2	96.2	96.2	96.2	96.2	96.2
TSLS	97.2	36.3	12.1	3.4	33.0	11.8	3.4
	95.5	2.7	0.2	0.0	2.5	0.2	0.0
	95.5	0.0	0.0	0.0	0.0	0.0	0.0
Oracle TSLS		98.4	99.4		98.4	99.4	
	-	97.1	98.1	-	97.1	98.1	-
		94.8	95.4		94.8	95.4	
post-ALasso	97.6	93.0	57.7	3.3	94.3	57.7	3.3
	95.9	96.8	21.1	0.0	96.9	21.3	0.0
	95.4	94.7	0.5	0.0	94.7	0.1	0.0
TSHT	98.9	57.6	36.9	4.1	57.9	36.4	3.4
	97.4	83.6	38.7	0.1	84.2	39.1	0.0
	94.9	95.0	5.7	0.0	95.0	5.6	0.0
MR-Egger	97.2	88.7	86.4	85.1	88.3	86.5	85.0
	96.4	90.1	86.6	84.2	90.0	86.7	84.2
	96.3	92.2	91.7	96.7	92.1	91.5	96.6

[†]: For the invalid IVs, TTF: IV assumption (3) (exclusion restriction) does not hold; TFF: both IV assumptions (2) and (3) (IV independence) do not hold.

Table 7: 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 three rows of results for each estimator correspond to sample sizes of $n = 500$, $n = 2000$ and $n = 10,000$ respectively.

	TTF†				TFF		
#invalid IV	0	3	6	10	3	6	10
Continuous exposure							
ALasso	0.0	2.3	3.9	0.0	3.1	5.3	0.0
	0.0	3.1	6.9	0.0	3.0	7.7	0.0
	0.0	3.0	8.0	0.0	3.0	8.1	0.0
sisVIVE	0.0	2.6	5.6	0.0	4.1	7.1	0.0
	0.0	3.9	8.0	0.0	4.2	8.2	0.0
	0.0	4.0	8.1	0.0	4.2	8.2	0.0
TSHT ($\hat{I}$)	0.0	0.8	2.1	0.0	2.4	3.1	0.0
	0.0	3.0	6.8	0.0	3.0	6.9	0.0
	0.0	3.0	6.3	0.0	3.0	6.3	0.0
TSHT ($\hat{S}$)	10.0	10.0	10.0	10.0	10.0	10.0	10.0
	10.0	10.0	10.0	10.0	10.0	10.0	10.0
	10.0	10.0	10.0	10.0	10.0	10.0	10.0
Binary exposure							
ALasso	0.1	3.3	4.3	0.2	3.3	4.4	0.2
	0.0	3.1	4.5	0.1	3.1	4.5	0.1
	0.0	3.0	5.3	0.1	3.0	5.4	0.1
sisVIVE	0.0	3.6	4.6	0.2	3.7	4.6	0.2
	0.0	4.2	5.4	0.3	4.2	5.4	0.3
	0.0	4.2	7.7	0.3	4.1	7.7	0.2
TSHT ($\hat{V}$)	0.0	0.4	0.3	0.0	0.4	0.3	0.0
	0.0	0.7	0.7	0.0	0.7	0.7	0.0
	0.0	2.7	2.8	0.0	2.7	2.8	0.0
TSHT ($\hat{S}$)	5.1	5.1	5.1	5.1	5.1	5.1	5.1
	3.2	3.2	3.2	3.2	3.2	3.2	3.2
	9.1	9.1	9.1	9.1	9.1	9.1	9.1

†: For the invalid IVs, `TTF`: IV assumption (3) (exclusion restriction) does not hold; `TFF`: both IV assumptions (2) and (3) (IV independence) do not hold.

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, 2007). From the 12,123 participants for whom genotype data was available, we restricted the sample to 7,738 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 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 et al. 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 follow-up. 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, and were strongly associated as a set with the 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.

Table 8 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

Table 8: Estimation of $\beta_{\text{t2d-ms}}$, the association between type 2 diabetes and memory score.

	$\hat{\beta}_{\text{t2d-ms}}$	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)

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=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.

