

Nature as a Trialist?

Deconstructing the Analogy Between Mendelian Randomization and Randomized Trials

Sonja A. Swanson,^{a,b} Henning Tiemeier,^{a,c,d} M. Arfan Ikram,^a and Miguel A. Hernán^{b,e,f}

Abstract: Mendelian randomization (MR) studies are often described as naturally occurring randomized trials in which genetic factors are randomly assigned by nature. Conceptualizing MR studies as randomized trials has profound implications for their design, conduct, reporting, and interpretation. For example, analytic practices that are discouraged in randomized trials should also be discouraged in MR studies. Here, we deconstruct the oft-made analogy between MR and randomized trials. We describe four key threats to the analogy between MR studies and randomized trials: (1) exchangeability is not guaranteed; (2) time zero (and therefore the time for setting eligibility criteria) is unclear; (3) the treatment assignment is often measured with error; and (4) adherence is poorly defined. By precisely defining the causal effects being estimated, we underscore that MR estimates are often vaguely analogous to per-protocol effects in randomized trials, and that current MR methods for estimating analogues of per-protocol effects could be biased in practice. We conclude that the analogy between randomized trials and MR studies provides further perspective on both the strengths and the limitations of MR studies as currently implemented, as well as future directions for MR methodology development and application. In particular, the analogy highlights potential future directions for some MR studies to produce more interpretable and informative numerical estimates.

(Epidemiology 2017;28: 653–659)

Mendelian randomization (MR) studies are often described as naturally occurring randomized trials in

which genetic factors are randomly assigned by nature.^{1,2} This analogy between MR and randomized trials is used to defend the superiority of MR analyses over other observational study analyses that require adjustment for measured confounders, and to position MR studies as the preferred source of evidence when trial data are not available. For example, recent clinical guideline reports purport MR studies are the best hope for estimating the benefits of screening or treating children for high cholesterol levels, implying MR studies are a welcome substitute for a decades-long trial that would be required to assess the impact on cardiovascular disease incidence.^{3,4}

When considered in depth, the analogy with randomized trials has profound implications for the design, conduct, reporting, and interpretation of MR studies. Best analytic practices in randomized trials should also be encouraged in MR studies, and methodologic advancements for one study design can potentially be adapted to advance the other design. Likewise, analytic practices that are discouraged in randomized trials should also be discouraged in MR studies.

Here, we deconstruct the oft-made analogy between MR studies and randomized trials. We focus on the consequences of the analogy for the design, conduct, reporting, and interpretation of MR studies. In particular, we explore the implications for clinical and public health decision-making of the meaning of effect estimates in MR studies. In so doing, we highlight the benefits of specifying the protocol of the hypothetical randomized trial that the MR study is replacing. Throughout, we use the term “treatment” as synonymous with “exposure.” We begin by defining the average causal effect of treatment assignment.

THE EFFECT OF TREATMENT ASSIGNMENT IN A RANDOMIZED TRIAL

The causal effect of treatment assignment in randomized trials is known as the intention-to-treat effect. Because assignment is randomized, eligible individuals assigned to each treatment group at “time zero” (i.e., baseline) are expected to be exchangeable (or comparable) on average. Thus, the intention-to-treat effect can be consistently estimated by an unadjusted intention-to-treat analysis that compares the observed outcome distributions across the trial arms. In the presence of differential loss to follow-up, the intention-to-treat analysis

Submitted 25 September 2016; accepted 25 May 2017.

From the ^aDepartment of Epidemiology, Erasmus MC, Rotterdam, The Netherlands; ^bDepartment of Epidemiology, Harvard T. H. Chan School of Public Health, Boston, Massachusetts; ^cDepartment of Child and Adult Psychiatry, Erasmus MC, Rotterdam, The Netherlands; ^dDepartment of Psychiatry, Erasmus MC, Rotterdam, The Netherlands; ^eDepartment of Biostatistics, Harvard T. H. Chan School of Public Health, Boston, Massachusetts; and ^fHarvard-MIT Division of Health Sciences and Technology, Harvard University and Massachusetts Institute of Technology, Boston, Massachusetts. Supported by the National Institutes of Health (R01 AI102634) and by a DynaHEALTH Grant (European Union H2020-PHC-2014; 633595). S.A.S. is supported by a NWO/ZonMW Veni Grant (91617066).

The authors report no conflicts of interest.

