JOURNAL CLUB
18-5-2017

Presenter: Dr Philip Bassey

INSTRUMENTAL VARIABLES II: IV APPLICATION FOR "PPP" AND GENETIC STUDIES

CLINICAL EPIDEMIOLOGY AND BIOSTATISTICS

CΣB

JOURNAL CLUB 2018

PHILIP ETABEE MACDONALD BASSEY

RESOURCE MATERIALS

I. Instrumental variables I: instrumental variables exploit natural variation in non-experimental data to estimate causal relationships

II. Preference based IV methods for the estimation of treatment effects: Assessing Validity & Interpreting Results

III. Instrumental variables II: instrumental variable application—in 25 variations, the physician prescribing preference generally was strong and reduced covariate imbalance

IV. Using multiple genetic variants as instrumental variables for modifiable risk factors

Outline of Presentation
3 Parts Presentation

1. Overview of instrumental variable(s) (IVs) – Natural variations-Covered by the first paper
2. Application of Instrumental variables in health services research –(The PPP Concept) – Covered in the second/third papers
3. Application of Instrumental variables in genetic studies- Covered in the fourth paper

J Clin Epidemiol. 2009 December ; 62(12): 1226–1232. doi:10.1016/j.jclinepi.2008.12.005.

Instrumental variables I: instrumental variables exploit natural variation in nonexperimental data to estimate causal relationships

Jeremy A. Rassen[a,b,c,*], M. Alan Brookhart[a,b], Robert J. Glynn[a,b,d], Murray A. Mittleman[b,c,e], and Sebastian Schneeweiss[a,b,c]

Introduction:

- Observational studies struggle with potential for bias from confounding by indication and other unmeasured risk factors
- The gold standard of study design for treatment evaluation is widely acknowledged to be the randomized controlled trial (RCT).
- The classic experimental method of establishing causality is to intervene in one group while leaving a second control group aside

Introduction:

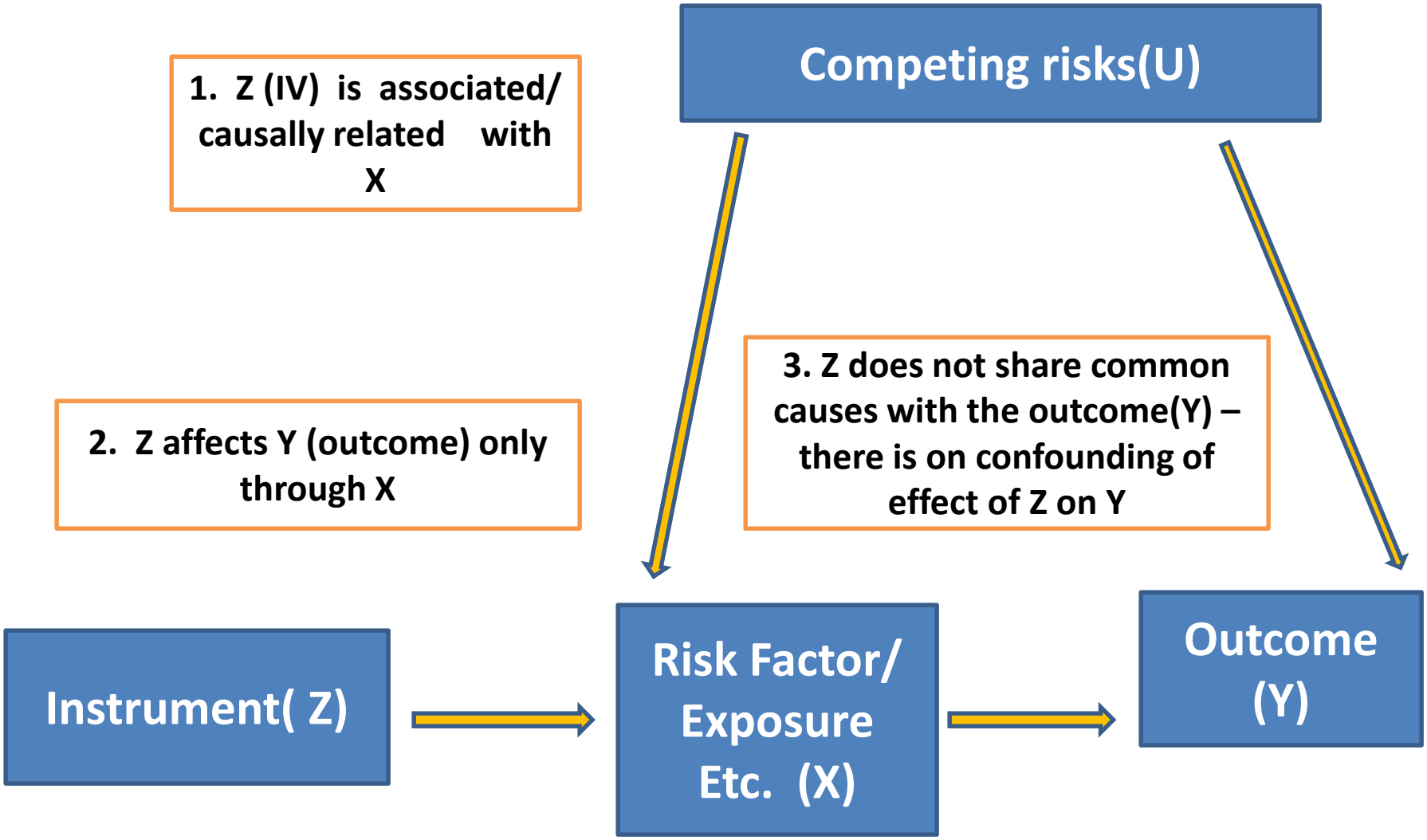
- For decades, economists have been using instrumental variable (IV) analysis as a method of causal inference in cases where an RCT is not possible and when an assumption of no unmeasured confounding is unwarranted.

- This article Instrumental Variable -1 outlines the theoretical framework, analytical method and the assumptions required for IV analysis

What is as instrumental variable (IV)?

- It is an unconfounded proxy for a study exposure that can be used to estimate a causal effect in the presence of unmeasured confounding.
- In the many cases where RCTs are impractical or unethical, instrumental variable (IV) analysis offers a nonexperimental alternative based on many of the same principles of RCTs.
- Instrumental variable (IV) analysis provides a method to obtain a potentially unbiased estimate of treatment effect, even in the presence of strong unmeasured confounding

Criteria for Instrumental Variables (IVs)



Assumptions of Instrumental variable

An IV (instrument) Z is defined as a variable that satisfies the following assumptions:

(1) Z (IV) is associated with X the exposure / risk factor of interest / intermediate variable

(2) Z affects the outcome Y only through X. [No direct effect of G on Y] –Exclusion restriction.

(3) Z is independent of the (unobserved) confounding factors U of the association between X and the outcome Y;

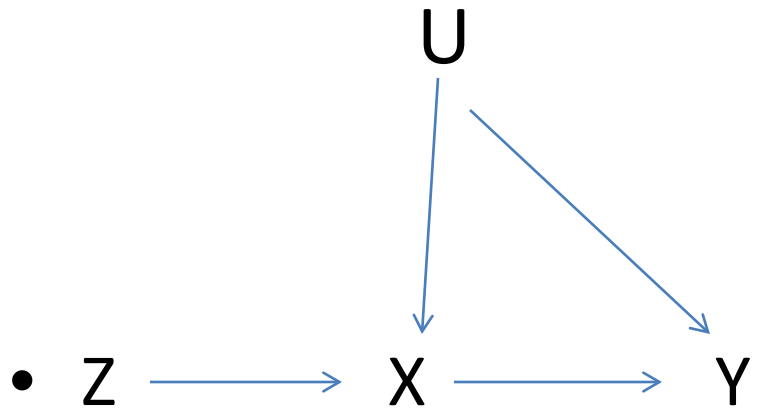


Table 1

Key points about IV analysis

An instrumental variable is a variable in nonexperimental data that can be thought to mimic the coin toss in a randomized trial.

If an appropriate and valid instrument is found, then the effects of measured and unmeasured confounding can be mitigated.

An IV analysis always has an experimental analog, however absurd the experiment sounds. The IV analysis is therefore based on "a natural experiment."

Assumption (1): The IV must predict treatment but that prediction does not have to be perfect. An IV that does a poor job of prediction is said to be weak.

Assumption (2): A valid IV will not be directly related to outcome, except through the effect of the treatment.

Assumption (3): A valid IV will also not be related to outcome through either measured or unmeasured paths.

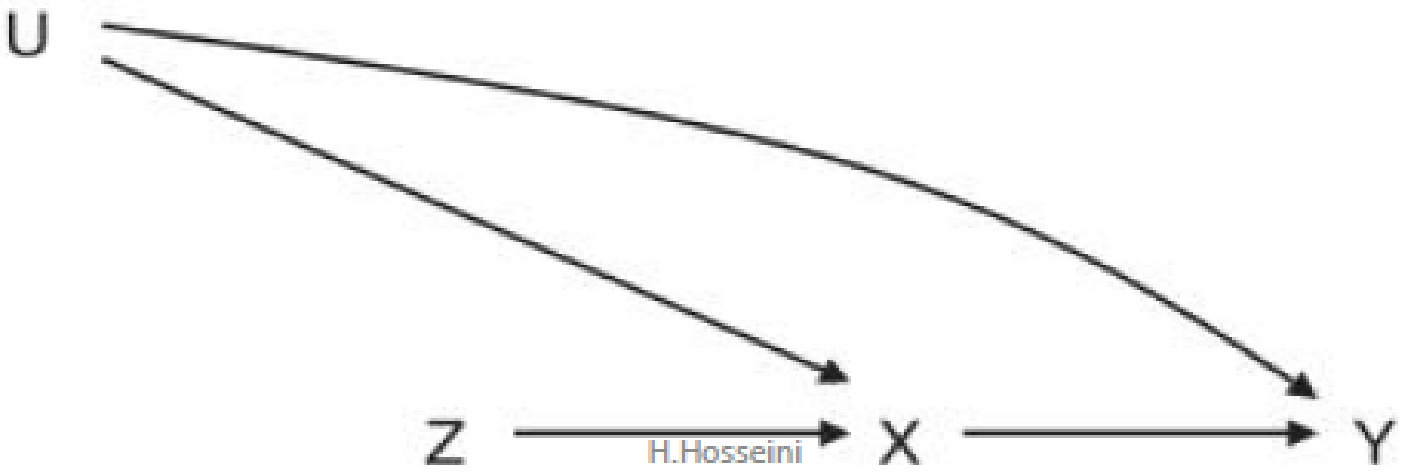
In a randomized trial, the assumptions are met by design in the act of randomization. In an IV analysis, these assumptions must be empirically checked to the extent possible or assumed based on context and subject matter knowledge.

In cases of treatment–effect homogeneity, IV studies estimate the effect on the marginal subject, the average treatment effect for patients whose treatment was determined by the instrument [14].

IVs, or instruments, in randomized experiments

For a typical trial

- Z is the randomization assignment indicator (eg, 1= treatment, 0=placebo),
- X is the actual treatment received (1 = treatment, 0= placebo),
- Y is the outcome, and
- U represents all factors (some unmeasured) that affect both the outcome and the decision to adhere to the assigned treatment.



Theory: Comparison between RCT and IV analysis

RCT

- Three categories of participants: Compliers; Noncompliers, Defiers
- Compliers randomly distributed in each of the arms provide the statistical information that will determine the effect measure of the study
- In RCTs **Blinding** removes the possibility of defiance

IV

- Also Three categories of subjects Compliers; Noncompliers, Defiers
- "Compliers" - marginal subjects whose treatment status is determined by the status of the instrument

(proximity/access to care} provide information about the effect of treatment, as they are the ones whose exposure was directly affected by the instrument.

Theory: Comparison between RCT and IV analysis

RCT

- **Independence and exclusion should be met by design.**
- **In randomized trials, the independence assumption and exclusion restriction are fundamentally unverifiable.**
- **Indeed, many of the problems with RCTs, such as poor randomization leading to treatment group imbalance, are empirical violations of independence or exclusion**
- **The ITT analysis provides an estimate of the treatment effect among the "compliers"**

IV

- **In IV designs independence & exclusion can be met using IV analysis**
- **In IV settings the independence assumption and exclusion restriction are also fundamentally unverifiable**
- **The exclusion restriction can be violated by the existence of common causes of both the instrument and the outcome, and is met only by assumption.**
- **IV analysis provides estimate of the effect of Rx among the marginal subjects (compliers). This estimate is scaled to a figure that reflects the effect of treatment had everyone in the population been marginal.**

Theory: Comparison between RCT and IV analysis

IV Assumptions:	RCT Compliance
1. Z has a causal effect on X	Condition is met in RCT- trial participants are more likely to be Rx if they were assigned to Rx
2. Z affects Y only through X { EXCLUSION RESTRICTION}	This is ensured by effective double blindness
3. Z does not share common causes with the outcome Y	This condition is ensured by the random assignment of Z

Distance to Specialty Provider as IV

McClellan, M., B. McNeil and J. Newhouse, *JAMA*, 1994.

"Does More Intensive Treatment of Acute Myocardial Infarction Reduce Mortality?"