8 Concluding Remarks

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. First, the approach bypasses the overly stringent orthogonality condition of MR-Egger. Furthermore, 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. Furthermore, as we have shown, MR GENIUS equally applies with continuous or binary outcome, continous or binary exposure and IV, multiple IVs, auxiliary pre-IV covariates and under both prospective and retrospective sampling designs. Finally, whereas adaptive lasso and TSTH require one or more model selection steps therefore compromising inferences that are uniform over the entire model of interest, MR GENIUS does not involve model selection, therefore bypassing this difficulty.

MR GENIUS also stands out from other methods because it does not require modeling the effects of invalid IVs on Y for consistent estimation of the effect of exposure, therefore allowing main effects and interactions among components of G to remain unrestricted in the outcome model. In the event of an interaction between A and G, such that equation (1) does not hold, it is straightforward to show that MR GENUIS estimates a certain weighted average of the causal effect. For instance, in the case of binary A, $\mu = E\left(\beta\left(G\right)w\left(G\right)\right)$ where $\beta\left(G\right) = E\left(Y|A=1,G\right) - E\left(Y|A=0,G\right)$ is the causal effect within levels of G and $w(G) = \left(G - E(G)\right)var(A|G) \times E\{\left(G - E(G)\right)var(A|G)\}^{-1}$; or equivalently $\mu = \beta\left(0\right) + \left(\beta\left(1\right) - \beta\left(0\right)\right) \times var(A|G=1)var(G) \times E\{\left(G - E(G)\right)var(A|G)\}^{-1}$. Therefore, MR GENUIS is guaranteed to be consistent under the null hypothesis of no conditional effect of exposure within levels of G provided there is no interaction between U and G in the outcome model. An R package which implements MR GENIUS is available at `github.com/bluosun/MR-GENIUS`.

In closing, we acknowledge certain limitations of MR GENIUS. First, the approach may be vulnerable to weak IV bias which may occur if $var(A|G)$ is weakly dependent on G, a possibility that was largely ruled out in this paper. 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.

9 Acknowledgment

Eric Tchetgen Tchetgen's work is funded by NIH grant R01AI104459. The Health and Retirement Study genetic data are sponsored by the National Institute on Aging (grant numbers U01AG009740, RC2AG036495, and RC4AG039029) and was conducted by the University of Michigan. The authors thank Frank Windmeijer for valuable discussions.

References

[1] Bowden, J., G.D. Smith, S. Burgess, (2015). Mendelian Randomization with Invalid Instruments: Effect Estimation and Bias Detection through Egger Regression, International Journal of Epidemiology 44, 512-525.

[2] Bowden J, Davey Smith G, Haycock PC, Burgess S.(2016) Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. Genet Epidemiol. 40:304–314.

[3] Chao, J.C. and Swanson, N.R., (2005). Consistent estimation with a large number of weak instruments. Econometrica, 73(5), pp.1673-1692.

[4] Davey Smith, G. and Ebrahim, S., (2003). 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease?. International journal of epidemiology, 32(1), pp.1-22.

[5] Demmer RT, Zuk AM, Rosenbaum M and Desvarieux M. Prevalence of diagnosed and undiagnosed type 2 diabetes mellitus among US adolescents: results from the continuous NHANES, 1999–2010. American Journal of Epidemiology 2013;178(7):1106-1113.

[6] Guo, Z., H. Kang, T. Cai and D. Small, (2016), Confidence Intervals for Causal effects with Invalid Instruments using Two-Stage Hard Thresholding, arXiv:1603.05224.

[7] Han, C., (2008), Detecting Invalid Instruments using L -GMM, Economics Letters 101, 285-287.

[8] Juster FT and Suzman R. An Overview of the Health and Retirement Study. The Journal of human resources 1995;30 (Special Issue on the Health and Retirement Study: Data Quality and Early Results):S7-S56.

[9] Kleiber, C. and Zeileis, A. (2008). Applied Econometrics with R. Springer-Verlag, New York. ISBN 978-0-387-77316-2.

[10] Kang, H. (2017). sisVIVE: Some Invalid Some Valid Instrumental Variables Estimator. R package version 1.4.

