# The Future of Mendelian Randomization Studies

**Lorentz Center, 13-17 December 2021**

**Research Presentations and Abstracts**

Research presentations

<table>
<thead>
<tr>
<th>Tuesday 9.00-10.45hr</th>
<th>Wednesday 14.15-16.00hr*</th>
<th>Thursday 14.15-16.00hr*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Co-chairs: Deborah Lawlor &amp; Kate Tilling</td>
<td>Co-chairs: Sonja Swanson &amp; Tom Palmer</td>
<td>Co-chairs: Eric Tchetgen Tchetgen &amp; Stephen Burgess</td>
</tr>
<tr>
<td><strong>9.00-9.15hr</strong></td>
<td><strong>14.15-14.30hr</strong></td>
<td><strong>14.15-14.30hr</strong></td>
</tr>
<tr>
<td>Identifying effects of multiple treatments in the presence of unmeasured confounding</td>
<td>Automated analysis of discrete IV with finite-sample guarantees</td>
<td>Falsification of MR analysis maternal alcohol &amp; offspring ADHD</td>
</tr>
<tr>
<td><strong>9.15-9.30hr</strong></td>
<td><strong>14.30-14.45hr</strong></td>
<td><strong>14.30-14.45hr</strong></td>
</tr>
<tr>
<td>Fatima Batool (Cambridge, UK)</td>
<td>Amy Mason (Cambridge, UK)</td>
<td>Zijun Gao (Stanford, US)</td>
</tr>
<tr>
<td>Dimensionality Reduction Methods to infer accurate causal risk factors using MVMR</td>
<td>Breaking down the barriers to non-linear Mendelian randomization</td>
<td>Biological pathways with genome-phenome summary data</td>
</tr>
<tr>
<td><strong>9.30-9.45hr</strong></td>
<td><strong>14.45-15.00hr</strong></td>
<td><strong>14.45-15.00hr</strong></td>
</tr>
<tr>
<td>Eleanor Sanderson (Bristol, UK)</td>
<td>Matt Tudball (Bristol, UK)</td>
<td>Ting Ye (University of Washington, US)</td>
</tr>
<tr>
<td>Estimation of causal effects of an exposure at multiple time points through Multivariable Mendelian randomization</td>
<td>Almost exact Mendelian randomization</td>
<td>Robust Mendelian randomization in the presence of many weak instruments and widespread horizontal pleiotropy</td>
</tr>
<tr>
<td><strong>9.45-10.00hr</strong></td>
<td><strong>15.00-15.15hr</strong></td>
<td><strong>15.00-15.15hr</strong></td>
</tr>
<tr>
<td>Break</td>
<td>Time away from screen</td>
<td>Time away from screen</td>
</tr>
<tr>
<td><strong>10.00-10.15hr</strong></td>
<td><strong>15.15-15.30hr</strong></td>
<td><strong>15.15-15.30hr</strong></td>
</tr>
<tr>
<td>Marie Sadler (Swiss Institute of Bioinformatics)</td>
<td>Jeremy Labrecque (Erasmus MC, Netherlands)</td>
<td>Rebecca Richmond (Bristol, UK)</td>
</tr>
<tr>
<td>Quantifying DNA methylation-to-complex trait effects mediated by transcript and protein levels</td>
<td>Observed age-varying genetic relationships for genetic variants commonly used in MR studies and implications for bias in Mendelian randomization estimates of lifetime effects</td>
<td>Using Mendelian randomization to provide new insights into the impact of sleep disturbance on diabetes</td>
</tr>
<tr>
<td><strong>10.15-10.30hr</strong></td>
<td><strong>15.30-15.45hr</strong></td>
<td><strong>15.30-15.45hr</strong></td>
</tr>
<tr>
<td>Wes Spiller (Bristol, UK)</td>
<td>Carolina Borges (Bristol, UK)</td>
<td>Ralph Moeller Trane (UW-Madison, US)</td>
</tr>
<tr>
<td>Estimating and visualising multivariable Mendelian randomization analyses within a radial framework</td>
<td>Challenges in conducting and interpreting Mendelian randomization studies testing the effect of molecular traits: the example of polyunsaturated fatty acids and cardiovascular diseases risk</td>
<td>Nonparametric Bounds in Two-Sample Summary-Data Mendelian Randomization: Some Cautionary Tales for Practice.</td>
</tr>
<tr>
<td><strong>10.30-10.45hr</strong></td>
<td><strong>15.45-16.00hr</strong></td>
<td><strong>15.45-16.00hr</strong></td>
</tr>
<tr>
<td>Jiao Luo (Leiden, Netherlands)</td>
<td>Fernando Hartwig (Pelotas, Brazil)</td>
<td>Linbo Wang (Toronto, Canada)</td>
</tr>
<tr>
<td>Depression &amp; Inflammatory Bowel Disease: Bidirectional two-sample MR</td>
<td>Generalized definition of ACE for binary &amp; continuous treatments</td>
<td>Fighting Noise with Noise: Mendelian Randomization with Pseudo Variables</td>
</tr>
</tbody>
</table>