- Medicare claims data identify admissions for AMI, 1987-91
- Treatment: Cardiac catheterization (marker for aggressive care)
- Outcome: Survival to 1 day, 30 days, 90 days, etc.
- Instrument: Indicator of whether the hospital nearest to a patient's residence does catheterizations.

Illustration : The differential difference hypothesis

- The study by McClellan et al
- Study context: An observation that some hospitals provide catheterization, whereas others do not (or do so only infrequently)
- Hypothesized that the patient's differential distance from catheterization-providing hospital may be a determinant of Rx .
- That the paramedic was more likely to go to the nearer hospital rather than select a farther one based on the availability of particular facilities
- Therefore, all things equal, patients living within short differential distances to catheterization-providing hospitals would be more likely to receive catheterization solely as a result of their proximity.

Table 2a

Association between catheterization (X) and death (D) (crude exposure to outcome association: RD = 0.150)

	Catheterization ($X+$)	No catheterization ($X-$)	Total
Died ($D+$)	100	25	125
Did not die ($D-$)	400	475	875
Total	500	500	1,000

Abbreviation: RD, risk difference.

Table 2b

Association between closeness to catheterization facility (Z) and death (D) (instrument to outcome association assuming quasi-randomization: RD = −0.100)

	Small Diff. Dist. to Cath. facility ($Z+$)	Large Diff. Dist. to Cath. facility ($Z-$)	
Died ($D+$)	6	119	125
Did not die ($D-$)	144	731	875
Total	150	850	1,000

Abbreviation: Diff. Dist.: differential distance; Cath., catheterization.

Table 2c

Association between closeness to catheterization facility (Z) and catheterization (X) (strength of instrument and amount of compliance: RD = 0.494)

	Small Diff. Dist. to Cath. facility ($Z+$)	Large Diff. Dist. to Cath. facility ($Z-$)	
Catheterization ($X+$)	138	362	500
No catheterization ($X-$)	12	488	500
Total	150	850	1,000

Analyzing the data: causal effect of the IV on the marginal subject illustrated with the study by McClellan et al.

- Based on the example of distance as a proxy for catheterization, the data from Table 2a (**crude RD = 0.150)** was reanalyzed by using "short differential distance" in place of "received catheterization" and "long differential distance" in place of "didn't receive catheterization" (Table 2b; **RD = –0.100)**
- Then the confounding effect of selection for catheterization and death was "supposedly " removed by the quasi-randomized treatment arising from the natural variation in the place where patients live.
- Therefore the analysis was moved from the treatment-based estimate to the IV-based estimate thereby switching the direction of the effect estimate.
- However the estimate of differential distance on catheterization may be muted because there might be a significant number of nonmarginal patients, patients for whom distance was not the factor that determined their treatment (Table 2c; **RD = 0.494).**

The simple calculation of the IV estimate on the RD scale is as follows:

$$\text{RD} = \frac{\text{Instrument}\rightarrow\text{Outcome association}}{\text{Instrument strength}}$$
$$= \frac{\text{Distance}\rightarrow\text{Death association}}{\text{Distance}\rightarrow\text{Catheterization association}}$$
$$= \frac{-0.100}{0.494} = -0.202$$

The numerator in the fraction is the IV-to-outcome relationship, and will ranges from −1 to 1; in a randomized study, the numerator is simply the ITT estimate.

The denominator is the scaling factor that accounts for compliance. A strong instrument will yield a rescaling factor toward ±1, whereas a weak instrument will be closer to zero.

Importantly, if any of the assumptions have been violated, scaling may magnify any bias from residual unmeasured confounding that is factored into the numerator

This fraction called Wald estimator, is useful for the most basic IV estimates.

INTENTION TO TREAT (ITT) & WALD IV ESTIMATOR

$$\text{ITT Estimator} = E[Y|Z=1] - E[Y|Z=0]$$

$$\text{IV Estimator} = \frac{E[Y|Z=1] - E[Y|Z=0]}{E[X|Z=1] - E[X|Z=0]}$$

WALD IV ESTIMATOR $= \dfrac{\text{Effect of the Instrument on the Outcome}}{\text{Effect of the Instrument on the Exposure}}$

J Clin Epidemiol. 2009 December ; 62(12): 1233–1241. doi:10.1016/j.jclinepi.2008.12.006.

Instrumental variables II: instrumental variable application—in 25 variations, the physician prescribing preference generally was strong and reduced covariate imbalance

Jeremy A. Rassen[a,b,c,*], M. Alan Brookhart[a,b], Robert J. Glynn[a,b,d], Murray A. Mittleman[b,c,e], and Sebastian Schneeweiss[a,b,c]

Over view of the paper:

- Provides a link with the first paper - Instrumental Variable-1

- Explores the alternative definitions of the physician prescribing preferences (PPP) proposed by **Brookhart et al**. and related work by other authors.

- Discusses possible analytic frameworks of IVs

Study Context

- Physician prescribing preference (PPP) has been used as an instrumental variable in clinical epidemiology
- They created 25 different PPP study algorithms from the IV instrument that was proposed by **Brookhart et al.** both in terms of making series of variations in the study design and cohort selection.
- For each variation, they assessed the IV's strength and the reduction in imbalance resulting from the application of the IV.
- They compared reductions in imbalance across the variations and assessed the overall relationship between strength and imbalance.

- **BEFORE PROCEEDING, LET US BRIEFLY EXAMINE THE BROOKHART CONCEPTS OF IVs**

The International Journal of Biostatistics

Volume 3, *Issue* 1 2007 *Article* 14

Preference-Based Instrumental Variable Methods for the Estimation of Treatment Effects: Assessing Validity and Interpreting Results

M. Alan Brookhart, *Division of Pharmacoepidemiology, Brigham and Women's Hospital & Harvard Medical School*
Sebastian Schneeweiss, *Division of Pharmacoepidemiology, Brigham and Women's Hospital & Harvard Medical School*

Overview of the study by Brookhart et al.

- They reviewed the use of Observational studies of prescription medications / medical interventions based on administrative data for clinical decision making.
- They queried the validity of such studies - because the data may not contain measurements of important prognostic variables that guide treatment decisions.
- Variables that are typically unavailable in administrative databases include lab values (e.g., serum cholesterol levels), clinical data (e.g., weight, blood pressure), aspects of lifestyle (e.g., smoking status, eating habits), and measures of cognitive and physical functioning.
- The threat of unmeasured confounding is thought to be particularly high in studies of intended effects because of the strong correlation between treatment choice and disease risk (Walker, 1996).

Study by Brookhart et al.

- The goal of Brookhart et al was to compare the effect of prescribing 2 classes of drugs (cyclooxygenase 2-[COX-2] selective and nonselective nonsteroidal antiinflammatory drugs [NSAIDs]) on gastrointestinal bleeding.

- The authors propose the "physician's prescribing preference" for drug class as the instrument, arguing that it meets conditions (i), (ii), and (iii).

- Because the proposed instrument is unmeasured, the authors replace it in their main analysis by the (measured) surrogate instrument "last prescription issued by the physician before current prescription."

Study by Brookhart et al.

Observational Study of Non-steroidal Anti-Inflammatory Drugs and GI bleeding risk in an elderly population (Brookhart et al, Epidemiology 2006)

- Compare short-term risk of GI outcomes between
  - Non-selective NSAIDs
  - COX-2 selective NSAIDs
- Coxibs are slightly less likely to cause GI problems
- Coxibs are likely to be selectively prescribed to patients at increased GI risk
- Classic problem of confounding by indication

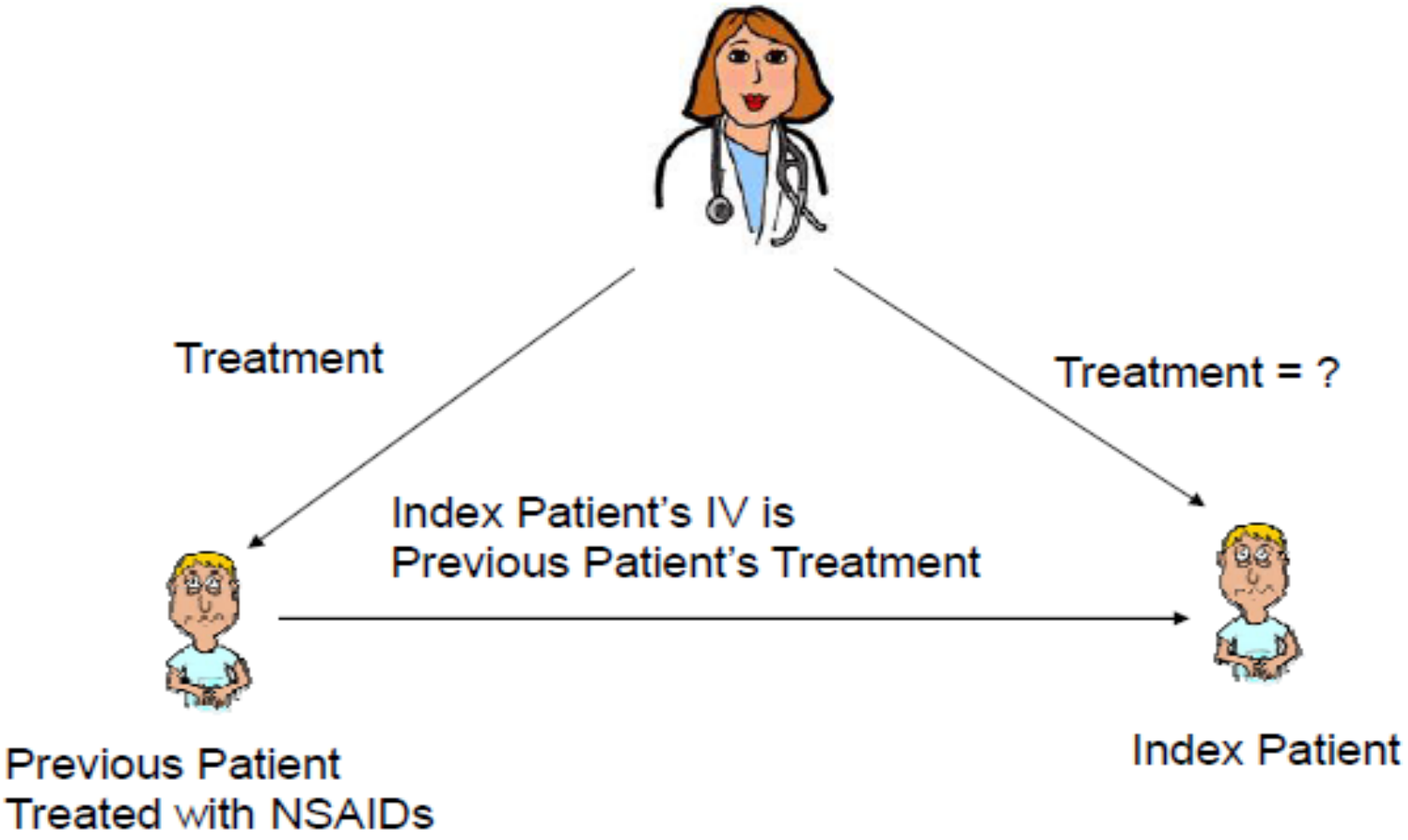
Characteristics of Cohort

Variable	Coxib	NS NSAID
Female Gender	86%	81%
Age > 75	75%	65%
Charlson Score>1	76%	71%
History of Hospitalization	31%	26%
History of Warfarin Use	13%	7%
History of Peptic Ulcer Disease	4%	2%
History of GI Bleeding	2%	1%
Concomitant GI drug use	5%	4%
History GI drug use	27%	20%
History of Rheumatoid Arthritis	5%	3%
History of Osteoarthritis	49%	33%

Patient's GI Risk

	Low	Moderate	High
			
		"Marginal Patient"	
 COX-2 Preferring Physician	NS NSAID	COXIB	COXIB
 NS NSAID Preferring Physician	NS NSAID	NS NSAID	COXIB

Estimating preference:



Instrument should be unrelated to observed patient risk factors

Variable	Coxib Pref Z=1	NS NSAID Pref Z=0
Female Gender	84%	84%
Age > 75	73%	72%
Charlson Score > 1	75%	73%
History of Hospitalization	29%	27%
History of Warfarin Use	12%	10%
History of Peptic Ulcer Disease	3%	3%
History of GI Bleeding	1%	1%
Concomitant GI drug use	5%	5%
History GI drug use (e.g., PPIs)	25%	24%
History of Rheumatoid Arthritis	4%	4%
History of Osteoarthritis	45%	41%

Instrument should be related to treatment

Last NSAID Prescription (IV)	Current Prescription (Actual Treatment)	
	Coxib X=1	**Non-Selective NSAID X=0**
Coxib Z=1	(73%)	(27%)
Non-Selective NSAID Z=0	(50%)	(50%)

With that brief background- We can proceed to Instrumental Variable 11

J Clin Epidemiol. 2009 December ; 62(12): 1233–1241. doi:10.1016/j.jclinepi.2008.12.006.