[11] Kang, H., Zhang, A., Cai, T. and Small, D., (2016). Instrumental variables estimation with some invalid instruments and its application to Mendelian randomization. Journal of the American Statistical Association, 111(513), pp.132-144.

[12] Lawlor, D.A., Harbord, R.M., Sterne, J.A., Timpson, N. and Davey Smith, G.,(2008). Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. Statistics in medicine, 27(8), pp.1133-1163.

[13] Leeb, H. and Pötscher, B.M., (2008). Sparse estimators and the oracle property, or the return of Hodges' estimator. Journal of Econometrics, 142(1), pp.201-211.

[14] Lewbel, A. (2012). Using Heteroscedasticity to Identify and Estimate Mismeasured and Endogenous Regressor Models. Journal of Business & Economic Statistics, 30(1), pp.67-80.

[15] Li, J., Fine, J. and Brookhart, A., (2015). Instrumental variable additive hazards models. Biometrics, 71(1), pp.122-130.

[16] Little, J., and Khoury, M. J. (2003). Mendelian Randomisation: A New Spin or Real Progress? The Lancet, 362, 930–931.

[17] Ma, Z., Xie, X. and Geng, Z. (2006). Collapsibility of distribution dependence. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 68(1), pp.127-133.

[18] Mailman MD, Feolo M, Jin Y, Kimura M, Tryka K, Bagoutdinov R, Hao L,Kiang A, Paschall J, Phan L, Popova N, Pretel S, Ziyabari L, Lee M, Shao Y, Wang ZY, Sirotkin K, Ward M,

Kholodov M, Zbicz K, Beck J, Kimelman M, Shevelev S, Preuss D, Yaschenko E, Graeff A, Ostell J, Sherry ST. The NCBI dbGaP Database of Genotypes and Phenotypes. Nat Genet. 2007 Oct; 39(10):1181-6.

[19] 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., Chakravarti, A. and Cho, J.H., (2009). Finding the missing heritability of complex diseases. Nature, 461, pp.747-753.

[20] 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, in press.

[21] Morris AP, Voight BF, Teslovich TM, et al. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. Nature genetics 2012;44(9):981.

[22] Mozumdar A and Liguori G. Persistent increase of prevalence of metabolic syndrome among US adults: NHANES III to NHANES 1999–2006. Diabetes Care 2011;34(1):216-219.

[23] 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(4), pp.1397-1405.

[24] Ofstedal MB, Fisher GG, Herzog AR. Documentation of cognitive functioning measures in the Health and Retirement Study. Ann Arbor, MI: Univ. of Michigan; 2005.

[25] Robins JM. (1997). Causal Inference from Complex Longitudinal Data. Latent Variable Modeling and Applications to Causality. Lecture Notes in Statistics (120), M. Berkane, Editor. NY: Springer Verlag, pp. 69-117.

[26] Smith, G.D. and Ebrahim, S., (2004). Mendelian randomization: prospects, potentials, and limitations. International journal of epidemiology, 33(1), pp.30-42.

[27] Staiger, D., and Stock, J. H. (1997), "Instrumental Variables Regression With Weak Instruments," Econometrica, 65, 557–586.

[28] Stock, J.H., Wright, J.H. and Yogo, M., (2002). A survey of weak instruments and weak identification in generalized method of moments. Journal of Business & Economic Statistics, 20(4), pp.518-529.

[29] Stock, J. H., and Wright, J. H. (2000), "GMM With Weak Identification," Econometrica, 68, 1055–1096.

[30] Tchetgen Tchetgen, E.J., Robins, J.M. and Rotnitzky, A., (2009). On doubly robust estimation in a semiparametric odds ratio model. Biometrika, 97(1), pp.171-180.

[31] Tchetgen Tchetgen, E.J., Walter, S., Vansteelandt, S., Martinussen, T. and Glymour, M., (2015). Instrumental variable estimation in a survival context. Epidemiology, 26(3), p.402.

[32] Wang, L. and Tchetgen Tchetgen, E. (2016). Bounded, efficient and multiply robust estimation of average treatment effects using instrumental variables. Journal of the Royal Statistical Society: Series B, in press.