*5 or 6 talks per session
- Each no more than 10-minute talk + 5 minutes discussion specific to talk.
- There is time throughout the program for participants to arrange to ‘meet’ virtually and discuss topics and this includes ‘meeting’ with speakers of these talks for further discussion.

We encourage speakers to have no more than 10 slides.
Chairs will be strict with time-keeping.
Abstracts for all talks are presented below.
Wang Miao (Peking University)

**Identifying effects of multiple treatments in the presence of unmeasured confounding**

In this talk, I will discuss two proposed approaches for confounding adjustment with multiple treatments. The auxiliary variables approach rests on an additional instrumental or confounder proxy variable, and the null treatments approach assumes that only an unknown subset of treatments are active. We provide sufficient conditions for identification, establish estimation, and illustrate with a data example where the goal is to jointly analyze the genotype, gene expression, and clinical data to identify important genes related to body weight. SNPs are used as IVs to make inference about the effects of multiple genes expressions on mouse obesity.

Fatima Batool (Cambridge, UK)

**Dimensionality Reduction Methods to infer accurate causal risk factors using MVMR**

In cases where Mendelian randomisation analysis is conducted using genetic variants from single gene region, it is not advised to apply usual MR methods due to high correlation among available genetic variants. Methods using pruning has been proposed in literature for the treatments of correlated instruments, with a caveat that selection of variants using pruning leads to causal estimates that are highly sensitive to the choice of variants selected resulting in somewhat fragile analysis. In this study, we have developed statistical dimensionality reduction methods to treat collinearity among variants. We have extended principal component analysis and factor analysis models for the multi-variable MR using various estimation techniques. We have applied methods to identify most likely risk factor among eotaxin-1, MCP-1, and MCP-3 for strokes using variants from chemokine gene cluster.

Eleanor Sanderson (Bristol, UK)

**Estimation of causal effects of an exposure at multiple time points through Multivariable Mendelian randomization.**

Mendelian Randomisation (MR) is a powerful tool in epidemiology which can be used to estimate the causal effect of an exposure on an outcome in the presence of unobserved confounding, by utilising genetic variants as instrumental variables (IVs) for the exposure. The effects obtained from MR studies are often interpreted as the lifetime effect of the exposure in question. However, the causal effects of many exposures are thought to vary throughout an individual’s lifetime and there may be periods that are more important for a particular outcome. Multivariable MR (MVMR) allows for multiple, potentially highly related, exposures to be included in an MR estimation. We explore the use of MVMR to estimate the effect of a single exposure at different time points in an individual’s lifetime on an outcome. We show that the effect of different time periods can be estimated through MVMR when the association between the genetic variants used as instruments and those time periods varies. We show that the effect estimated of each time periods obtained through this estimation is the effect of increasing the exposure trajectory by a unit at the time point considered conditional on the genetic predicted level of exposure at the other time points included in the estimation. This may include effects of the exposure through time periods that have not been included in the estimation but that are associated with the genetically predicted trajectory for the exposure. We illustrate the method with an application to the estimation of the effect of early and later life BMI on anorexia and smoking behaviour.