Instrumental variables II: instrumental variable application—in 25 variations, the physician prescribing preference generally was strong and reduced covariate imbalance

Jeremy A. Rassen[a,b,c,*], M. Alan Brookhart[a,b], Robert J. Glynn[a,b,d], Murray A. Mittleman[b,c,e], and Sebastian Schneeweiss[a,b,c]

RECALL :

- **Brookhart et al.** had proposed that an individual physician's preference for prescribing one drug over another is an IV that predicts which drug a patient will be treated with.
- From the examination of physician prescribing patterns they deduced that the variation they observed may be an instrument under the assumption that PPP is unrelated to outcome.
- The preference at the time of seeing the patient was determined by the treatment a doctor chose for the previous patient who was treated in his or her practice and who also required a new prescription for one of the study drugs

Overview of the Paper-Key Points of Instrumental Variable-11

- The instrumental variable here is the Physician Prescribing Preference (PPP)
- Emphasis on reliable and consistent estimates of effect
- Achieving IV validity by reducing covariate imbalance
- Study was therefore aimed at exploring ways of achieving covariate balance and the improving the strength of the instrument

Objective of the study- To:

- Examine the covariate balance and instrument strength in 25 formulations of the PPP IV in two cohort studies.

- Explore variations in the simple definition of PPP by changing the PPP algorithm through the application of restriction and stratification schemes

- Evaluate each variation based on the IV strength and reduction in imbalance.

Study Design

Application of the PPP IV to assess antipsychotic medication (APM) use and subsequent death within 180 days among two cohorts of elderly patients in two different locations.

Method /Modalities

(i) They varied the measurement of the PPP

(ii) Performed cohort restriction and stratification.

(iii) Modeled risk differences with two-stage least square regression

(iv) Assessed the balance of the covariates using the Mahalanobis distance

Varying the IV Tool

Even though the use of the previous patient's treatment to estimate preference has the advantage of quickly registering any changes in preference, two issues arise:

(i) The previous patient's treatment may not reflect the doctor's true preference

(ii) The simple IV as specified may not possess the required **strength** and **validity**.

1. Varying the measurement of the PPP IV Tool

- Note that Brookhart et al. had proposed the simple technique for measuring a physician's preference which Rassen et al. termed the “base case”.
- The “base case” is considered to be the reference cohort that are on the existent treatment preferences / regimens
- Base cohort : had no restrictions and physician's previous prescription was used as instrument [Reference group]
- In all instances, they chose single, dichotomous IVs for interpretability and comparability.

Steps in varying the study design and physician prescribing preference formulation

Rassen et al. designed variations on the "base case" that were meant to exercise the definition of the PPP measure and to create contrasts in strength and validity by modeling:

(1) preference assignment algorithm
(2) source population
(3) stratification criteria

Method- Variation of -study design Cont'd

- They also expanded the time window to calculate preference from more than just the last new prescription filled.
- They used the previous two, three, and four new prescriptions, and set different targets for prescribing consistency
- E.g. in the case of four prescriptions, they considered that "any of the four," "half of the four," and "all of the four" were conventional rather than atypical APMs.
- They hypothesized that expanding the window would increase balance in treatment groups by creating a better, more stable estimate of true underlying preference and therefore better quasi-randomization of patients to two predicted treatment groups (arms)

Methods-Cont'd

Base Case:

Base cohort with no restrictions and physician's previous prescription as instrument

1. Preference assignment algorithm changes

1A. Lenient criteria

- P1: At least **1** conventional APM Rx within last 2 Rx's
- P2: At least **1** conventional APM Rx within last 3 Rx's
- P3:At least **1** conventional APM Rx within last 4 Rx's

1.B. Strict criteria

- P4: **2** conventional APM Rx's within last **2** Rx's
- P5: **3** conventional APM rx's within last **3** rx's
- P6: 4 conventional APM rx's within last **4** rx's

1.C. Moderate criteria

- P7: At least 2 conventional APM rx's within last 3 rx's
- P8: At least 2 conventional APM rx's within last 4 rx's

2. Cohort restrictions

- ***2.A. Cohort restriction based on doctor characteristics***
- **R1:** Doctor has a very high-volume practice
- R2: Doctor has a high-volume practice
- R3: Doctor has a low-volume practice
- R4: Doctor sees many older patients

Methods-Cont'd

- R5 : Doctor sees many younger patients
- R6 : Doctor is a primary care physician
- R7: Doctor is a specialist
- R8: Doctor graduated before 1980 (PA[b])
- R9: Doctor graduated after 1980 (PA[b]):

2.B. Cohort restriction based on patient characteristics

- R10: Patient above median patient age
- R11:Patient below median patient age
- R12 : Patient in the middle quartiles of age

2.C. Cohort restriction based on patient and doctor characteristics

- R13: Patient is older than the median age in the doctor's practice

Stratification changes

- S1: Last patient was in the same age category
- S2: Last patient was also above/ below the median patient age
- S3: Last patient was also above/ below the median patient age within doctor's practice
- S4: Last patient was in the same quartile of propensity score

Illustrated example-context

- They performed an example study of initiation of APM therapy and the associated risk of short-term mortality.
- APMs are categorized into two groups: conventional (older) and atypical (newer) agents
- They are widely used off-label to control behavioral disturbances in demented elderly patients.
- Previous studies have found increased rates of death among users of atypical antipsychotic agents as compared with placebo
- Nonrandomized studies have indicated that both types of APMs increase risk of death in the elderly, with the atypical drugs showing lesser risk than the conventional ones

Study Population & Setting:

- Two cohorts of patients aged 65 years and older who initiated APM treatment.
- The first cohort was drawn from Pennsylvania (PA)'s Pharmaceutical Assistance Contract for the Elderly (PACE), a drug assistance program for the state's low-income seniors, between 1994 and 2003.
- The second cohort was drawn from all British Columbia (BC) residents aged 65 years or more between 1996 and 2004.
- Patients with existing cancer diagnoses were excluded

Drug exposures, study outcomes, and measured patient characteristics

- They defined the exposed group to be initiators of conventional APM treatment and compared them with a referent group of initiators of atypical APM therapy
- Outcome was defined as death within 180 days from drug initiation.
- The baseline characteristics of the patients was defined based on the 6 months before each subject's index date and included coexisting illnesses and use of health care services
- All dates were measured to the level of day; events occurring on the same day were ordered randomly.

Statistical models:

- Two-stage least squares (2SLS) models were used to estimate risk differences
- All IV models were run in Stata Version 9 using the ivreg2 module
- They applied the robust function to estimate the standard errors to account for clustering within physician practices using the sandwich estimator

How to Estimate the Effect of Treatment Using an IV

- IV analysis is typically done using a 2-stage least-squares estimation.
- In order to assess the strength of an instrument, an F test can be used in first-stage regression which predicts treatment as a function of the instrument and covariates.

- The F test is used to test the hypothesis that $\alpha 1$ is significantly different from 0 .

- An F statistic >10 is often used as a 'rule of thumb' to indicate that an instrument is not weak , but this may not be the case if multiple instruments are available.

- A partial r2 (the square of the correlation between the instrument and the treatment adjusted for other covariates) can also be used to assess the proportion of the variance explained by the addition of the IV to the regression model. A large partial r2 is an indication that the instrument makes a large contribution to the prediction of treatment .

Description of the Two-Stage Least-Squares Regression

Stage 1: regression: $T_i = \alpha_0 + \alpha_1 Z_i + \nu_i$ (1)
where T_i = treatment; Z_i = IV; ν_i = error term for treatment $\alpha_1 \neq 0$; Z_i and T_i can be either continuous or binary (can also adjust for measured confounders).

Stage 2: regression: $Y_i = \beta_0 + \beta_1 \hat{T} + e_i$ (2)
where $\hat{T} = \hat{\alpha}_0 + \hat{\alpha}_1 Z_i$; Y_i = outcome; $\hat{T}_i$ = estimated treatment effect; e_i = error term for outcome (can also adjust for measured confounders).

Substituting equation 1 into equation 2:

$$Y_i = \gamma_0 + \gamma_1 Z_i + \varepsilon_i$$

where

$$\gamma_0 = \beta_0 + \eta_1 \hat{\alpha}_0;\ \gamma_1 = \beta_1 \hat{\alpha}_1;\ \varepsilon_i = \beta_1 \nu_1 + e_i.$$

In order to estimate the *direct treatment effect* (β_1) of treatment (T_i) on outcome (Y_i), we need to use the information from equations 1 and 2:

$$\beta_1 = \hat{\gamma}_1 / \hat{\alpha}_1.$$

- In the first stage, a regression estimate ($\hat{\alpha}1$) is obtained by regressing treatment (T) on the IV (Z) in equation 1 (can also adjust for relevant measured confounders).

- In the second stage, the predicted value of the treatment ($\hat{T}$) is used in a regression of the outcome (Y) on treatment ($\hat{T}$) (can once again adjust for relevant measured confounders) to obtain an estimate $\hat{\Upsilon}1 = \beta 1\hat{} / \alpha 1$ in equation 2.

- This 2-stage approach eliminates the bias that would have occurred in a conventional regression of outcome on actual treatment received using our observational data.

- The estimated direct treatment effect ($\beta 1$) is calculated as the ratio of $\hat{\Upsilon}1 / \hat{\alpha} 1$.

Dichotomous Outcomes and Relative Measures of Effect

- The simple Wald estimator and the linear structural equation models can be used with dichotomous outcomes.
- The linear structural models require the use of appropriate software to conduct inference, correctly specified models, and the predicted values of exposure in the 0-1 range.
- However, in medicine and epidemiology interest often focuses on ratio measures such as relative risks or rates. IV approaches based on the Wald estimator or linear structural equation models yield estimates of an absolute measure of effect (e.g., a risk difference).
- A variety of IV approaches can be used to estimate relative measures of effect, and each imposes somewhat different assumptions.

IV Estimation Using Stata

- let outcome be the *outcome*,
- *exp* be the exposure,
- *iv* be the instrument,
- *age* be a continuous age variable,
- *sex* be an indicator for male sex (1=male, 0=female),
- and *c1, c2, and c3* be three dichotomous confounders.

simple crude and adjusted models

using ordinary least squares (OLS) estimation

```
. reg outcome exp

      Source |       SS       df       MS              Number of obs =   78731
-------------+------------------------------           F(  1, 78729) =  165.66
       Model |  3.17645234     1  3.17645234            Prob > F      =  0.0000
    Residual |  1509.58325 78729  .019174424           R-squared     =  0.0021
-------------+------------------------------           Adj R-squared =  0.0021
       Total |   1512.7597 78730  .019214527           Root MSE      =  .13847
```

Crude RD

outcome	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
exp	**-.0158182**	**.001229**	**-12.87**	**0.000**	**-.018227**	**-.0134094**
_cons	.0227949	.0005525	41.26	0.000	.0217121	.0238778

```
. reg outcome exp age sex c1 c2 c3

      Source |       SS       df       MS              Number of obs =   78730
-------------+------------------------------           F(  6, 78723) =  218.16
       Model |  24.7418045     6  4.12363408            Prob > F      =  0.0000
    Residual |  1488.01751 78723  .018901941           R-squared     =  0.0164
-------------+------------------------------           Adj R-squared =  0.0163
       Total |  1512.75932 78729  .019214766           Root MSE      =  .13748
```

Adj RD

outcome	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
exp	**-.0206018**	**.0012311**	**-16.73**	**0.000**	**-.0230147**	**-.0181889**
age	.0012435	.0000408	30.49	0.000	.0011635	.0013234
sex	-.0047138	.0010838	-4.35	0.000	-.0068381	-.0025896
c1	-.0126105	.0012686	-9.94	0.000	-.015097	-.010124
c2	.003701	.0022383	1.65	0.098	-.000686	.008088
c3	-.0030277	.0011144	-2.72	0.007	-.0052119	-.0008434
_cons	-.0455348	.0029075	-15.66	0.000	-.0512335	-.0398362

simple ***ivreg*** of exposure, instrument, and outcome

- The bold line shows the desired point estimates: an **absolute risk difference of -1.2 per 100** people, with a **95% confidence interval of -0.5 to -1.9 per 100**. Adding the IV has moved the point estimate toward the null, but increased the standard error by a factor of three.

```
. ivreg outcome (exp = iv)


Instrumental variables (2SLS) regression

      Source |       SS       df       MS              Number of obs =   63053
-------------+------------------------------           F(  1, 63051) =   12.32
       Model |   1.9715631     1   1.9715631           Prob > F      =  0.0004
    Residual |  1241.48864 63051  .019690229           R-squared     =  0.0016
-------------+------------------------------           Adj R-squared =  0.0016
       Total |   1243.4602 63052  .019721186           Root MSE      =  .14032

------------------------------------------------------------------------------
     outcome |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         exp |  -.0122035   .0034773    -3.51   0.000    -.019019   -.005388
       _cons |   .0220232   .0007775    28.32   0.000     .0204993    .0235472
------------------------------------------------------------------------------
Instrumented:  exp
Instruments:   iv
------------------------------------------------------------------------------
```

simple model with age, sex, and three major covariates adjusted

- In this case, adjusting for covariates made little difference: the point estimate changed from -1.2 per 100 to -1.5 per 100.

```
. ivreg outcome (exp = iv) age sex c1 c2 c3

Instrumental variables (2SLS) regression

      Source |       SS       df       MS              Number of obs =   63052
-------------+------------------------------           F(  6, 63045) =  154.85
       Model |  20.9205856     6  3.48676426            Prob > F      =  0.0000
    Residual |  1222.53921 63045  .019391533           R-squared     =  0.0168
-------------+------------------------------           Adj R-squared =  0.0167
       Total |   1243.4598 63051  .019721492           Root MSE      =  .13925

------------------------------------------------------------------------------
     outcome |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         exp |  -.0154683   .0034772    -4.45   0.000    -.0222835   -.0086531
         age |   .0012849   .0000469    27.42   0.000     .0011931    .0013768
         sex |  -.0053853   .0012283    -4.38   0.000    -.0077928   -.0029779
          c1 |  -.0123998   .0014531    -8.53   0.000    -.0152479   -.0095518
          c2 |    .002082   .0025857     0.81   0.421     -.002986     .00715
          c3 |  -.0041958   .0012639    -3.32   0.001     -.006673   -.0017187
       _cons |  -.0478727   .0032737   -14.62   0.000    -.0542892   -.0414561
------------------------------------------------------------------------------
Instrumented:  exp
Instruments:   age sex c1 c2 c3 iv
------------------------------------------------------------------------------
```