[33] White, K., Mondesir, F. L., Bates, L. M., & Glymour, M. M. (2014). Diabetes risk, diagnosis, and control: do psychosocial factors predict hemoglobin A1c defined outcomes or accuracy of self-reports?. Ethnicity & disease, 24(1), 19-27.

[34] Windmeijer, F., Farbmacher, H., Davies, N. and Smith, D., (2016). On the Use of the Lasso for Instrumental Variables Estimation with Some Invalid Instruments (No. 16/674). Department of Economics, University of Bristol, UK.

[35] Wu Q, Tchetgen Tchetgen EJ, Osypuk TL, White K, Mujahid M, Glymour MM. Combining direct and proxy assessments to reduce attrition bias in a longitudinal study. Alzheimer Dis. Assoc. Disord. 2012;27:207–212.

[36] Yavorska, O. O. and Burgess, S. (2017). Mendelian randomization: an r package for performing mendelian randomization analyses using summarized data. International Journal of Epidemiology, page dyx034.

Appendix

A1. Proof of Lemma 5

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

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

because

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

where we used the fact that

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

see for example Tchetgen Tchetgen et al (2009). It is straightforward to verify that the

$$\theta = -\ln\left(1 - \frac{E\left[\left\{G - E(G|A=0)\right\}\left\{A - E(A|G=0)\right\}Y\exp\left\{-\varphi_g\left(G\right)A\right\}\right]}{E\left[\left\{G - E(G|A=0)\right\}\left\{A - E(A|G=0)\right\}AY\exp\left\{-\varphi_g\left(G\right)A\right\}\right]}\right)$$

Next,

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

Likewise

$$
\begin{aligned}
& E\left[\{G-E(G|A=0)\}\{A-E(A|G=0)\}AY\exp\left\{-\varphi_g\left(G\right)A\right\}\right] \\
& = E\left[\{G-E(G|A=0)\}\{A-E(A|G=0)\}\exp\left(\beta_a\right)E(Y|A=0,G,U)A\exp\left\{-\varphi_g\left(G\right)A\right\}\right]
\end{aligned}
$$

Therefore

$$
\begin{aligned}
& \frac{E\left[\{G-E(G|A=0)\}\{A-E(A|G=0)\}Y\exp\left\{-\varphi_g\left(G\right)A\right\}\right]}{E\left[\{G-E(G|A=0)\}\{A-E(A|G=0)\}AY\exp\left\{-\varphi_g\left(G\right)A\right\}\right]} \\
& = \frac{\left(\exp\left(\beta_a\right)-1\right)E\left[\{G-E(G|A=0)\}\{A-E(A|G=0)\}AE\left(Y|A=0,G,U\right)\exp\left\{-\varphi_g\left(G\right)A\right\}\right]}{\exp\left(\beta_a\right)E\left[\{G-E(G|A=0)\}\{A-E(A|G=0)\}E(Y|A=0,G,U)A\exp\left\{-\varphi_g\left(G\right)A\right\}\right]} \\
& = \frac{\left(\exp\left(\beta_a\right)-1\right)}{\exp\left(\beta_a\right)}
\end{aligned}
$$

$$\begin{aligned}\theta &= -\ln\left(1 - \frac{(\exp(\beta_a) - 1)}{\exp(\beta_a)}\right)\\ &= -\ln\exp(-\beta_a)\\ &= \beta_a\end{aligned}$$

provided that

$$\begin{aligned}&E\left[\{G - E(G|A=0)\}\{A - E(A|G=0)\}E(Y|A=0,G,U)A\exp\{-\varphi_g(G)A\}\right]\\ &= E\left[\{G - E(G|A=0)\}\{A - E(A|G=0)\}\beta_g(G)A\exp\{-\varphi_g(G)A\}\right]\\ &\neq 0\end{aligned}$$

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

A2. Variance estimation with single IV

The estimating equation in (6) 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

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

The estimation procedure satisfies the joint conditions $m_\theta\left(\hat{\theta}\right) = 0$. 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

$$\sqrt{n}\left(\hat{\theta} - \theta_0\right) = -\left[\left.\frac{\partial m_\theta(\theta)}{\partial\theta}\right|_{\theta^*}\right]^{-}\sqrt{n}m_\theta(\theta_0), \tag{18}$$