Correspondence: Sonja A. Swanson, Department of Epidemiology, Erasmus MC, P.O. Box 2040, 3000 CA Rotterdam, The Netherlands. E-mail: s.swanson@erasmusmc.nl.

Copyright © 2017 Wolters Kluwer Health, Inc. All rights reserved.

ISSN: 1044-3983/17/2805-0653

DOI: 10.1097/EDE.0000000000000699

often requires adjustment for prognostic factors that predict loss to follow-up.⁵

THE ANALOGUE OF THE EFFECT OF TREATMENT ASSIGNMENT IN MR STUDIES

When investigators say that MR studies are “natural randomized experiments,” they are proposing that genetic factors are analogous to the randomized treatment assignment in a trial. Therefore, the effect of the genetic variant in a MR study is analogous to the intention-to-treat effect (the effect of treatment assignment) in a randomized trial. Under this analogy, the effect of the genetic variant can be estimated via an “intention-to-treat” analysis that compares the observed outcome distributions across the groups with different genetic variants. However, there are four key threats to this analogy, and therefore to any analogous estimation of the effect of treatment assignment.

First, groups with different values of the genetic variant may not be exchangeable (i.e., comparable) in MR studies, for example because of population stratification and linkage disequilibrium.^{1,2,6,7} Noncomparability can also occur if there is kinship in the study population that is unknown to the study investigators, a problem known as “cryptic relatedness.”⁸ Such themes have been widely discussed in the MR literature. As is fairly standard practice in the current MR literature, measuring and adjusting for measures of population stratification (e.g., ethnicity) or other sources of noncomparability mitigates this concern,⁹ but, as in an observational study with potential confounding, the concern is never eliminated.

Second, “time zero” of follow-up may not be well defined in a MR study. In randomized trials, time zero is the time at which concomitantly (i) treatment is assigned, (ii) eligibility criteria are met, and (iii) the outcome events of interest start to be counted.¹⁰ In MR studies, there is no clear time zero at which (i), (ii), and (iii) co-occur. Treatment assignment (i) might be (arguably) conceptualized as the time of assignment to the genetic variant, that is, conception, but many MR studies apply postconception eligibility criteria (ii) and then begin recording outcomes (iii) later in life. Of course, applying eligibility criteria subsequent to treatment assignment would make little sense in a true randomized trial. Some examples of MR studies applying “postconception eligibility criteria” include: studies of disease progression restricted to persons with the disease¹¹; studies restricted to pregnant women^{12,13}; and studies using metabolism- or addiction-related genetic variants restricted to persons who report alcohol, tobacco, or other substance use at or above a minimum frequency.^{13,14}

Thus, MR studies are the analogue of randomized trials with a lag of several years or decades between the time of randomization, the time of eligibility, and the time during which the outcomes start to be recorded (Figure). Therefore, the effect estimate may be biased if the genetic variant is associated with the probability of being alive (e.g., either through the treatment of interest¹⁵ or another mechanism) and

eligible at the time events start to be recorded,¹⁰ just like the intention-to-treat effect estimate in a trial may be biased if similar restrictions were made to the analytic sample based on postrandomization events. Such bias is more likely to be consequential in MR studies for which the genetic variant is strongly associated with the eligibility criteria or studies in the elderly or involving genetic variants with strong effects on longevity.¹⁶ For example, estimates of the effect of variants in the FTO gene (a gene strongly related to obesity and perhaps also survival) are expected to be more biased when studying outcomes in the elderly than in young adults.

Third, the genetic variant is not always causal. When the data for MR studies come from genome-wide association studies, the measured genetic polymorphisms are usually a proxy to the causal variant. Thus the genetic variant in the MR study is not an analogue of the treatment assignment in a randomized trial, but rather the analogue of a misclassified version of treatment assignment. This misclassification does not occur in randomized trials, which have accurate recordings of subjects’ treatment assignments. Thus an “intention-to-treat” analysis does not estimate the causal effect of the proxy genetic variant, which is confounded by the causal genetic variant itself. If the causal genetic variant were known, information on the linkage between the measured noncausal variant and the underlying causal variant might be used to minimize this concern. In general, using a proxy of the causal variant affects the identifiability and interpretation of some types of treatment effects beyond the intention-to-treat analogue.^{17,18}

Fourth, the definition of “adherence” is unclear in MR studies. It is therefore difficult to assess how far the “intention-to-treat” effect is from the effect of the intended treatment strategies. In a randomized trial, the protocol specifies the treatment strategies under study, and nonadherence is defined as a deviation from the protocol. In the presence of low adherence, the intention-to-treat effect estimate will be interpreted cautiously. In MR studies, however, the treatment strategies are not explicitly described, which prevents us from unambiguously determining the degree of adherence, and therefore compromises the interpretation of the intention-to-treat effect.