TABLE 2

Characteristics of adults 65 years and older in British Columbia and Pennsylvania stratified by type of APM received

Characteristic	British Columbia		Pennsylvania	
Atypical treatment	Conventional treatment	Atypical treatment	Conventional treatment	
Number of new drug starts	23,785	12,756	12,031	8,056
Age (mean)	80.32	79.89	83.58	83.30
Male (%)	35.1	39.7	15.1	20.1
Cardiac arrhythmia (%)	0.1	0.0	1.4	1.4
Cerebrovascular disease (%)	9.9	10.8	30.2	28.3
Congestive heart failure (%)	6.0	8.4	30.4	31.8
Hypertension (%)	24.1	22.3	64.2	57.2
Diabetes (%)	13.8	15.0	26.3	25.5
Myocardial infarction (%)	2.3	2.7	3.3	3.4
Other ischemic heart disease (%)	2.7	3.8	23.8	28.3
Other cardiovascular disorders (%)	16.6	20.2	57.7	55.4
Dementia (%)	12.6	9.7	19.0	7.8
Delirium (%)	8.4	7.4	15.2	11.7
Mood disorders (%)	25.3	15.6	35.5	21.8
Psychotic disorders (%)	16.7	11.2	24.4	21.7
Other psychiatric disorders (%)	4.5	3.1	7.9	5.7
Nursing home residence in previous 180 days (%)	26.8	31.0	20.2	15.5
Number of drugs used (mean)	7.34	7.36	7.82	6.65

Table 3. Comparison of Risk Difference Models in 3 Cohorts of Patients, Pennsylvania (1994–2003) and British Columbia, Canada (1996–2004)[a]

Exposure:	COX-2 Inhibitor		Conventional APM		Conventional APM	
Referent:	Nonselective NSAID		Atypical APM		Atypical APM	
Outcome:	Severe Gastrointestinal Complications		Death		Death	
Population:	Pennsylvania		Pennsylvania		British Columbia	
	RD × 100	95% CI	RD × 100	95% CI	RD × 100	95% CI
Crude OLS model	0.19	−0.03, 0.41	2.69	1.65, 3.73	4.46	3.69, 5.23
Adjusted OLS model	−0.07	−0.30, 0.16	3.91	2.68, 5.13	3.55	2.74, 4.37
2-stage least squares model[b]	−1.28	−2.56, 0.01	7.69	1.26, 14.12	4.00	0.94, 7.06
Logistic/OLS model[b,c]	−1.36	−2.58, -0.15	7.64	1.55, 13.74	4.84	1.80, 7.88
3-stage model[b,c]	−1.35	−2.53, −0.17	7.53	1.83, 13.24	4.76	1.81, 7.72
Crude probit marginal effects[d] model	0.19	−0.03, 0.41	2.69	1.68, 3.70	4.46	3.69, 5.23
Adjusted probit marginal effects[d] model	−0.05	−0.26, 0.17	3.51	2.32, 4.69	3.20	2.42, 3.97
IV-based probit marginal effects[d] model[b,c]	−1.41	−3.14, 0.32	8.94	1.64, 16.24	3.88	0.67, 7.08

Table 4

Difference in risk of all-cause mortality within 180 days of initiation of conventional versus atypical APM treatment.

Population and variation	Events in conventional APM group	Events in atypical APM group	Unadjusted OLS estimate	Age/sex-adjusted OLS estimate	Fully adjusted OLS estimate[a]	IV analysis estimate
British Columbia						
Base case (unrestricted)	1,806	2,307	4.46 (3.69, 5.23)	4.49 (3.75, 5.22)	3.55 (2.74, 4.37)	4.00 (0.94, 7.06)
Restricted to PCPs (R6)	1,735	2,115	4.24 (3.41, 5.06)	4.48 (3.68, 5.28)	3.59 (2.70, 4.48)	3.11 (−0.57, 6.79)
Pennsylvania						
Base case (unrestricted)	1,307	1,628	2.69 (1.65, 3.73)	2.47 (1.46, 3.49)	3.91 (2.68, 5.13)	7.69 (1.26, 14.12)
Restricted to PCPs (R6)	960	1,129	2.39 (1.07, 3.71)	2.29 (0.98, 3.60)	4.32 (2.71, 5.93)	5.34 (−3.53, 14.21)

NOTE. The values within brackets are 95% confidence intervals. Risk differences are expressed per 100 patients.

Abbreviations: APM, antipsychotic medication; OLS, ordinary least squares; IV, instrumental variable; PCP, primary care physician.

[a] Adjusted for age, sex, race, year of treatment, and history of diabetes, arrhythmia, cerebrovascular disease, myocardial infarction, congestive heart failure, hypertension, other ischemic heart disease, other cardiovascular disorders, dementia, delirium, mood disorders, psychotic disorders, other psychiatric disorders, antidepressant use, nursing home residence, and hospitalization.

Table 5

Characteristics of the first-stage instrumental variable regression model for each study variation

	British Columbia					Pennsylvania				
	First-stage models					First-stage models				
Study variation	*N*	Unadj. IV to treatment OR	Adj. IV to treatment OR[a]	First-stage partial *F* statistic[a]	Partial r^2 value study variation	*N*	Unadj. IV to treatment OR	Adj. IV to treatment OR[a]	First-stage partial *F* statistic[a]	Partial r^2 value[a]
Base case	31,976	6.15	3.80	909	0.075	13,131	6.60	3.29	428	0.054
R1	24,085	6.68	3.98	613	0.079	4,577	8.06	3.19	250	0.052
R2	29,741	6.37	3.86	814	0.076	9,198	7.46	3.30	539	0.055
R3	2,235	3.97	2.66	78	0.042	3,933	5.19	2.81	153	0.038
R4	28,205	6.44	3.88	753	0.076	9,024	7.64	3.36	545	0.057
R5	3,771	4.44	2.93	142	0.051	4,107	5.08	2.79	158	0.037
R6	27,352	5.65	3.37	736	0.062	8,602	6.38	3.10	428	0.048
R7	4,462	8.79	5.27	88	0.099	4,184	7.08	3.62	275	0.062
R8	—	—	—	—	—	6,538	6.76	3.51	425	0.061
R9	—	—	—	—	—	6,148	6.19	3.10	306	0.048
R10	16,774	6.51	4.00	589	0.078	5,432	6.51	3.59	343	0.060
R11	12,369	4.71	3.19	347	0.058	5,256	5.43	2.81	214	0.039
R12	14,914	5.91	3.80	531	0.073	5,362	6.12	3.35	302	0.054
R13	12,585	6.00	3.90	420	0.074	4,149	6.27	3.63	252	0.058
S1	22,242	4.70	3.21	448	0.055	7,092	5.20	2.93	296	0.040
S2	29,143	3.33	2.72	465	0.042	10,688	2.74	2.28	287	0.026
S3	28,563	3.14	2.64	421	0.040	9,608	2.66	2.26	244	0.025
S4	31,976	6.15	3.80	909	0.075	13,131	6.57	3.27	744	0.054
P1 (R1)	24,085	6.73	4.10	582	0.072	4,577	8.62	3.48	214	0.045
P2 (R1)	24,085	6.56	4.01	478	0.057	4,577	7.89	3.15	130	0.028
P3 (R1)	24,085	6.72	4.10	450	0.048	4,577	7.67	3.09	92	0.020
P4 (R1)	24,085	7.37	4.41	607	0.074	4,577	7.06	3.01	208	0.044

Result & Conclusion

Results:

- Partial r^2 ranged from 0.028 to 0.099. PPP generally alleviated imbalances in nonpsychiatry-related patient characteristics, and the overall imbalance was reduced by an average of 36% (±40%) over the two cohorts.

Conclusion:

- In the study setting, most of the 25 formulations of the PPP IV were strong IVs and resulted in a strong reduction of imbalance in many variations.
- The association between strength and imbalance was mixed.

Part 3: Application of Instrumental Variables in Genetic Studies - Mendelian Randomization

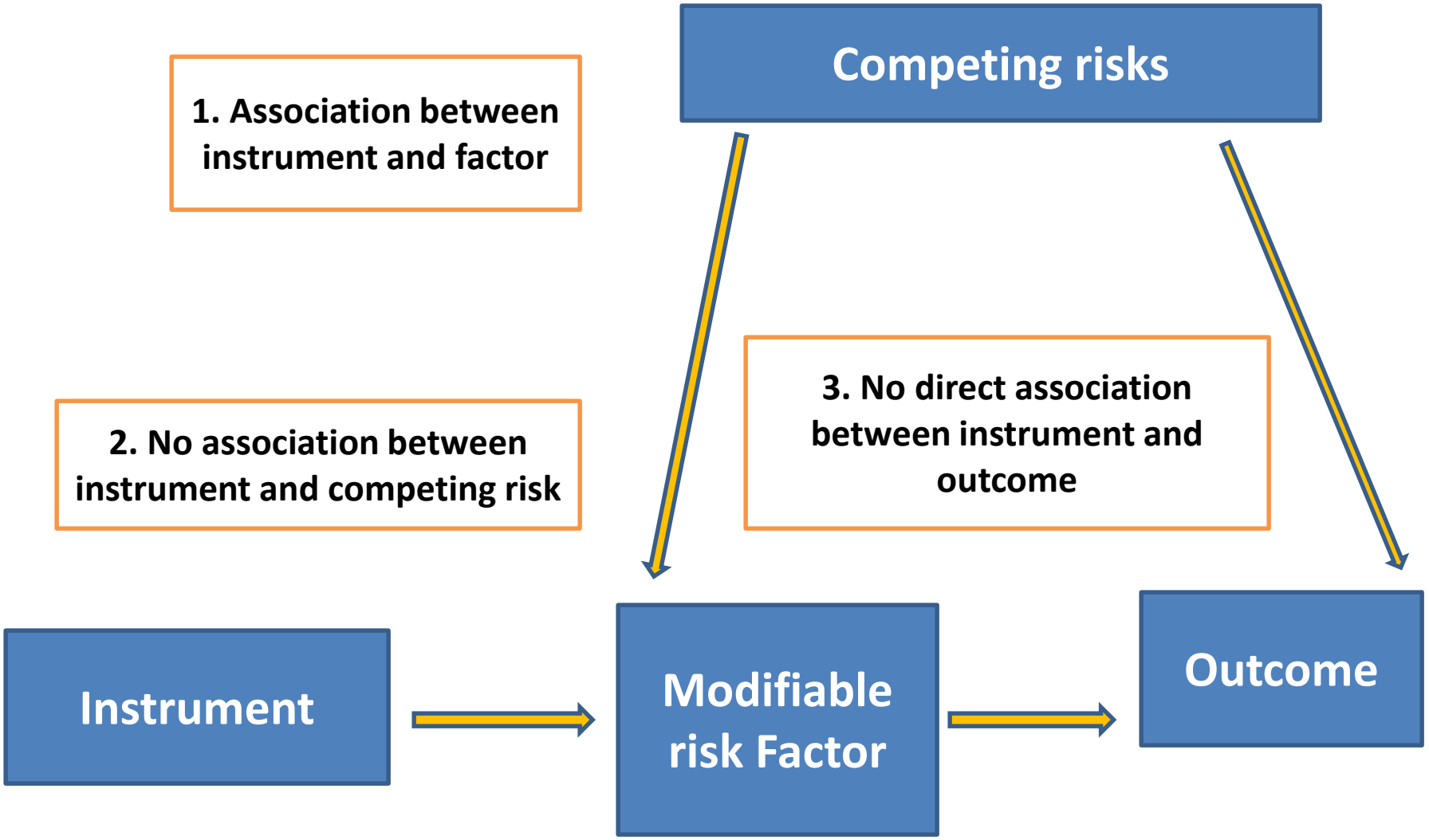
Statistical Methods in Medical Research
21(3) 223–242