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

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

while for the "bread" matrix

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

$$\mathbb{P}_n\begin{bmatrix} -1 & 0_{1\times 2} & 0 \\ 0_{2\times 1} & -\left\{\left.(1,G)'\frac{\partial}{\partial\psi}E(A|G;\psi)\right|_{\psi^*}\right\} & 0_{2\times 1} \\ & \left\{\left.\frac{\partial\widehat{U}}{\partial\mu}\right|_{\mu^*},\left.\frac{\partial\widehat{U}}{\partial\psi}\right|_{\psi^*},\left.\frac{\partial\widehat{U}}{\partial\beta_a}\right|_{\beta_a^*}\right\} & \end{bmatrix},$$

where

$$\frac{\partial}{\partial\psi}E(A|G;\psi) = \begin{cases}(1,G), & \text{for continuous } A \\ \frac{\exp{(1,G)\psi}}{1+\exp{(1,G)\psi}}\left(1-\frac{\exp{(1,G)\psi}}{1+\exp{(1,G')\psi}}\right)(1,G), & \text{for binary A (logit model)},\end{cases}$$

and

$$\begin{aligned}\frac{\partial\widehat{U}}{\partial\mu} &= -(A-E(A|G;\psi))(Y-\beta_aA) \\ \frac{\partial\widehat{U}}{\partial\psi} &= -(G-\mu)(Y-\beta_aA)\frac{\partial}{\partial\psi}E(A|G;\psi) \\ \frac{\partial\widehat{U}}{\partial\beta_a} &= -(G-\mu)(A-E(A|G;\psi))A.\end{aligned}$$

Assume that the matrix $B(\theta_0)$ is non-singular, 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

$$\begin{aligned}&\sqrt{n}\left(\hat{\theta}-\theta_0\right) \xrightarrow{d} \\ &N\left(0, B(\theta_0)^{-}E\left[\tilde{m}(\theta_0)\tilde{m}(\theta_0)'\right]B(\theta_0)^{-\prime}\right). \qquad (19)\end{aligned}$$

Replacing the expected values in (19) 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 $\hat{\theta}$.

A3. Variance estimation with multiple IVs

Let $\widehat{\beta}_a$ be the solution to (12) with optimal weight $\widehat{W}_{opt} = \mathbb{P}_n\left[\widehat{U}\left(\beta_a\right)\widehat{U}\left(\beta_a\right)'\right]^{-}$ where T^{-} denotes the generalized inverse of matrix T. The empirical moment conditions $\widehat{U}\left(\beta_a\right)$ in (12) 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

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

be the k and $(k+1)$ empirical moment conditions of obtaining $\left(\hat{\mu}, \hat{\psi}\right)$ respectively. For iterated or continuously updated GMM procedures in which β_a is estimated simultaneously with the optimal weight, the first order condition of (12) is

$$m_{\beta_a}(\beta_a) = \left\{\mathbb{P}_n\left[\frac{\partial\widehat{U}\left(\beta_a\right)}{\partial\beta_a}\right]\right\}'\widehat{W}_{opt}(\beta_a)\mathbb{P}_n\left[\widehat{U}\left(\beta_a\right)\right] + o_p\left(n^{-1/2}\right).$$

The two-stage procedure solution satisfies the joint moment conditions

$$m_\theta\left(\hat{\theta}\right) = \left(m_\mu\left(\hat{\mu}\right), m_\psi\left(\hat{\psi}\right), m_{\beta_a}\left(\hat{\beta}_a\right)\right)' = 0, \quad \hat{\theta} = \left(\hat{\mu}, \hat{\psi}, \hat{\beta}_a\right). \tag{20}$$

Assume standard regularity conditions and expand $\hat{\theta}$ around the true parameter value θ_0 yields

$$\sqrt{n}\left(\hat{\theta} - \theta_0\right) = -\left[\left.\frac{\partial m_\theta\left(\theta\right)}{\partial \theta}\right|_{\theta^*}\right]^{-} \sqrt{n} m_\theta\left(\theta_0\right), \tag{21}$$