For example, several MR studies used variants in the FTO gene to study the effect of obesity on depression (more precise effect definitions are considered below).^{19–21} One such study reported only 22% of persons with the highest-risk allele (“randomized to treatment”) and 15% of persons with the lowest-risk allele (“randomized to control”) were overweight or obese at some point during the study’s 19-year period.^{19,20} If we define the treatment strategies of interest as “becoming overweight or obese” versus “not becoming overweight or obese,” then this MR study had substantial “nonadherence” because the genetic variant is only weakly associated with the treatment level. On the other hand, if we define the treatment strategies as “increasing body mass index (BMI) by at least two units” versus “decreasing BMI or increasing BMI by two units or less” over the study period, then the MR study might

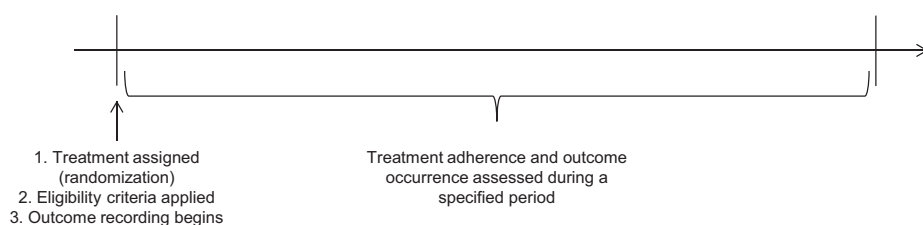
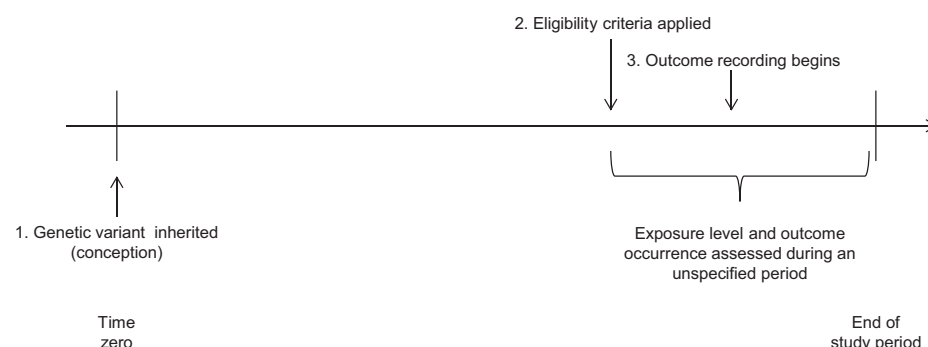
A Randomized Trial**B Mendelian Randomization Study**

FIGURE. Timing of treatment assignment, application of eligibility criteria, recording of treatment adherence, and recording of outcome occurrences in (A) a randomized trial and (B) a hypothetical Mendelian randomization study. The time between beginning and end of study period in a randomized trial is often weeks or possibly years; the time scale in a Mendelian randomization study is often years or decades.

have high adherence if each FTO allele does increase BMI by more than two units in most people, or low adherence if FTO only increases BMI by more than two units in a minority of people. That adherence is poorly defined in MR studies does not affect our ability to estimate the effect of the genetic variant (i.e., the intention-to-treat analog), but it limits its utility and interpretability.

These four threats to the analogy between MR studies and randomized trials (nonguaranteed exchangeability; unclear time zero; proxy randomizers; and undefined adherence) question the validity or interpretability of the effect of genetic variants in “intention-to-treat” analyses of MR studies. MR studies, however, are often not focused on the effect of the genetic variants themselves because the goal is not generally to inform decision-making that involves direct genetic manipulation.²² Instead, MR studies use genetic variation to study other treatment effects that are the analogue of per-protocol effects in randomized trials. We next turn to defining these treatment effects, for which the implications of these four threats to the analogy are compounded and further threats emerge.