Marie Sadler (Swiss Institute of Bioinformatics)

**Quantifying DNA methylation-to-complex trait effects mediated by transcript and protein levels**

High-dimensional omics datasets provide valuable resources to determine the causal role of molecular traits in mediating the path from genotype to phenotype. Making use of quantitative trait loci (QTL) and
genotype-wide association study (GWAS) summary statistics, we developed a three-sample multivariable Mendelian randomization (3S-MVMR) framework to estimate the proportion of DNAm-to-trait causal effects that is mediated through the cis-regulation of transcript and protein levels. Applying our method on a genome-wide scale to 50 complex traits, we found that on average 37.8% (95% CI: [36.0%-39.5%]) of DNAm-to-trait effects were mediated through transcripts in the cis-region, while only 15.8% (95% CI: [11.9%-19.6%]) are mediated through proteins in cis. DNAm sites typically regulate multiple transcripts, and decrease gene expression for only 53.4% of the ~47,000 significant DNAm-transcript pairs. Notable differences in the transcript and protein QTL architectures were detected with only 22% of protein levels being significantly causally driven by their corresponding transcript levels. Results revealed several regulatory mechanisms, e.g. DNAm cg10385390 (chr1:8’022’505) increases the risk of irritable bowel disease by reducing PARK7 transcript and protein expression. Additionally, DNAm at cg09070378 (chr1:161’183’762) was found to decrease the risk of asthma by reducing FCER1G expression, a gene part of the KEGG pathway for asthma. The proposed integrative framework identified putative causal chains through omics layers providing a powerful tool to map GWAS signals while also indicating that molecular mechanisms can be more complex than what the central dogma of biology would suggest.

Jiao Luo (Leiden)

Depression and Inflammatory Bowel Disease: A Bidirectional two-sample Mendelian Randomization Study

Background and Aims: Observational studies have suggested a bidirectional association between depression and inflammatory bowel disease (IBD), including Crohn’s disease (CD) and ulcerative colitis (UC). However, it remains unclear whether observed associations are casual due to the difficulties of determining sequential temporality. We investigated the association between depression and IBD by using bidirectional two-sample Mendelian randomization (MR).
Methods: Independent genetic variants for depression and IBD were selected as instruments from published genome-wide association studies (GWAS) among individuals of predominantly European ancestry. Summary statistics for instrument-outcome associations were retrieved from three separate databases for both depression (Psychiatric Genomics Consortium, FinnGen, and UK Biobank), and IBD (the largest GWAS meta-analysis, FinnGen, and UK Biobank), respectively. MR analyses included inverse-variance weighted method, weighted-median estimator, MR-Egger regression, and sensitivity analyses of Steiger filtering and MR PRESSO. From either direction, analyses were performed per outcome database and were subsequently meta-analyzed using fixed-effect model.

Results: Genetically predicted depression (per log-odds ratio increase) was associated with a higher risk of IBD; odds ratios (95% confidence interval) for IBD, CD and UC were 1.20 (1.05, 1.36), 1.29 (1.07, 1.56) and 1.22 (1.01, 1.47) in a combined sample size of 693,183 (36,507 IBD cases), 212,172 (13,714 CD cases) and 219,686 (15,691 UC cases) individuals, respectively. In contrast, no association was observed between genetically influenced IBD and depression in 534,635 individuals (71,466 depression cases).

Conclusions: Our findings corroborated a causal association of depression on IBD, which may impact the clinical decision on the management of depression in patients with IBD. Though our results did not support a causal effect of IBD on depression, further investigations are needed to clarify the effect of IBD activity on depression (with different symptomology).

Richard Gou (Cambridge, UK)
Automated analysis of discrete IV with finite-sample guarantees

We propose a framework for the analysis of discrete instrumental variable models based on convex optimization and a finite-sample bound on the multinomial likelihood ratio test. The framework "automates" the often laborious workflow from (partial) identification to statistical inference. We illustrate the method with the Vietnam draft lottery data. Using the lottery number as an instrument, we study the effect of military service on annual earnings.