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

$$\sqrt{n} m_\theta\left(\theta_0\right) =$$

$$\begin{bmatrix} I_{(2k+1)\times(2k+1)} & 0_{(2k+1)\times k} \\ 0_{1\times(2k+1)} & \left\{\mathbb{P}_n\left[\left.\frac{\partial \widehat{U}(\beta_a)}{\partial \beta_a}\right|_{\beta_{a0}}\right]\right\}' \widehat{W}_{opt}(\beta_{a0}) \end{bmatrix} \sqrt{n}\mathbb{P}_n \begin{bmatrix} G - \mu_0 \\ (1, G')'\left[A - E(A|G;\psi_0)\right] \\ U(\beta_{a0}) \end{bmatrix} + o_p(1).$$

Let

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

so that

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

Then

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

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

and by Slutsky's theorem

$$\begin{aligned}\sqrt{n} m_\theta\left(\theta_0\right) &\xrightarrow{d} \begin{bmatrix} I_{(2k+1)\times(2k+1)} & 0_{(2k+1)\times k} \\ 0_{1\times(2k+1)} & \Lambda'\Omega^- \end{bmatrix} N(0, E\left[\tilde{m}(\theta_0)\tilde{m}(\theta_0)'\right]) \\ &= M(\theta_0)N(0, E\left[\tilde{m}(\theta_0)\tilde{m}(\theta_0)'\right]).\end{aligned}$$

Next consider the "bread" matrix

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

$$\left[\begin{array}{ccc} -I_{k\times k} & 0_{k\times(k+1)} & 0_{k\times 1} \\ 0_{(k+1)\times k} & -\mathbb{P}_n\left\{(1, G')'\frac{\partial}{\partial\psi}E(A|G;\psi)\Big|_{\psi^*}\right\} & 0_{(k+1)\times 1} \\ \multicolumn{3}{c}{\left\{\mathbb{P}_n\left[\left.\frac{\partial\widehat{U}(\beta_a)}{\partial\beta_a}\right|_{\beta_a^*}\right]\right\}'\widehat{W}_{opt}(\beta_a^*)\mathbb{P}_n\left\{\left.\frac{\partial\widehat{U}}{\partial\mu}\right|_{\mu^*}, \left.\frac{\partial\widehat{U}}{\partial\psi}\right|_{\psi^*}, \left.\frac{\partial\widehat{U}}{\partial\beta_a}\right|_{\beta_a^*}\right\} + o_p(1)} \end{array}\right],$$

where

$$\frac{\partial}{\partial\psi}E(A|G;\psi) = \begin{cases} (1, G'), & \text{for continuous } A \\ \frac{\exp(1,G')\psi}{1+\exp(1,G')\psi}\left(1 - \frac{\exp(1,G')\psi}{1+\exp(1,G')\psi}\right)(1, G'), & \text{for binary A (logit model),} \end{cases}$$

and

$$\begin{aligned}\frac{\partial\widehat{U}}{\partial\mu} &= -I_{k\times k}(A - E(A|G;\psi))(Y - \beta_a A) \\ \frac{\partial\widehat{U}}{\partial\psi} &= -(G-\mu)(Y - \beta_a A)\frac{\partial}{\partial\psi}E(A|G;\psi) \\ \frac{\partial\widehat{U}}{\partial\beta_a} &= -(G-\mu)(A - E(A|G;\psi))A.\end{aligned}$$

Assume that the matrix $B(\theta_0)$ is non-singular, 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

$$\begin{aligned}&\sqrt{n}\left(\hat{\theta}-\theta_0\right) \xrightarrow{d} \\ &N\left(0, B(\theta_0)^{-} M(\theta_0) E\left[\tilde{m}(\theta_0)\tilde{m}(\theta_0)'\right] M(\theta_0)' B(\theta_0)^{-\prime}\right)\end{aligned} \tag{22}$$

In practice, replacing the expected values in (22) 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 $E\left[\tilde{m}(\theta_0)\tilde{m}(\theta_0)'\right]$ 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 (10)