THE PER-PROTOCOL EFFECT OF TREATMENT STRATEGIES IN A RANDOMIZED TRIAL

When there is low adherence to the assigned treatment, patients and clinicians may be interested in the effect that would have been observed if all participants had adhered to their assigned treatment strategy as specified in the protocol, that is, the per-protocol effect.²³ Protocols may mandate two types of treatment strategies: strategies that require a single intervention at time zero (e.g., a one-dose vaccine, surgery,

colonoscopy), and strategies that are sustained during the follow-up of the trial (e.g., a daily medication taken throughout follow-up). Below we consider these two cases separately.

The per-protocol effect of baseline interventions can sometimes be validly estimated via instrumental variable methods. An instrumental variable is a variable that (i) is associated with the treatment, (ii) causes the outcome at most through treatment, and (iii) shares no causes with the outcome. Treatment assignment in a randomized trial meets condition (i) because treatment assignment affects actual treatment, and is expected to meet condition (iii) by randomization. Condition (ii), known as the “exclusion restriction,” is expected to hold in double-blind placebo-controlled trials, but not necessarily when the trial is not blinded or a placebo effect is possible.²⁴ Condition (ii) may also be unreasonable in trials comparing two active treatments, because treatment assignment can affect noncompliance (e.g., taking neither treatment) which in turn may affect the outcome.²⁵

The use of an instrumental variable to estimate the per-protocol effect of a baseline intervention also requires a strong “effect homogeneity” condition,^{18,26} which cannot be guaranteed by design. In practice, such homogeneity conditions are often considered biologically implausible, as they essentially require that unmeasured prognostic factors cannot modify the treatment effect.^{18,24,27,28} Therefore, many investigators replace that homogeneity condition by a monotonicity condition, which allows them to estimate the per-protocol effect for only a subgroup of the trial participants rather than the “global” per-protocol effect in the entire study population. This “local” average treatment effect is the per-protocol effect in the unknown subgroup of “complier” participants who

would have complied with the assigned treatment level had they been assigned to either randomization arm.²⁹ For dichotomous instruments and treatments, monotonicity means that there are no “defiers,” that is, no patients who would have taken the opposite treatment level of treatment assignment in either randomization arm, which is a reasonable assumption in many trials.

An alternative that requires neither homogeneity nor monotonicity is to compute bounds (rather than point estimates) for the per-protocol effect in the entire study population. Such bounds are often so wide in practice that they cannot inform whether treatment is helpful, harmful, or has no effect.^{26,30} A more thorough discussion of when these assumptions are expected to hold in a randomized trial with a baseline intervention is provided elsewhere.²⁴

Many trials, however, assign patients to sustained strategies, for example, rather than instructing patients to take or not take a single treatment once, we instruct them to take, or refrain from taking, a medication repeatedly over a period, and include instructions for when to change or stop that treatment. This means that adherence, too, can vary over time. In these trials, the per-protocol effect is the effect that would have been observed had all patients adhered to their assigned treatment strategy for the duration of the study. (Note there is no analogue of the local average treatment effect for sustained strategies.)¹⁸

Generally, classical instrumental variable methods cannot be used (and are not used) to compare sustained treatment strategies, for reasons we explain further below. Extensions of instrumental variable methods based on g-estimation of structural nested models^{26,31} can be used, but they have rarely applied. Such generalizations of instrumental variable methods require detailed and unverifiable assumptions about the effect of treatment, and can only be employed when there is comprehensive data collection on (time-varying) adherence for each individual during follow-up.

THE ANALOGUE OF THE PER-PROTOCOL EFFECT OF TREATMENT STRATEGIES IN MR STUDIES

MR studies frequently include a classical instrumental variable analysis to estimate the effect of the treatment of interest using the genetic factor as the instrumental variable.⁹ Much has been written about whether the conditions required for a valid instrumental variable analysis are reasonable for MR studies, with much focus on the plausibility of the “exclusion restriction” condition (ii).^{2,6,7,9,32} For example, a genetic variant with pleiotropic effects may affect the outcome through pathways other than the treatment of interest, and thus violate this condition. The “exclusion restriction” in MR studies should be explicitly justified on a case-by-case basis—just like in trials without double-blinding. Likewise, any other evoked assumptions using classical instrumental variable approaches (e.g., monotonicity; homogeneity) should

also be justified on a case-by-case basis. This requirement is even more important for MR studies with noncausal genetic instruments, which result in hard to interpret instrumental variable estimates.¹⁷

The use of classical instrumental variable estimation suggests that the goal of MR studies is the comparison of baseline interventions. However, most MR studies estimate the effects of biomarkers (e.g., C-reactive protein, blood pressure, obesity) or behaviors (e.g., diet, substance use) sustained over the entire life-course. While the treatment strategies of interest are not usually rendered explicit in MR studies (because there is no explicit protocol), the interest in sustained treatment strategies is implied by the common phrasing that MR studies are of “lifetime effects of genetically conferred” treatment levels. That is, the appropriate analogy between MR and trials is almost certainly with trials comparing sustained treatment strategies and therefore potentially time-varying treatments.