Amy Mason (Cambridge, UK)
Breaking down the barriers to non-linear Mendelian randomization

Availability of summarised data has democratised genetic epidemiology by enabling researchers from around the world to perform analyses using world-leading datasets. This has facilitated two/three-sample Mendelian Randomization with large sample sizes, and made analyses easy to replicate. However, summary methods rely on an assumption of a linear effect of the exposure on the disease and Mendelian randomization methods can estimate non-linear exposure—outcome relationships typically require individual-level data. This talk will discuss a new package for implementing fractional polynomial and piecewise-linear methods on stratified summarised data that can either be estimated from individual-level data using the package or supplied by a collaborator. This method facilitating analyses where individual-level data cannot easily be shared, and additionally increasing reproducibility as summarised data can be reported. This talk will demonstrate the method’s viability for a range of potential exposure-outcome relationships via simulated data and show the method in a practical setting in UK Biobank data, evaluating the evidence for non-linear effects of LDL-cholesterol on Coronary Heart Disease. I will also discuss potential extensions to more dimensions and how the initial phase of the analysis could be performed blindly on a central server (a so-called “walled garden”), allowing greater accessibility to large data sources with access to obtain stratified summarised data and non-linear analyses given to a wider set of users.

Matt Tudball (Bristol, UK)
Almost exact Mendelian randomization
Mendelian randomization (MR) explores the causal effect of an exposure on an outcome via the transmission from parents to offspring of exposure-modifying genetic variants. MR is therefore best justified in a within-family design with parent-offspring pairs or trios, which is increasingly feasible due to growing data availability. Existing methods for within-family MR rely on specifying phenotype models for the exposure and outcome, leaving the role of meiosis implicit.

However, meiosis has been thoroughly studied and modelled in genetics dating back to Haldane (1919). We propose a statistical framework that enables meiosis models to be used as the “reasoned basis for inference” in MR. Specifically, we develop an approach to inference based on exact hypothesis testing, first described in Fisher’s original proposal for randomized experiments (Fisher, 1935). We therefore make explicit the common analogy between MR and a randomized controlled trial. Furthermore, we develop a rigorous graphical framework for describing within-family MR, which we use to identify sufficient confounder adjustment sets.

In addition to the conceptual advantages, our randomization-based inference has several practical advantages too. First, it sidesteps the need for correctly specifying phenotype models, although a better model will often lead to more powerful tests. We demonstrate via simulation that propensity scores obtained from the underlying meiosis model can form powerful test statistics. Second, our approach is robust to arbitrarily weak instruments. Finally, by using our sufficient adjustment sets, it is robust to biases arising from population structure, assortative mating, dynastic effects and several forms of pleiotropy.

Jeremy Labrecque (Erasmus MC, Netherland)

**Observed age-varying genetic relationships for genetic variants commonly used in MR studies and implications for bias in Mendelian randomization estimates of lifetime effects**

Background: Estimates from conventional Mendelian randomization (MR) analyses can be biased when the genetic variants proposed as instruments vary over age in their relationship with the exposure.

Methods: For four exposures commonly studied using MR, body mass index (BMI), alcohol consumption, C-reactive protein (CRP) and low density lipoprotein cholesterol (LDL), we assessed the degree to which their relationship with genetic variants commonly used as instruments varies by age using flexible, spline-based models in UK Biobank data. Using these models, we then estimated how biased MR estimates would be due to age-varying relationships using plasmode simulations assuming different exposure windows. In separate simulations, we examined the degree to which multivariate MR could recover the lifetime effect when genetic variants have time-varying relationships with the exposures.

Results: We found that most genetic variants had age-varying relationships with the exposure for which they are a proposed instrument. Body mass index and LDL cholesterol had the most variation while alcohol consumption had very little. This variation over age led to small potential biases when estimating the effect of alcohol consumption and C-reactive protein and large potential biases when estimating the effect of BMI and LDL cholesterol. Biases were larger with longer exposure windows and when assuming a Gaussian exposure window rather than a linear exposure window. We also found that when the exposure is only measured at two time points, multivariate MR remains largely biased for estimates of the lifetime effect.

Conclusions: Time-varying genetic relationships can, for some genetic variant-exposure pairs, lead to large biases in estimates using Mendelian randomization. Future Mendelian randomization studies should check for this and, when present, carefully interpret results with potential bias in mind.