DOI: 10.1177/0962280210394459
smm.sagepub.com

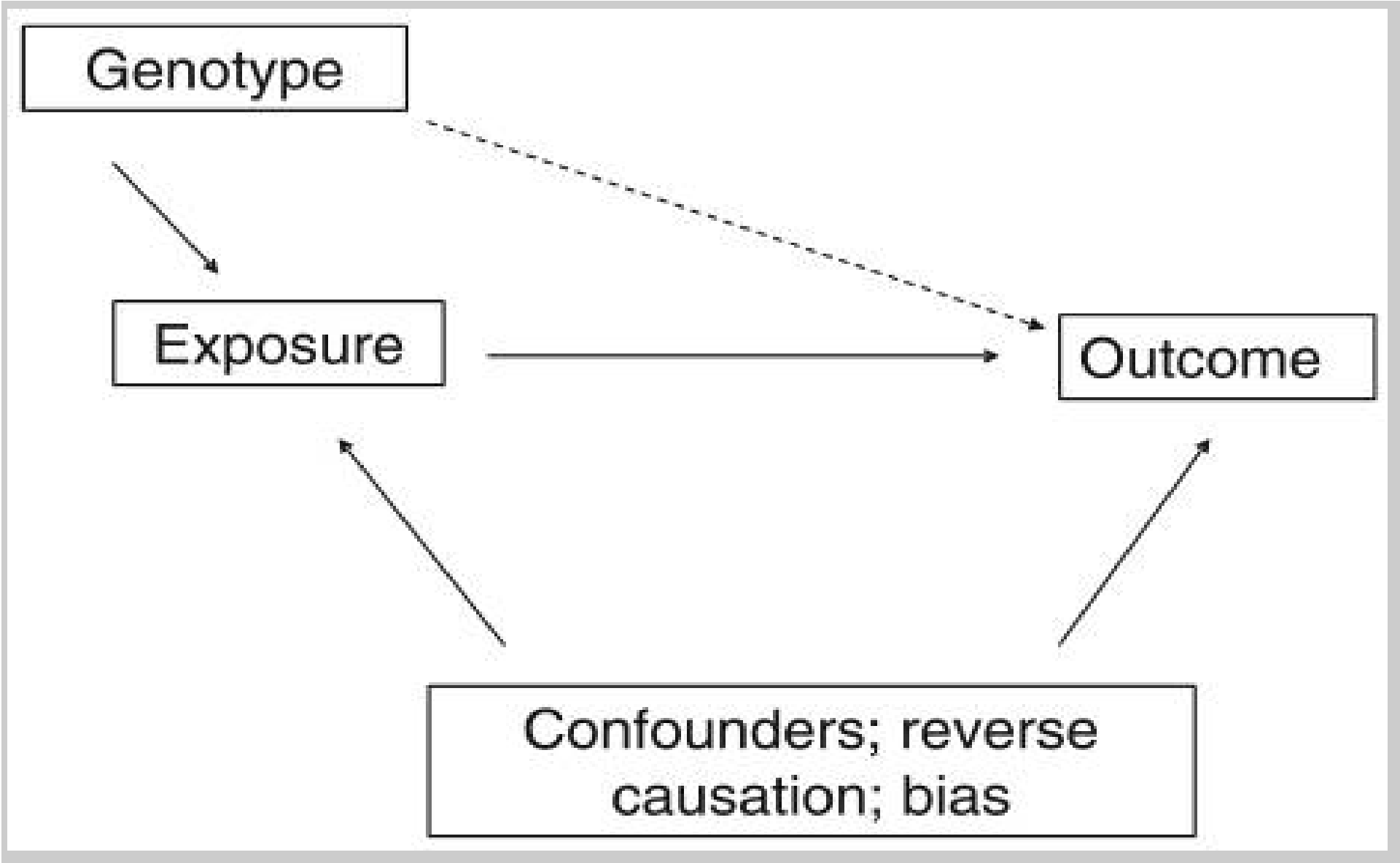
Using multiple genetic variants as instrumental variables for modifiable risk factors

Tom M Palmer,[1] Debbie A Lawlor,[1]
Roger M Harbord,[2] Nuala A Sheehan,[3] Jon H Tobias,[4]
Nicholas J Timpson,[1] George Davey Smith[1] and
Jonathan AC Sterne[2]

Criteria for Instrumental Variables (IVs)



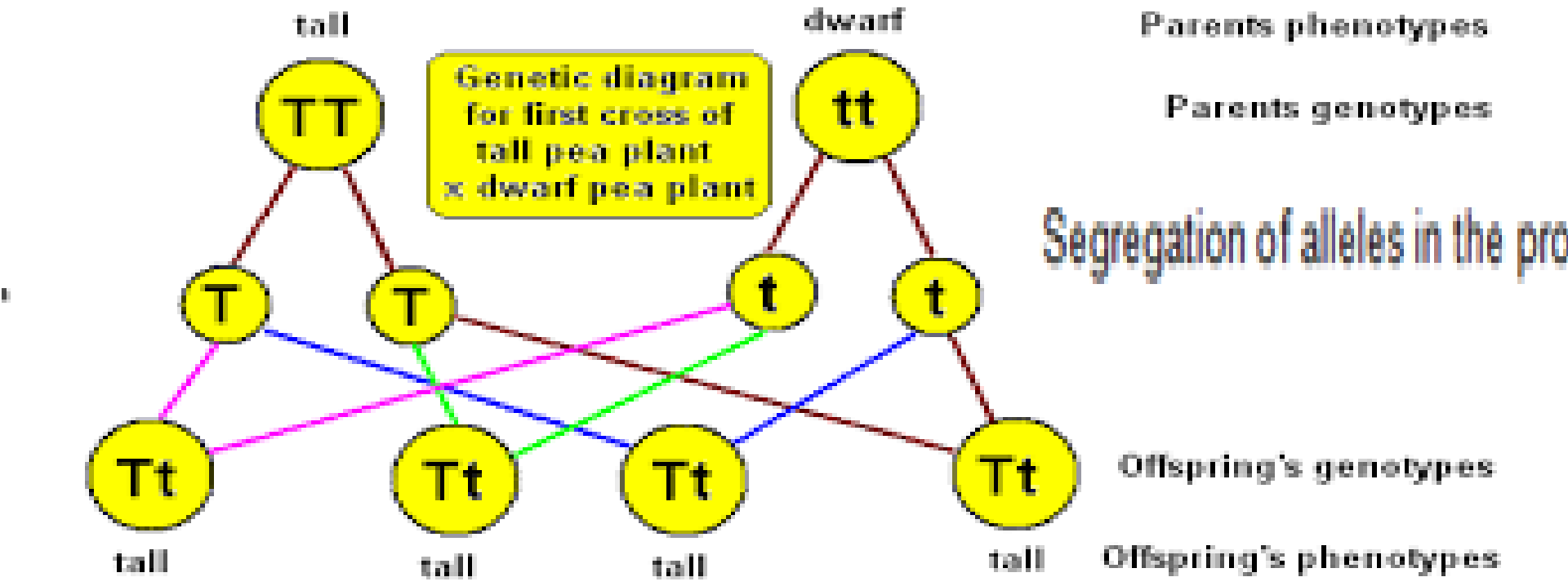
Mendelian randomization as an instrumental variables approach



Refresh Genetics 101 (Basic concepts of genetics)

Mendel's principles (laws) of inheritance

1. the principle of segregation
2. the principle of independent assortment



Overview-What is Mendelian randomization?

- Mendelian randomization technique (MRT) -the use of DNA (genetic) variants as instrumental variables to make epidemiological causal inferences about the effect of modifiable factors on health and disease-related outcomes in the presence of unobserved confounding of the relationship of interest in observational data.

- Mendelian randomization is “instrumental variable” analysis using genetic instruments”

Principles of Mendelian randomization?

- MRT is based on the principle that if a DNA variant is known to directly affect an intermediate phenotype.
- The phenotype could be a variant in the promoter of a gene encoding a biomarker that affects its expression
- If intermediate phenotype truly contributes to the disease, then the DNA variant should be said to be associated with the disease to the extent predicted by:

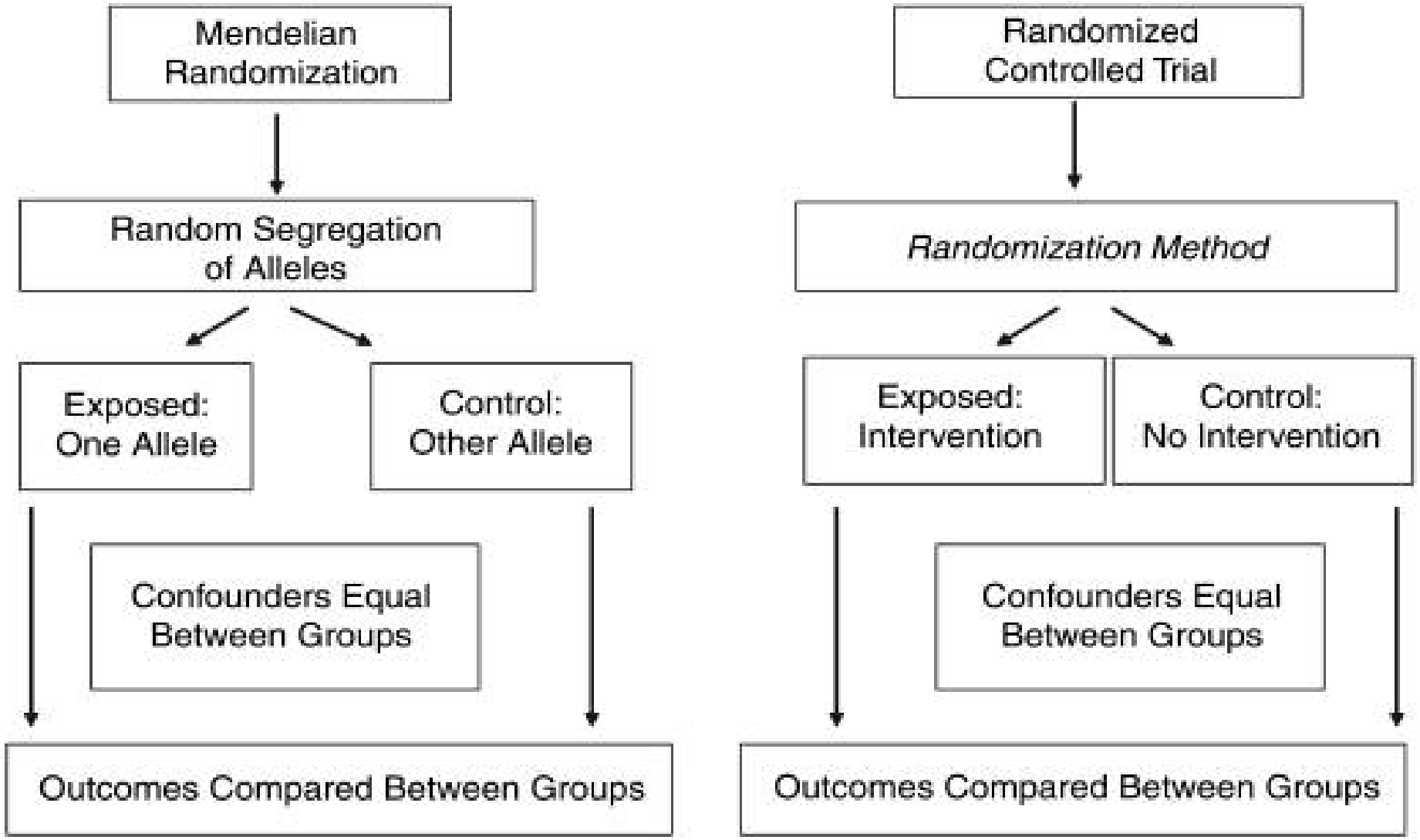
(1) the size of the effect of the variant on the phenotype

(2) the size of the effect of the phenotype on the disease

Application of Mendelian randomization?

- Use of Mendelian randomization is growing rapidly.
- However, using genetic variants as IVs poses statistical challenges.
- Particularly, there is a need for large sample sizes because of the relatively small proportion of variation in risk factors typically explained by genetic variants

Mendelian randomization and randomized controlled trial designs compared



Key points of Mendelian Randomization?

- The MR study design can be likened to a prospective randomized clinical trial in that the randomization for each individual occurs at the moment of conception
- At conception—genotypes of DNA variants are randomly ‘‘assigned’’ to gametes during meiosis, a process that should be impervious to the typical confounders observed in observational epidemiological studies.

Key points of Mendelian randomization-cont'd

- Genetic variants are ideal candidates for IVs, as genes are typically specific in function and ideally affect a single risk factor
- Genetic variation is determined at conception, so no reverse causation of an outcome on a genetic variant is possible.
- Genetic markers used as IVs are usually single nucleotide polymorphisms (SNPs)

Structure of the article by Palmer et al.

- **Section 1:** Description of the instrumental variable assumptions and introduction of an illustrative Mendelian randomisztion analysis with the presentation of separate IV estimates for four instruments
- **Section 2**: Discussion of the use of multiple instruments to help address some of the genetic and statistical issues that can affect Mendelian randomisation analyses
- **Sections 3 and 4**: Results of the simulation studies
- **Section 5**: Comparison of the IV estimates using multiple instruments and allele scores
- **Section 6**: Assessment of the impact of missing data

- **Section 7:** Discussion of the implications of the findings.

Illustrative Example of MRT:

- Illustration of Mendelian randomization using an example of four adiposity-associated genetic variants as IVs for the causal effect of fat mass on bone density, based on data of 5509 children enrolled in the ALSPAC birth cohort study

.

STUDY SETTING

The Avon Longitudinal Study of Parents and Children (ALSPAC), also known as Children of the 90s, is a world-leading birth cohort study, charting the health of 14,500 families in the Bristol area.