We now see an important incongruence: the target effect in MR studies is analogous to the per-protocol effect in a randomized trial with sustained treatment strategies, while the common analysis of MR studies is analogous to an instrumental variable estimation in a randomized trial with baseline interventions. This incongruence between the research question and the data analysis is rarely acknowledged in the interpretation of results from MR studies.

To see why this incongruence is a problem, consider a randomized trial of aspirin treatment to estimate the effect of taking aspirin continuously unless contraindications arise (vs. taking no aspirin) on risk of stroke over a 5-year period. Because adherence is expected to change over the follow-up, an instrumental variable analysis cannot generally be used to estimate the per-protocol effect in this trial without bias. If the investigators decided to choose a single measurement of adherence at a random time during follow-up (e.g., at year 2) and apply an instrumental variable analysis based on that single measurement, the resulting estimate would be uninterpretable because adherence in year 2 may be a poor proxy measure for overall adherence (e.g., patients who adhered in year 2 may stop adhering in year 3). In fact, the “exclusion restriction” assumption would be violated because aspirin use at any time may affect future stroke risk. This means that the instrumental variable analysis based on that single measurement would also be inappropriate if the goal was to measure the effect of adherence at year 2 only. Similarly, an instrumental variable analysis in an MR study with a single measurement of the treatment during the follow-up will generally not produce interpretable effect estimates. Any sound analytic strategy for estimating an analogue of the per-protocol effect requires either measuring changes in adherence throughout follow-up, or that the treatment values are constant throughout time.

There is a related challenge to the analogy between MR studies and randomized trials. The per-protocol effect in randomized trials is the effect that would have been observed if all participants had adhered to their assigned treatment strategy

as specified in the protocol. Trying to identify an analogue of a “per-protocol” effect in MR studies is intrinsically related to defining an analogue of the protocol, that is, to providing a clear definition of the treatment strategies of interest. As discussed above, in MR studies that propose variants in the FTO gene as the “treatment assignment” for obesity, the obesity-related “protocol” is undefined. Those who inherit the high-risk variant may be assigned at conception to a treatment strategy of continuous obesity throughout the life-course, to a treatment strategy of progressive weight gain after a childhood of normal weight, etc. Each of these treatment strategies will result in different effect estimates and it is unclear which one of them corresponds to the one targeted by the MR study. Because of this, it is all the more difficult to describe the resulting bias from an MR study using a classical instrumental variable analysis to study a sustained treatment strategy: without first addressing the ambiguity in the per-protocol effect of interest, we cannot logically address whether our estimates are valid or approximately valid.

TESTING FOR A NON-NULL AVERAGE EFFECT OF TREATMENT STRATEGIES

Some investigators insist the goal of MR studies (and, by reversing the analogy, perhaps randomized trials) is not to estimate an effect, but to test whether the treatment has a null average causal effect. In fact, the initial description of MR concepts by Katan³³ in 1986 emphasized methods for testing. In both randomized trials and MR studies, an “intention-to-treat” analysis suffices to test this null hypothesis. However, for such a test to be valid, the effect of treatment assignment must be weaker than the effect of treatment. This will occur when the effect of treatment is in the same direction for all individuals (i.e., a “monotonic” treatment effect) and treatment assignment is indeed an instrumental variable. Thus, if the goal is testing, an instrumental variable analysis is not required but an instrumental variable still is. Of note, testing whether a treatment effect is non-null requires generally fewer assumptions (and less data—we do not need to measure treatment levels) than estimating that same treatment effect, but both require specifying the treatment effect of interest for meaningful discussions of validity.³⁴

All we need for testing purposes is the intention-to-treat effect, yet even for testing we must consider the four above-mentioned threats to the trial analogy in MR. Specifically, the potentially low adherence severely limits the statistical power of any test of the effect of the treatment of interest in MR studies.^{35–37} Moreover, the misclassification in treatment assignment may not affect the validity of a test (if the misclassification is random, potentially conditional on measured covariates), but the statistical power of that test will nonetheless be low.³⁴

EMBRACING THE TRIAL ANALOGY: THE FUTURE OF MR STUDIES?