Carolina Borges (Bristol, UK)

**Challenges in conducting and interpreting Mendelian randomization studies testing the effect of molecular traits: the example of polyunsaturated fatty acids and cardiovascular diseases risk.**
Background: Despite early interest in the cardiovascular effects of polyunsaturated fatty acids (PUFA), the evidence linking PUFA to cardiovascular diseases (CVDs) is still controversial. Genetic variants regulating fatty acid desaturases genes (e.g. FADS1), coding for rate-limiting enzymes in PUFA biosynthesis (e.g. D5D), can be used as causal anchors to investigate the involvement of PUFA in CVD aetiology.

Aims: We used Mendelian randomization (MR) to explore the effect of higher D5D activity on a wide range of CVDs in up to 1,153,768 European ancestry individuals. In addition, we explored three key scenarios that could lead to spurious MR findings (i.e. horizontal pleiotropy, population structure, and selection bias).

Methods: We used summary-data MR to investigate the effect of higher D5D activity (proxied by the ratio of arachidonic acid to dihomo-γ-linoleic acid) on CVDs risk using rs174546 as the genetic instrument. To assess the plausibility of bias by horizontal pleiotropy, we used genetic colocalization and multivariable Mendelian randomization jointly modelling the expression of multiple genes within the FADS1 locus. To assess confounding by residual population structure, we compared the association of rs174546 with established CVDs risk factors (i.e. LDL-cholesterol, triglycerides, systolic blood pressure, glycated haemoglobin, smoking, and body mass index) between unrelated individuals and within-siblings (up to 68,691 sibships). To explore selection bias, we carried out a positive control Mendelian randomization analyses of established risk factors on CVDs risk.

Results: Our main findings suggest that higher D5D activity is related to higher risk of coronary artery disease, ischemic stroke, heart failure, atrial fibrillation, peripheral artery disease, venous thromboembolism, and aortic valve stenosis. Multivariable MR confirmed that main findings were driven by higher FADS1 expression (rather than by the expression of other genes in the region) in multiple tissues and genetic colocalization pointed out a shared genetic variant between genetic signals for D5D activity and LDL-cholesterol. The relation between rs174546 and CVDs risk factors was broadly similar when comparing unrelated individuals and siblings. In the positive control analyses, we observed the expected effect of risk factors on CVDs risk.

Conclusions: MR findings indicate that lifelong exposure to higher D5D activity is related to higher risk of several CVDs among Europeans. Sensitivity analyses support this interpretation and indicate LDL-cholesterol as a potential mediating trait between PUFA biosynthesis and CVDs risk.

Fernando Hartwig (Pelotas, Brazil)

A generalized definition of the average causal effect for both binary and continuous treatments

One of the main tasks of causal inference is estimating well-defined causal parameters. One of the main causal parameters is the average causal effect (ACE) – the expected value of the individual level causal effects in the target population. For binary treatments, the individual level causal effect is defined as contrast between potential outcomes. For continuous outcomes, however, there are many such contrasts in finite samples, thus hampering their use for summarizing the "overall" causal relationship. I will present a generalized version of the ACE, where individual level causal effects are defined as the partial derivative (with respect to the treatment) of the individual level causal dose-response function evaluated at treatment value that the individual has. This definition is equivalent to the conventional definition for binary treatments, but also incorporates continuous treatments.

Lizzie Diemer (Harvard, US)

Falsification of a Mendelian randomization analysis of the effect of maternal alcohol consumption during pregnancy on offspring ADHD symptoms

Background: Like all causal inference methods, Mendelian randomization (MR) requires unverifiable assumptions. However, the MR model implies a set of inequalities that can be used to falsify, but not
verify, such assumptions. While these inequalities are rarely used in practice, their potential utility in the context of MR analyses with multiple proposed genetic instruments remains underexplored.

Aim: To explore whether the instrumental inequalities could be used to falsify proposed genetic instruments for the effect of maternal alcohol consumption during pregnancy on offspring ADHD.