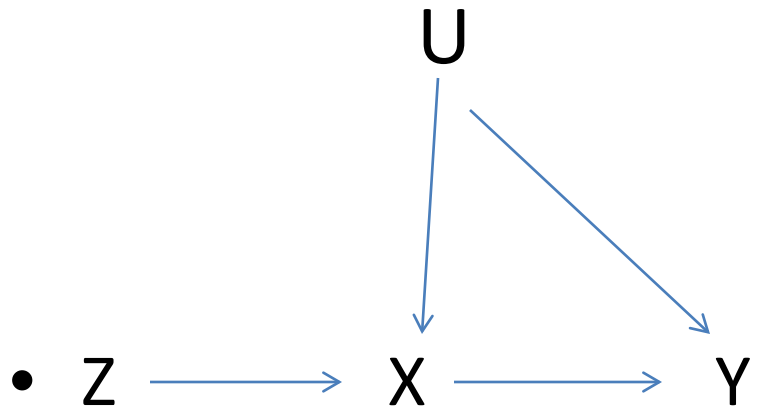
Section1: Instrumental variable assumptions

An IV (instrument) Z – genotype is defined as a variable that satisfies the following assumptions:

(1) It is associated with the risk factor (phenotype or intermediate variable) of interest X;

(2)It affects the outcome Y only through X. [No direct effect of Z on Y] –Exclusion restriction.

(3) It is independent of the (unobserved) confounding factors U of the association between X and the outcome Y



Section2: Illustrative Mendelian randomisation analysis: Single instrument estimates

- Investigation of the causal effect of fat mass on bone mineral density (BMD) using four genotypes known to be associated with adiposity from previous GWAS.
- A previous study using SNPs associated with the FTO and MC4R genes as IVs. found a positive effect of fat mass on BMD
- The authors concluded that higher fat mass caused increased accrual of bone mass in childhood.

Section2: Illustrative Mendelian randomisation analysis: Single instrument estimates , Cont'd

Current study is therefore to consider:

a) whether the IV estimates from the separate instruments are of similar magnitude;

b) whether use of multiple instruments increases the precision of IV estimates;

c) the use of allele scores as IVs; and

d) the impact of missing data on IV estimates

2.1. Data

- The illustrative example used data from the Avon Longitudinal Study of Parents and Children (ALSPAC).
- ALSPAC is a longitudinal, population-based birth cohort study that recruited 14 541 pregnant women resident in Avon, UK, with expected dates of delivery 1 April 1991 to 31 December 1992
- Out of this 13 988 live born infants survived to at least one year of age.
- Children eligible for inclusion in the analysis:

(1) had DNA available for genotyping;

(2) attended the research clinic at age 9 and

(3) had complete data on height and dual energy X-ray densitometry (DXA) scan-determined total fat mass and total BMD.

2.2. Selection of genotypes

- Eleven adiposity-related SNPs identified in previous GWAS have been genotyped in ALSPAC.
- Four SNPs, namely FTO (rs9939609), MC4R (rs17782313), TMEM18 (rs6548238) and GNPDA2 (rs10938397), that had the strongest association with adiposity in previous studies were chosen a priori for the IV analysis.
- Functional studies are required to ascertain the specific biological pathways through which these polymorphisms affect adiposity.
- However studies have shown that the pathways to greater adiposity are likely to involve influences on diet/appetite or physical activity.

3. Assessment of the IV assumptions

For the assessment of the IV assumptions they assumed:

- That the underlying mechanisms by which they influence diet or physical activity differ for each of the variants under consideration.
- Although current knowledge about their function is limited, their location on different chromosomes suggests that their influences may indeed be independent.

Encoded IV assumptions in a directed acyclic graph (DAG)

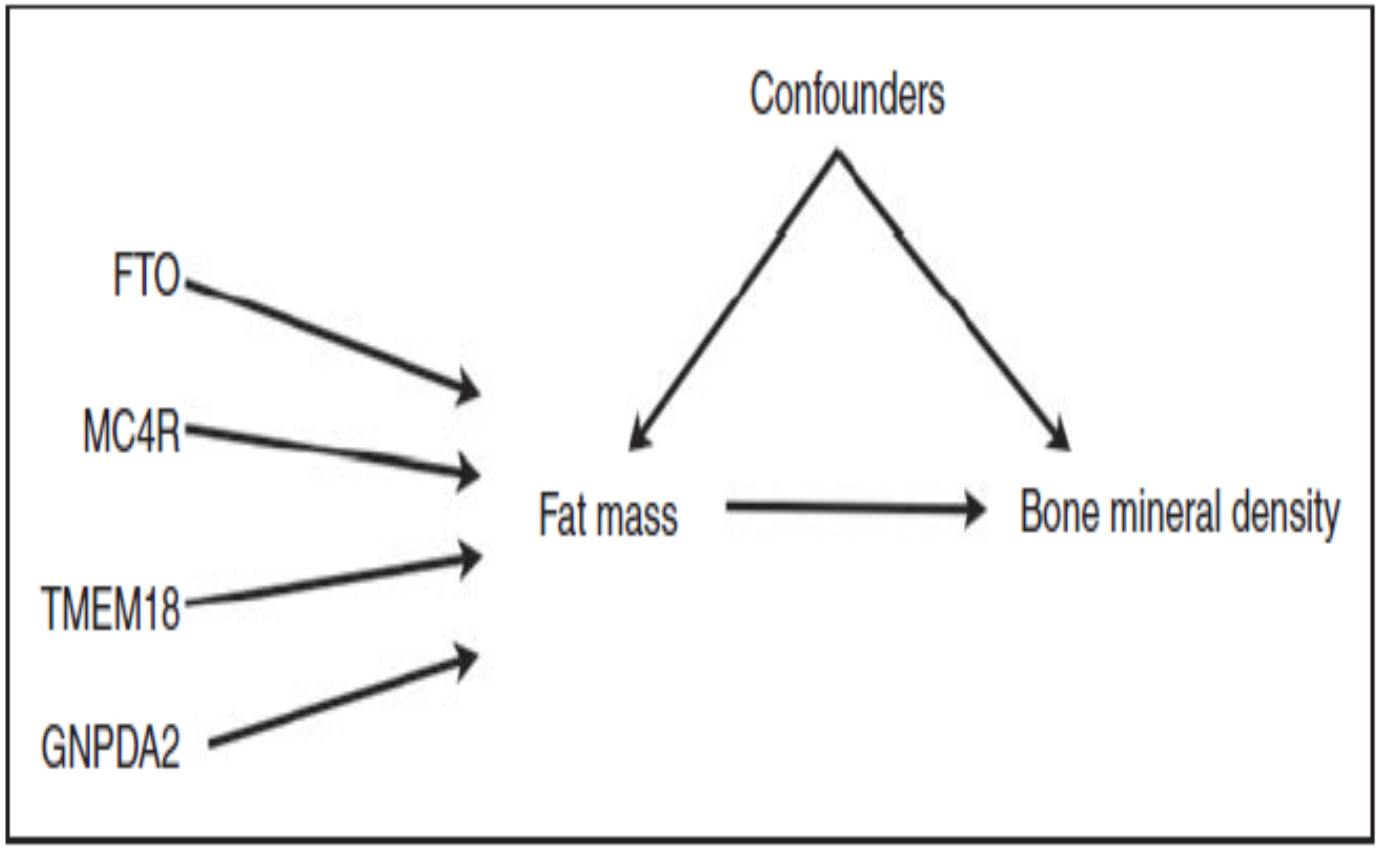


Figure 1. DAG for a Mendelian randomisation analysis using four genetic variants as instrumental variables for the effect of fat mass on bone mineral density.

Statistical methods:-Parametric data

- Fat mass and BMD were positively skewed and were log transformed.
- To account for sex and age differences in fat mass and BMD, age and sex standardised z-scores of log transformed fat mass and BMD were used in the analysis.
- Height and height-squared were included as covariates in analyses.
- They exponentiated parameter estimates to derive ratios of geometric mean BMD per standard deviation (SD) increase in log fat mass.
- Analyses were performed in Stata 11.0.

Statistical methods : Genetic data

- Genotypes were incorporated into IV models assuming an additive genetic model for the genotypes coded 0, 1 and 2
- They used the two-stage least squares (TSLS) for IV estimation
- The estimator was implemented in the user written Stata command ivreg
- The Hausman test of endogeneity was used to compare the difference between the ordinary-least-squares (OLS) and TSLS estimates using the user-written Stata command ivendog.
- In models including multiple instruments the Sargan test of over-identification available in the ivreg2 command, was used to test the joint validity of the instruments

Two-stage analysis

- The causal association can be estimated using a two-stage approach. With continuous outcomes, this is known as two-stage least squares (2SLS)
- In 2SLS, a linear regression of the risk factor is fitted on the IVs (*G–X regression), and secondly a* linear regression of the outcome on the fitted values for the risk factor from the first stage regression (*ˆX –Y regression).*
- *The 2SLS estimate (ˆ β2SLS) is the coefficient* for the increase in outcome per unit increase in risk factor.
- With binary outcomes, an analogous estimate has been proposed, called a two-stage , pseudo-2SLS - two-stage predictor substitution or Wald-type estimator

2 Stage least Squares Analysis

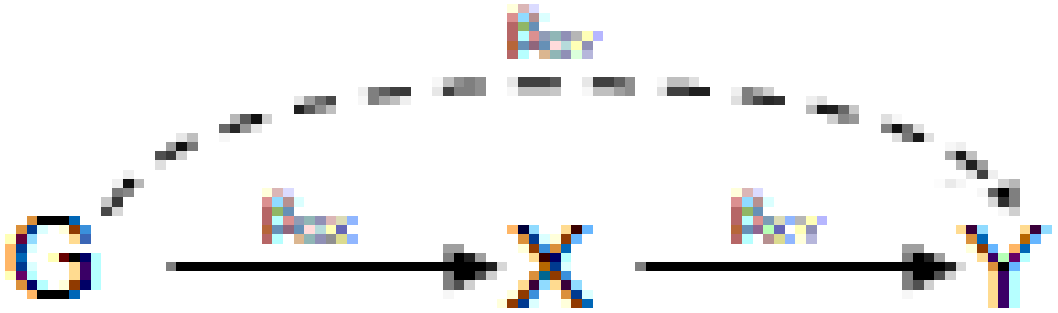
- This replaces the second linear ˆG –*Y regression with a logistic regression. With a single instrument,*
- the 2SLS and two-stage methods estimators coincide with the ratio of coefficients from the appropriate G–*Y regression (linear or logistic) divided by the coefficient from the G–X regression*
- There are several difficulties with this approach. Firstly, the fitted values for the risk factor are plugged into the second-stage regression without accounting for Secondly, the distribution of the causal parameter is assumed to be normal

Estimation of causal association

- If all associations are linear and subject to interactions, the causal effect of a factor on an outcome can be estimated by the ratio of :

$$\frac{\text{Regression coeff. of outcome (Y) on instrument(G)}}{\text{Regression coeff. of factor(X) on instrument (G)}}$$

$$= \beta_{GY} / \beta_{GX} = \beta_{XY}$$



2.4. Results for separate instruments:

Table 1. Study participant characteristics, total eligible children $N = 5509$

	N (%)	Mean (SD), geometric mean (95% CI) or *N* (%)	HWE *p*-value for genotypes
Gender: *N*(%) Female	5509 (100%)	2713 (49.3%)	
Age: Mean (SD) years	5509 (100%)	9.88 (0.32)	
BMD: geometric mean (95% CI) g/cm^2	5509 (100%)	0.902 (0.900, 0.903)	
Fat mass: geometric mean (95% CI) g	5509 (100%)	7209 (7100, 7320)	
Height: mean (SD) cm	5509 (100%)	139.6 (6.3)	
FTO (rs9939609):	5091 (92%)	TT = 0: 868 (37%) TA = 1: 2413 (47%) AA = 2: 810 (16%)	0.51
MC4R (rs17782313):	5412 (98%)	TT = 0: 3115 (58%) TC = 1: 2017 (37%) CC = 2: 280 (5%)	0.04
TMEM18 (rs6548238):	5323 (97%)	CC = 0: 3705 (70%) CT = 1: 1465 (28%) TT = 2: 153 (3%)	0.57
GNPDA2 (rs10938397):	5303 (96%)	AA = 0: 1731 (33%) AG = 1: 2604 (49%) GG = 2: 968 (18%)	0.84

HWE: Hardy–Weinberg Equilibrium.