A pessimistic reader may wonder whether MR studies are ever appropriate. By deconstructing the analogy with

randomized trials, we described multiple problems with effect estimation in MR studies. On the other hand, the trial analogy provides opportunities for improving future MR studies and innovating MR methodology. Here we provide some suggestions.

Embracing the trial analogy in MR studies requires an explicit specification of the protocol of the trial that the MR study attempt to emulate, that is, the target trial.³⁸ The main features of this protocol are eligibility criteria, treatment strategies under study, assignment procedures, follow-up period, outcome, causal contrasts of interest, and analysis plan. These features are described in the protocols of randomized trials, but only partially or vaguely explained in most MR reports. In particular, the treatment strategies and causal contrasts are ambiguously defined and do not always map onto the analysis performed.

Explicit specification of the target trial protocol has two key by-products. First, it can help protect against biases, and help us understand which problems most urgently require solutions in a given study. Second, it helps reposition MR studies to address questions of clinical or public health interest. For example, rather than making imprecise statements about the effect of obesity, future MR studies that specify a target trial protocol (and use appropriate instrumental variable methods for time-varying treatments) could potentially address explicit and important questions concerning when and to what extent various weight-loss interventions over the life-course may be beneficial. Of course, explicit specification of the target trial protocol is also beneficial for non-MR observational studies, because, by decreasing the vagueness in study questions, it facilitates the interpretation of numerical estimates.^{38–40} These themes are perhaps all the more important when considering some of the recent advancements in MR designs and analyses,^{32,41,42} which we discuss briefly in the Appendix.

Regardless of whether the goal is to estimate the effect of treatment assignment, or to test or estimate the effect of a sustained treatment strategy, investigators should consider the four following questions in the context of each MR application. (1) Are likely sources of confounding (i.e., noncomparability across populations with each level of the genetic variant) measured and appropriately adjusted for? If not, consider augmenting with sensitivity analyses, which are especially important for instrumental variable analyses, as small amounts of noncomparability can lead to counterintuitively large biases.⁴³ (2) Is there a unified time zero at which eligibility criteria are applied, treatment is assigned, and outcome recording begins? If no, consider augmenting the analysis with adjustment for measured covariates related to selection (e.g., via inverse-probability weighting), as well as consider using negative controls, simulations, and/or sensitivity analyses to better understand plausible biases.^{10,16} (3) Is the genetic variant known to have a causal effect on the treatment? If no, be careful in interpretation of effect estimates and likewise consider whether the sample size is sufficiently large to mitigate the lowered power

and the possibility of weak instrument biases.¹⁸ And, (4) is there substantial (and well defined) “nonadherence”, that is, is the association between the genetic variant and treatment strategy weak? If yes, consider whether the sample size is sufficiently large to mitigate the lowered power and the possibility of weak instrument bias, and use appropriate analytic techniques for weak instruments.³⁷ Beyond these four queries, investigators should, of course, consider the specific assumptions underlying their analytic goals of testing and/or estimating particular treatment effects. Such considerations can be guided by prior reporting guidelines on instrumental variable methods and for randomized trials.^{27,38,44}

The analogy between MR studies and randomized trials is often used to promote MR studies as the ideal approach for estimating causal effects when randomized trials are neither feasible nor practice. Indeed, MR methods have a lot of promise for advancing our understanding of the effects of treatment strategies when unmeasured confounding appears insurmountable. However, before we can fully answer when and whether an MR analysis is valid for estimating a causal effect, we have to define the effect we are trying to estimate. Here, we have presented some limits of and opportunities for using MR studies. All told, the analogy between MR studies and trials is not a simple strength of the MR design, but rather a complex and helpful framework for understanding and improving the validity of MR research.

REFERENCES

- Nitsch D, Molokhia M, Smeeth L, DeStavola BL, Whittaker JC, Leon DA. Limits to causal inference based on Mendelian randomization: a comparison with randomized controlled trials. *Am J Epidemiol*. 2006;163:397–403.
- Burgess S, Thompson SG. *Mendelian Randomization: Methods for Using Genetic Variants in Causal Estimation*. Boca Raton, FL: Chapman & Hall/CRC Press; 2015.
- Gidding SS, Daniels SR, Kavey RE; Expert Panel on Cardiovascular Health and Risk Reduction in Youth. Developing the 2011 Integrated Pediatric Guidelines for Cardiovascular Risk Reduction. *Pediatrics*. 2012;129:e1311–e1319.
- Urbina EM, de Ferranti SD. Lipid screening in children and adolescents. *JAMA*. 2016;316:589–591.
- Hernán MA, Hernández-Díaz S, Robins JM. Randomized trials analyzed as observational studies. *Ann Intern Med*. 2013;159:560–562.
- Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum Mol Genet*. 2014;23:R89–R98.
- VanderWeele TJ, Tchetgen Tchetgen EJ, Cornelis M, Kraft P. Methodological challenges in Mendelian randomization. *Epidemiology*. 2014;25:427–435.
- Voight BF, Pritchard JK. Confounding from cryptic relatedness in case-control association studies. *PLoS Genet*. 2005;1:e32.
- Boef AG, Dekkers OM, le Cessie S. Mendelian randomization studies: a review of the approaches used and the quality of reporting. *Int J Epidemiol*. 2015;44:496–511.
- Hernán MA, Sauer BC, Hernández-Díaz S, Platt R, Shrier I. Specifying a target trial prevents immortal time bias and other self-inflicted injuries in observational analyses. *J Clin Epidemiol*. 2016;79:70–75.
- Simon KC, Eberly S, Gao X, et al.; Parkinson Study Group. Mendelian randomization of serum urate and Parkinson disease progression. *Ann Neurol*. 2014;76:862–868.
- Zhang G, Bacelis J, Lengyel C, et al. Assessing the causal relationship of maternal height on birth size and gestational age at birth: a Mendelian randomization analysis. *PLoS Med*. 2015;12:e1001865.
- Lewis SJ, Araya R, Smith GD, et al. Smoking is associated with, but does not cause, depressed mood in pregnancy—a Mendelian randomization study. *PLoS One*. 2011;6:e21689.
- Lawlor DA, Nordestgaard BG, Benn M, Zuccolo L, Tybjaerg-Hansen A, Davey Smith G. Exploring causal associations between alcohol and coronary heart disease risk factors: findings from a Mendelian randomization study in the Copenhagen General Population Study. *Eur Heart J*. 2013;34:2519–2528.
- Boef AG, le Cessie S, Dekkers OM. Mendelian randomization studies in the elderly. *Epidemiology*. 2015;26:e15–e16.
- Swanson SA, Robins JM, Miller M, Hernán MA. Selecting on treatment: a pervasive form of bias in instrumental variable analyses. *Am J Epidemiol*. 2015;181:191–197.
- Swanson SA, Hernán MA. The challenging interpretation of instrumental variable estimates under monotonicity. *Int J Epidemiol*. 2017.
- Hernán MA, Robins JM. Instruments for causal inference: an epidemiologist’s dream? *Epidemiology*. 2006;17:360–372.
- Kivimäki M, Jokela M, Hamer M, et al. Examining overweight and obesity as risk factors for common mental disorders using fat mass and obesity-associated (FTO) genotype-instrumented analysis: The Whitehall II Study, 1985–2004. *Am J Epidemiol*. 2011;173:421–429.
- Glymour MM, Tchetgen Tchetgen EJ, Robins JM. Credible Mendelian randomization studies: approaches for evaluating the instrumental variable assumptions. *Am J Epidemiol*. 2012;175:332–339.
- Bjørngaard JH, Carslake D, Lund Nilsen TI, et al. Association of body mass index with depression, anxiety and suicide—an instrumental variable analysis of the HUNT Sstudy. *PLoS One*. 2015;10:e0131708.
- Claussnitzer M, Dankel SN, Kim KH, et al. FTO obesity variant circuitry and adipocyte browning in humans. *N Engl J Med*. 2015;373:895–907.
- Hernán MA, Hernández-Díaz S. Beyond the intention-to-treat in comparative effectiveness research. *Clin Trials*. 2012;9:48–55.
- Swanson SA, Holme Ø, Løberg M, et al. Bounding the per-protocol effect in randomized trials: an application to colorectal cancer screening. *Trials*. 2015;16:541.
- Robins JM. Correction for non-compliance in equivalence trials. *Stat Med*. 1998;17:269–302; discussion 387.
- Robins JM. The analysis of randomized and nonrandomized AIDS treatment trials using a new approach to causal inference in longitudinal studies. In: Sechrest L, Freeman H, Mulley A, eds. *Health Service Research Methodology: A Focus on AIDS*. Washington, DC: US Public Health Service; 1989;113–159.
- Swanson SA, Hernán MA. Commentary: how to report instrumental variable analyses (suggestions welcome). *Epidemiology*. 2013;24:370–374.
- Swanson SA, Hernán MA. Think globally, act globally: an epidemiologist’s perspective on instrumental variable estimation. *Stat Sci*. 2014;29:371–374.
- Angrist JD, Imbens GW, Rubin DB. Identification of causal effects using instrumental variables. *J Am Stat Assoc*. 1996;91:444–455.
- Balke A, Pearl J. Bounds on treatment effects for studies with imperfect compliance. *J Am Stat Assoc*. 1997;92:1171–1176.
- Robins JM. Correcting for non-compliance in randomized trials using structural nested mean models. *Comm Stat*. 1994;23:2379–2412.
- Burgess S, Bowden J, Fall T, Ingelsson E, Thompson SG. Sensitivity analyses for robust causal inference from Mendelian randomization analyses with multiple genetic variants. *Epidemiology*. 2017;28:30–42.
- Katan MB. Apolipoprotein E isoforms, serum cholesterol, and cancer. *Lancet*. 1986;1:507–508.
- Didelez V, Sheehan N. Mendelian randomization as an instrumental variable approach to causal inference. *Stat Methods Med Res*. 2007;16:309–330.
- Davies NM, von Hinke Kessler Scholder S, Farbmacher H, Burgess S, Windmeijer F, Smith GD. The many weak instruments problem and Mendelian randomization. *Stat Med*. 2015;34:454–468.
- Burgess S, Thompson SG. Bias in causal estimates from Mendelian randomization studies with weak instruments. *Stat Med*. 2011;30:1312–1323.
- Bound J, Jaeger D, Baker R. Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *J Am Stat Assoc*. 1995;90:443–450.
- Hernán MA, Robins JM. Using big data to emulate a target trial when a randomized trial is not available. *Am J Epidemiol*. 2016;183:758–764.