2.4. Results for separate instruments:

Table 2. Associations of genotypes with potential confounding factors

Genetic variant	Covariate (unit) (*N*)	Number of risk alleles 0	1	2	Regression coefficient* (95% CI), *p*-value
Continuous confounding factors		Mean (95% CI)	Mean (95% CI)	Mean (95% CI)	
TO	Height (cm) (5091)	139.5 (139.2, 139.7)	139.6 (139.3, 139.8)	139.8 (139.4, 140.3)	0.18 (−0.07, 0.42), $p=0$
	Lean mass (g) (2515)	24426 (24218, 24634)	24620 (24439, 24800)	24593 (24287, 24899)	104 (−74, 283), $p=0.25$
MC4R	Height (cm) (5412)	139.7 (139.4, 139.9)	139.5 (139.2, 139.8)	140.1 (139.4, 140.9)	0.01 (−0.28, 0.29), $p=0$
	Lean mass (g) (2685)	24548 (24387, 24708)	24636 (24438, 24834)	24910 (24362, 25458)	128 (−78, 334), $p=0.22$
MEM18	Height (cm) (5323)	139.7 (139.5, 139.9)	139.5 (139.1, 139.8)	139.3 (138.3, 140.3)	−0.24 (−0.56, 0.08), $p=0$
	Lean mass (g) (2640)	24770 (24622, 24917)	24286 (24053, 24519)	24017 (23293, 24740)	−447 (−679, −215), $p<$
NPDA2	Height (cm) (5303)	139.5 (139.3, 139.8)	139.6 (139.4, 139.9)	139.7 (139.3, 140.1)	0.10 (−0.14, 0.34), $p=0$
	Lean mass (g) (2625)	24596 (24382, 24810)	24655 (24479, 24832)	24525 (24234, 24816)	−21 (−198, 155), $p=0.8$
Categorical confounding factors		*n*/*N* (%)	*n*/*N* (%)	*n*/*N* (%)	Odds ratio* (95% CI), *p*-va
TO	MEA (2421) HHSC (2329)	139/857 (16%)	189/1161 (16%)	69/403 (17%)	1.03 (0.88, 1.20), $p=0.726$ Chi-squared $p=0.038$
MC4R	MEA (2591) HHSC (2485)	255/1492 (17%)	155/971 (16%)	25/128 (20%)	0.99 (0.83, 1.18), $p=0.929$ Chi-squared $p=0.432$
MEM18	MEA (2543) HHSC (2438)	314/1765 (18%)	107/705 (15%)	4/73 (5%)	0.74 (0.60, 0.92), $p=0.006$ Chi-squared $p=0.556$
NPDA2	MEA (2532) HHSC (2432)	151/838 (18%)	203/1236 (16%)	69/458 (13%)	0.90 (0.77, 1.04), $p=0.159$ Chi-squared $p=0.754$

MEA: Mother's highest educational achievement is a binary variable derived from the groups 0 = CSE, O-level, Vocational and 1 = A-level and degree.
HHSC: Head of household social class coded as categorical variable I, II, III non-manual, III manual, IV and V.
*Assuming an additive genetic model.

2.4. Results for separate instruments:

Table 3. OLS and IV estimates of the effect of fat mass on bone mineral density (BMD) based on complete case analysis, $N = 4796$[a]

Method	First stage regression coefficient (95% CI)	First stage R^2	First stage F-statistic	Ratio of geometric mean BMD[b] (95% CI)	SE of estimate (log scale)	Hausman test p-value	Sargan test P-value
OLS	NA	NA	NA	1.22 (1.19, 1.26), $p < 0.001$	0.014	NA	NA
IV: SNP(s) used as IV							
FTO	0.11 (0.08, 0.15)	0.0082	39.83	1.44 (1.05, 1.97), $p = 0.024$	0.16	0.300	NA
MC4R	0.09 (0.05, 0.13)	0.0037	17.85	2.33 (1.34, 4.05), $p = 0.003$	0.28	0.006	NA
TMEM18	−0.06 (−0.11, −0.02)	0.0016	7.47	2.27 (0.98, 5.28), $p = 0.056$	0.43	0.089	NA
GNPDA2	0.05 (0.01, 0.09)	0.0016	7.57	0.98 (0.47, 2.03), $p = 0.953$	0.37	0.540	NA
FTO, MC4R	NA	0.0119	29.92	1.67 (1.27, 2.19), $p < 0.001$	0.14	0.020	0.11
FTO, MC4R, TMEM18	NA	0.0136	21.95	1.73 (1.34, 2.24), $p < 0.001$	0.13	0.010	0.22
FTO, MC4R, TMEM18, GNPDA2	NA	0.0153	18.59	1.63 (1.28, 2.06), $p < 0.001$	0.12	0.013	0.16
Unweighted allele score (4 SNPs)	0.06 (0.04, 0.08)	0.0069	33.15	1.40 (0.99, 1.98), $p = 0.055$	0.18	0.430	NA
Weighted allele score (4 SNPs)	0.19 (0.15, 0.24)	0.0153	74.35	1.63 (1.29, 2.07), $p < 0.001$	0.12	0.012	NA

[a]Analyses adjusted for height and height squared.

[b]For a 1 unit increase in z-score of age and gender standardised fat mass.

Section 3: Using multiple instruments to address potential biases in Mendelian randomization analyses

- Population stratification, linkage disequilibrium and pleiotropy have been identified as factors that could bias Mendelian randomization analyses

- The use of multiple instruments to address issues they raise.

Section 3: Using multiple instruments to address potential biases in Mendelian randomization analyses

- Comparison of IV estimates from independent genetic variants is analogous to comparing the results of RCTs of different classes of blood pressure lowering drugs, which lower blood pressure by different mechanisms.
- If the effect of the drug on stroke risk in each RCT is proportional to the direction and magnitude of its effect on blood pressure,
- It strengthens the evidence for a causal link between blood pressure and stroke risk, and against the drugs having effects on stroke risk through other mechanisms.

Section 4: Statistical issues relating to use of multiple instruments in Mendelian randomization analyses

- **Over-identification** -the situation when there is more than one instrument for a single risk factor of interest or, more generally, when there are more instruments than endogenous variables.
- In such circumstances testing the 'over-identification restriction' checks the joint validity of multiple instruments by testing whether they give the same estimates when used singly or in linear combination.
- Two commonly used tests of over-identification; the Hansen test and the Sargan test.

Section 4.2 : Finite sample bias and instrument strength

- IV estimators such as TSLS are asymptotically unbiased but biased in finite samples, with such bias inversely proportional to the amount of phenotypic variability explained by the instrument.
- Two closely related measures of this are the first-stage regression F-statistic and coefficient of determination R2.
- It is important to report these. If measured confounders are included then the partial R2 and F-statistics for the instruments should be reported.

Section 4.2 : Finite sample bias and instrument strength- Cont'd

- In Mendelian randomisation the first stage R2 is the proportion of risk factor variability explained by genotype. The relationship between the F and R2 statistics is given by:

$$F = \frac{R^2/k}{(1 - R^2)/(n - k - 1)}.$$

where k is the number of parameters in the model (in this case instruments). The relative bias of the TSLS estimator to the OLS estimator is related to the inverse of the F-statistic.

Section 4.2 : Finite sample bias and instrument strength-cont'd

Hahn and Hausman gave a simplified version of the relative bias as approximately the inverse of the F-statistic

$$\frac{\text{bias TSLS}}{\text{bias OLS}} \approx \frac{k}{nR^2}$$

As R2 increases the relative bias of TSLS decreases, but including additional instruments that do not increase the first stage R2 increases the relative bias of TSLS.

A first stage F-statistic less than 10 is often taken to indicate a weak instrument, although this is not a strict limit but a rule of thumb drawn from simulation studies.

4.3 Statistical power

- Genotypic effects on phenotypes are typically small, so Mendelian randomization analyses can require very large sample sizes to obtain adequate power.
- When multiple instruments are used in the TSLS estimator, the resulting IV estimate can be viewed as the efficient linear combination of the separate IV estimates; provided that each instrument is valid
- Use of multiple instruments will increase the precision of the IV estimate compared with the separate IV estimates

4.4 Use of an allele score as an instrumental variable

- An allele score is a weighted or unweighted sum of the number of ‘risk’ alleles across several genotypes: weights are usually based on each genotype’s effect on the phenotype.
- Use of such scores is becoming more common in gene–disease association studies.
- To justify the use of an allele score the genotypes should have an approximately additive effect on the risk factor.
- For an unweighted score they should also have similar per allele effects

5.1. Multiple instrument simulations

5.1 Simulation 1: non-weak instruments

Data were simulated as follows, where G_1, G_2 and G_3 are genotype variables coded additively, X is the risk factor, Y the disease outcome, U the unmeasured confounder and subscript i denotes a subject:

$$G_{1i} \sim \text{Bin}(2, 0.3),\ G_{2i} \sim \text{Bin}(2, 0.3),\ G_{3i} \sim \text{Bin}(2, 0.3), \text{ and,}$$

$$U_i \sim N(0, 1),$$

$$X_i = 0.55G_{1i} + 0.4G_{2i} + 0.25G_{3i} + U_i \text{ and } Y_i = \beta X_i + U_i.$$

The values of the coefficients on the genotypes were chosen so that G_1 explained the most variability in X, followed by G_2 and G_3. The value of the causal effect of X on Y, β, was set to 1. We monitored the estimates of β from the following models:

(1) OLS estimate of the regression of Y on X,
(2) TSLS using G_1 as the instrument,
(3) TSLS using G_1 and G_2 as instruments,
(4) TSLS using G_1–G_3 as instruments,
(5) TSLS using an unweighted allele score of G_1–G_3 as an instrument,
(6) TSLS using a weighted allele score of G_1–G_3 as an instrument.

5.2 Simulation 1: results

Table 4. Simulation 1 (non-weak instruments): results (Monte Carlo standard error reported in brackets beside each estimate)

Model	Average bias	MSE	Average SE	Coverage	Average R^2	Average F	Average absolute TSLS/OLS bias ratio
1. OLS	0.8194 (0.00005)	0.6714 (0.00009)	0.0054 (7 E–7)	0	NA	NA	NA
2. TSLS G_1	–0.0019 (0.0004)	0.0016 (0.00002)	0.03991 (0.00003)	0.9523 (0.0021)	0.1163 (0.0001)	581.41 (0.504)	0.0022 (0.0005)
3. TSLS G_1 & G_2	–0.00004 (0.0003)	0.0010 (0.00002)	0.03215 (0.00002)	0.9467 (0.0022)	0.1898 (0.0001)	474.09 (0.333)	0.0001 (0.0004)
4. TSLS G_1–G_3	0.00084 (0.0003)	0.0009 (0.00001)	0.0301 (0.00002)	0.9487 (0.0022)	0.2212 (0.0001)	368.41 (0.243)	0.0012 (0.0004)
5. TSLS allele score G_1–G_3	–0.00098 (0.0003)	0.0010 (0.00002)	0.0316 (0.00002)	0.9486 (0.0022)	0.1981 (0.0001)	990.22 (0.685)	0.0010 (0.0004)
6. TSLS weighted allele score G_1–G_3	0.00084 (0.0003)	0.0009 (0.00001)	0.0301 (0.00002)	0.9492 (0.0022)	0.2212 (0.0001)	1105.43 (0.730)	0.0012 (0.0004)

MSE: mean squared error, SE: standard error, TSLS: two-stage least squares, OLS: ordinary least squares.

5.2. Results from simulation- cont'd

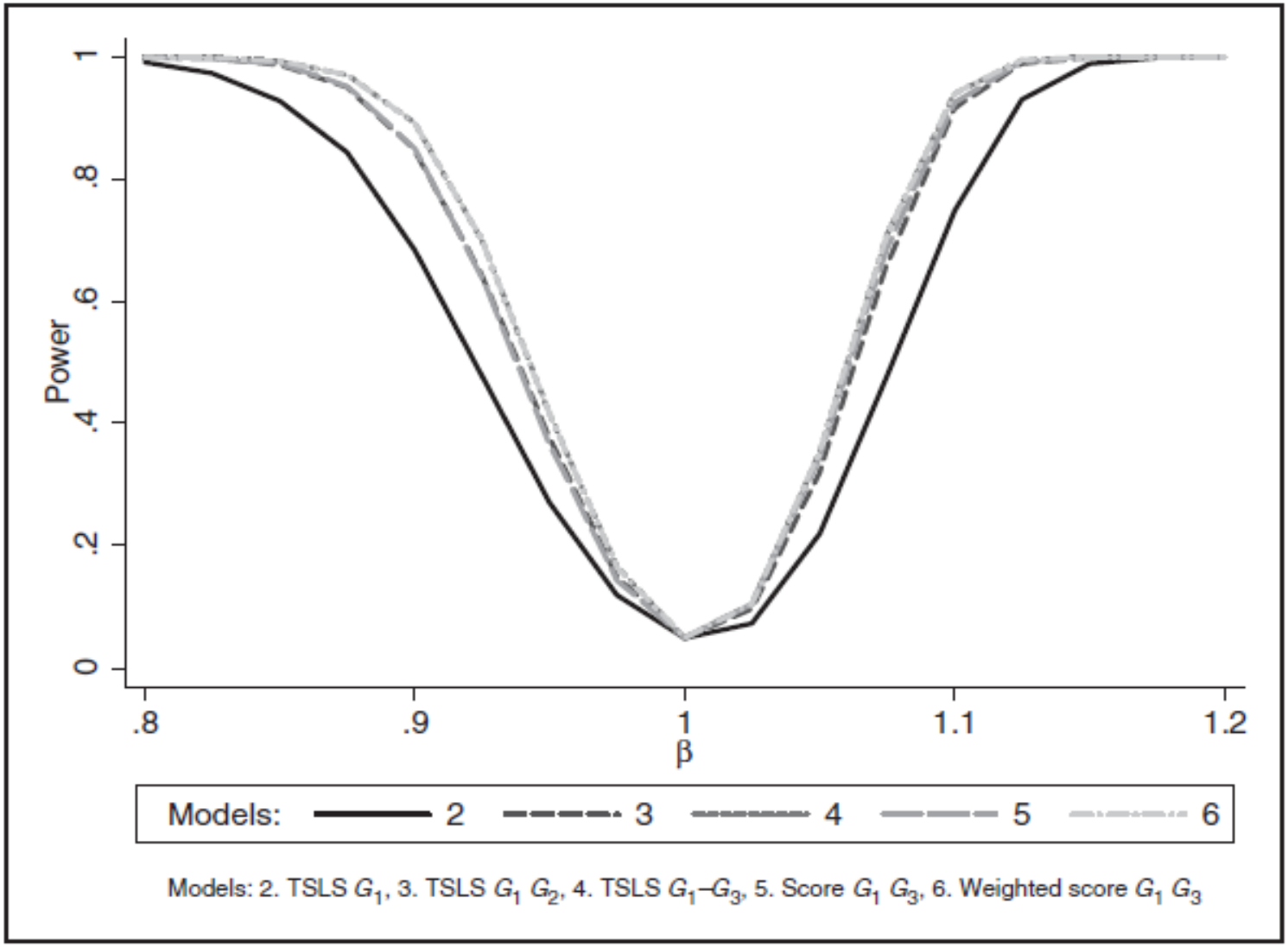


Figure 2. Simulation I (non-weak instruments): power curves.