39. Hernán MA. Does water kill? A call for less casual causal inferences. *Ann Epidemiol.* 2016;26:674–680.
40. Kaufman JS. There is no virtue in vagueness: Comment on: causal identification: a charge of epidemiology in danger of marginalization by Sharon Schwartz, Nicolle M. Gatto, and Ulka B. Campbell. *Ann Epidemiol.* 2016;26:683–684.
41. Burgess S, Timpson NJ, Ebrahim S, Davey Smith G. Mendelian randomization: where are we now and where are we going? *Int J Epidemiol.* 2015;44:379–388.
42. Swanson SA. Commentary: Can we see the forest for the IVs? Mendelian randomization studies with multiple genetic variants. *Epidemiology.* 2017;28:43–46.
43. Jackson JW, Swanson SA. Toward a clearer portrayal of confounding bias in instrumental variable applications. *Epidemiology.* 2015;26:498–504.
44. Baiocchi M, Cheng J, Small DS. Instrumental variable methods for causal inference. *Stat Med.* 2014;33:2297–2340.

APPENDIX

SOME EXTENSIONS TO THE ANALOGY BETWEEN RANDOMIZED TRIALS AND MENDELIAN RANDOMIZATION

MR Studies with a Two-sample Design

MR studies sometimes employ “two-sample” designs such that one study population is used to estimate the association between the genetic variants and treatment and a second study population is used to estimate the association between the genetic variants and outcome.^{32,41} Even

beyond the fundamental question raised in the main text—what treatment effect is this design really targeting?—such a design is peculiar when drawing a trial analogy. Specifically, it is like using the adherence observed in one trial to adjust the intention-to-treat effect estimate in another trial.

MR Studies with Multiple Genetic Variants

The main text considers the analogy between randomized trials and MR studies with one genetic variant, but MR studies increasingly involve multiple genetic variants. This setting is often described as analogous to a meta-analysis of multiple randomized trials that are all happening simultaneously in the same set of individuals.⁴¹ That is, each genetic variant is a treatment assignment in its own randomized trial with its own level of adherence, which partly depends on the other simultaneous trials. Although recent advancements in MR methods can address some important limitations of classical instrumental variable analyses with weak or imperfect instruments,^{2,32,41} all such techniques developed thus far are limited to settings for baseline treatments only.⁴² Therefore, MR studies using data on multiple genetic variants would suffer from the same limitations as inappropriately applying classical instrumental variable analyses in settings with time-varying treatments.