5.3 Simulation 2: non-weak and weak instruments

Data were simulated with four IVs as follows such that G_1 and G_2 had F-statistics greater than 10 and G_3 and G_4 had F-statistics less than 10. The variables were simulated as: $G_{1i} \sim \text{Bin}(2,0.4)$, $G_{2i} \sim \text{Bin}(2,0.2)$, $G_{3i} \sim \text{Bin}(2,0.2)$, $G_{4i} \sim \text{Bin}(2,0.4)$, and, $U_i \sim N(10,1)$, $X_i = 0.1G_{1i} + 0.1G_{2i} + 0.05G_{3i} + 0.05G_{4i} + U_i$ and $Y_i = \beta X_i + U_i$. The value of the causal effect of X on Y, β, was set to 1. We monitored the estimates of β from the following models:

(1) OLS estimate from regression of Y on X;
(2) TSLS estimate using G_1 as the IV;
(3) TSLS estimate using G_1 and G_2 as the IVs;
(4) TSLS estimate using G_1, G_2, G_3 and G_4 as the IVs;
(5) TSLS estimate using an unweighted allele score of G_1 and G_2 as the IV;
(6) TSLS estimate using a weighted allele score of G_1 and G_2 as the IV;
(7) TSLS estimate using an unweighted allele score of G_1–G_4 as the IV;
(8) TSLS estimate using a weighted allele score of G_1–G_4 as the IV.

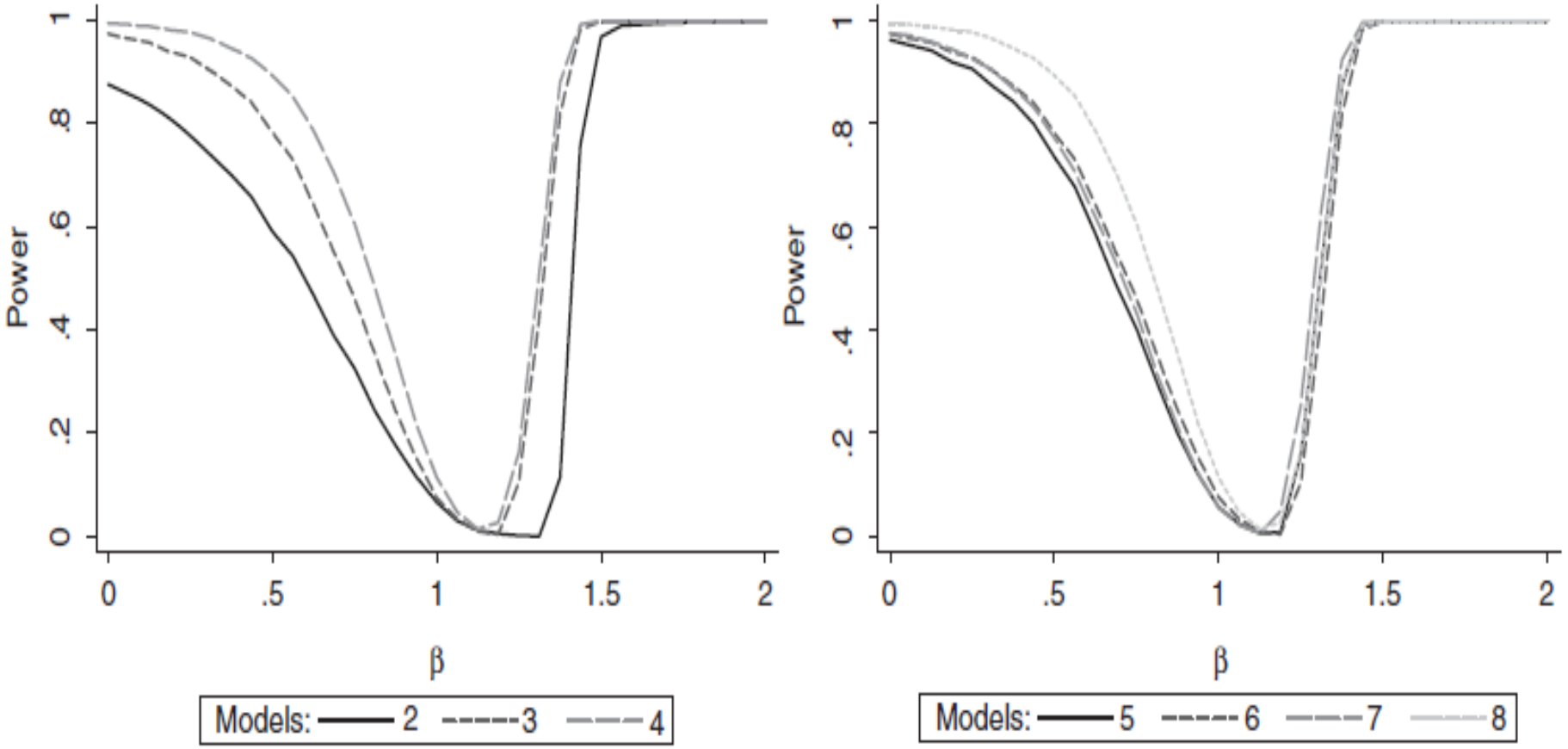
5.4. Simulation 2:Results for joint weak & strong instruments

Table 5. Simulation 2 (non-weak and weak instruments): results (Monte Carlo standard error in brackets beside each estimate)

Model	Average bias	MSE	Average SE	Coverage	Average R^2	Average F	Av. absolute TSLS/OLS bias ratio
1. OLS	0.990 (0.00001)	0.980 (0.00003)	0.0014 (1.9E-7)	0 (0)	NA	NA	NA
2. TSLS G_1	−0.047 (0.0025)	0.067 (0.003)	0.237 (0.0015)	0.93 (0.0025)	0.005 (0.00002)	24.92 (0.099)	0.047 (0.003)
3. TSLS G_1 & G_2	0.001 (0.0017)	0.028 (0.0006)	0.164 (0.0006)	0.92 (0.0027)	0.008 (0.00003)	20.99 (0.065)	0.001 (0.002)
4. TSLS G_1–G_4	0.040 (0.0013)	0.020 (0.0003)	0.137 (0.0004)	0.89 (0.0031)	0.011 (0.00003)	13.50 (0.036)	0.041 (0.001)
5. TSLS allele score G_1 & G_2	−0.026 (0.0018)	0.032 (0.0007)	0.172 (0.0006)	0.94 (0.0024)	0.008 (0.00003)	40.99 (0.128)	0.027 (0.002)
6. TSLS weighted allele score G_1 & G_2	0.001 (0.0017)	0.028 (0.0006)	0.164 (0.0006)	0.92 (0.0027)	0.008 (0.00003)	41.99 (0.129)	0.001 (0.002)
7. TSLS allele score G_1–G_4	−0.024 (0.0016)	0.027 (0.0006)	0.160 (0.0005)	0.94 (0.0024)	0.009 (0.00003)	45.91 (0.136)	0.024 (0.002)
8. TSLS weighted allele score G_1–G_4	0.040 (0.0013)	0.020 (0.0003)	0.137 (0.0004)	0.89 (0.0031)	0.011 (0.00003)	54.01 (0.145)	0.041 (0.001)

MSE: mean squared error, SE: standard error, TSLS: two-stage least squares, OLS: ordinary least squares.

5.4 Simulation 2: results- cont'd



Models: 2. TSLS G_1, 3. TSLS G_1 G_2, 4. TSLS G_1–G_4, 5. Score G_1 G_2, 6. Weighted score G_1 G_2, 7. Score G_1–G_4, 8. Weighted score G_1–G_4

Figure 3. Simulation 2 (non-weak and weak instruments): power curves.

6. Multiple instrument estimates and assessment of missing data

6.1 Multiple instrument estimates

Table 3. OLS and IV estimates of the effect of fat mass on bone mineral density (BMD) based on complete case analysis, $N = 4796$[a]

Method	First stage regression coefficient (95% CI)	First stage R^2	First stage *F*-statistic	Ratio of geometric mean BMD[b] (95% CI)	SE of estimate (log scale)	Hausman test *p*-value	Sargan test P-value
OLS	NA	NA	NA	1.22 (1.19, 1.26), $p < 0.001$	0.014	NA	NA
IV: SNP(s) used as IV							
FTO	0.11 (0.08, 0.15)	0.0082	39.83	1.44 (1.05, 1.97), $p = 0.024$	0.16	0.300	NA
MC4R	0.09 (0.05, 0.13)	0.0037	17.85	2.33 (1.34, 4.05), $p = 0.003$	0.28	0.006	NA
TMEM18	−0.06 (−0.11, −0.02)	0.0016	7.47	2.27 (0.98, 5.28), $p = 0.056$	0.43	0.089	NA
GNPDA2	0.05 (0.01, 0.09)	0.0016	7.57	0.98 (0.47, 2.03), $p = 0.953$	0.37	0.540	NA
FTO, MC4R	NA	0.0119	29.92	1.67 (1.27, 2.19), $p < 0.001$	0.14	0.020	0.11
FTO, MC4R, TMEM18	NA	0.0136	21.95	1.73 (1.34, 2.24), $p < 0.001$	0.13	0.010	0.22
FTO, MC4R, TMEM18, GNPDA2	NA	0.0153	18.59	1.63 (1.28, 2.06), $p < 0.001$	0.12	0.013	0.16
Unweighted allele score (4 SNPs)	0.06 (0.04, 0.08)	0.0069	33.15	1.40 (0.99, 1.98), $p = 0.055$	0.18	0.430	NA
Weighted allele score (4 SNPs)	0.19 (0.15, 0.24)	0.0153	74.35	1.63 (1.29, 2.07), $p < 0.001$	0.12	0.012	NA

[a]Analyses adjusted for height and height squared.

[b]For a 1 unit increase in z-score of age and gender standardised fat mass.

6.2 Assessment of missing data

Table 6. IV estimates of the effect of fat mass on bone mineral density (BMD) using all available data[a]

SNPs used as instrumental variable	N	First stage regression coefficient (95% CI)	First stage R^2	First stage F-statistic	Ratio of geometric mean BMD[b] (95% CI)	SE of estimate (log scale)	Hausman test p-value	Sargan test p-value
OLS	5509	NA	NA	NA	1.22 (1.18, 1.25), $p < 0.001$	0.014	NA	NA
IV: SNP(s) used as IV								
FTO	5091	0.12 (0.08, 0.15)	0.0088	45.35	1.41 (1.05, 1.89), $p = 0.023$	0.15	0.320	NA
MC4R	5412	0.09 (0.05, 0.13)	0.0037	19.95	2.42 (1.42, 4.12), $p = 0.001$	0.27	0.002	NA
TMEM18	5323	−0.06 (−0.11, −0.02)	0.0013	6.99	2.17 (0.92, 5.12), $p = 0.077$	0.44	0.130	NA
GNPDA2	5303	0.05 (0.01, 0.08)	0.0013	6.90	0.92 (0.42, 2.01), $p = 0.84$	0.40	0.463	NA
FTO, MC4R	5007	NA	0.0125	31.61	1.60 (1.24, 2.07), $p < 0.001$	0.13	0.029	0.221
FTO, MC4R, TMEM18	4881	NA	0.0138	22.75	1.69 (1.32, 2.17), $p < 0.001$	0.13	0.006	0.227

[a]Analyses adjusted for height and height squared.

[b]For a 1 unit increase in z-score of age and gender standardised fat mass.

Conclusion

- The illustrative Mendelian randomisation analysis confirmed a positive causal effect of adiposity (fat mass) on BMD the result suggested that the size of this effect was larger than that estimated by ignoring unmeasured confounding and using ordinary least squares, based on the Hausman endogeneity test.
- The SE of the IV estimate decreased by around 20% using all four genotypes, compared with the SE of the IV estimate using only the genotype with the strongest effect on risk factor. Such a reduction in SE corresponds to a 56% increase in sample size.
- With increasing availability of multiple genetic variants associated with the same risk factor or disease outcome, it is becoming common for genetic association studies to report associations with allele scores.
- Before an allele score is used as an IV the joint validity of the SNPs should be assessed using an over-identification test.

CLOSING COMMENTS

Mendelian randomization has potential shortcomings:

(1) The technique is only as reliable as the robustness of the estimates of the effect sizes of the variant on the phenotype and of the phenotype on disease

(2) It assumes that the DNA variant does not influence the disease by means other than the intermediate phenotype being studied (pleiotropy), which may not be true.

Nevertheless, Mendelian randomization has the potential to be as informative as a traditional randomized clinical trial.