Methods: We applied the instrumental inequalities to models proposing 11 maternal genetic variants as instruments for the effect of maternal alcohol consumption during pregnancy on offspring ADHD symptoms in the Norwegian Mother, Father, and Child Study (MoBa) and the Avon Longitudinal Study of Parents and Children (ALSPAC).

Results: When comparing any to no alcohol consumption in ALSPAC, the instrumental inequalities held (meaning we did not falsify the model) for all genetic variants individually, some combinations of 2-4 SNPs jointly, and no combinations of 5 or more genetic variants. In MoBa, the instrumental inequalities held for only 8 of the 11 genetic variants individually, 9 combinations of 2 genetic variants, and no combinations of 3 or more genetic variants.

Conclusions: The instrumental inequalities detected violations of the MR model for at least 6 of the proposed genetic instruments in ALSPAC, and at least 8 of the proposed genetic instruments in MoBa. From a methodologic standpoint, this suggests the instrumental inequalities may be more useful for falsification in planned MR analyses than was previously understood. Substantively, these findings also indicate a need for further investigation of the relationship between maternal genetic variants related to alcohol use and offspring psychiatric health.

Zijun Gao (Stanford, US)

Discovering biological pathways with genome-phenome summary data

In the big data era, there is an increasing scale of genome-phenome data available for understanding latent biological mechanisms. However, there are two challenges in uncovering biological pathways: (1) due to privacy concerns, the access to individual-level data is often restricted and only the summary statistics are publicly available; (2) phenotypes are influenced not only by genetic mediators but also by environmental factors. In this project, we develop a method for exploring biological pathways using genome-phenome summary data. In particular, we propose a procedure using signal and noise SNPs to disentangle unobserved genomic and environmental factors, extract genetic components, and form clusters of genotypes and phenotypes corresponding to biological pathways. We then propose a bootstrap aggregation (bagging) approach to increase the procedure's stability and reproducibility. We apply the method to a Metabolomics dataset and the UK Biobank database and find phenotype-genotype clusters suggesting new biological pathways.

Ting Ye (University of Washington, US)

Robust Mendelian randomization in the presence of many weak instruments and widespread horizontal pleiotropy

Mendelian randomization (MR) has become a popular approach to study the effect of a modifiable exposure on an outcome by using genetic variants as instrumental variables (IVs). Two distinct challenges persist in MR: (i) each genetic variant explains a relatively small proportion of variance in the exposure and there are many such variants, a setting known as many weak IVs; and (ii) many genetic variants may have direct effects on the outcome not through the exposure, or in genetic terms, when there exists widespread horizontal pleiotropy. To address these two challenges simultaneously, we propose two novel estimators, the debiased inverse-variance weighted (dIVW) estimator for summary-data MR and the GENIUS-MAWII estimator for individual-data MR, and we establish their statistical properties. We conclude by demonstrating these two methods in simulated and real datasets.

Rebecca Richmond (Bristol, UK)
Using Mendelian randomization to provide new insights into the impact of sleep disturbance on diabetes

Sleep disturbance, characterized by abnormal sleep duration, quality or timing, is common in modern society and has been linked to poor glucose control. However, it is currently unclear whether sleep disturbance is causally related to hyperglycaemia and type 2 diabetes. We have applied Mendelian randomization to investigate this issue, primarily using the UK Biobank study, which includes data on sleep traits, HbA1c (glycated haemoglobin), glucose, as well as prevalent and incident diabetes. We have triangulated evidence across multivariable regression, one-sample and two-sample Mendelian randomization; applied a new ‘collider-correction’ method which allows bias from weak instruments and unbalance horizontal pleiotropy to be addressed in a one-sample setting; compared estimates based on self-reported and accelerometer-derived sleep traits; investigated complementary approaches including Genomic Structural Equation Modeling to assess multivariate genetic associations among sleep and glycaemic traits; and performed non-linear MR methods to investigate the potential non-linear relationships with sleep duration. We have found that self-reported insomnia has a consistent adverse effect on HbA1c levels, but no robust effects of other self-reported or accelerometer-derived traits. While there was some evidence for a J-shaped relationship between self-reported sleep duration and HbA1c, evidence supporting a nonlinear association between accelerometer-derived sleep duration and HbA1c was not clear.

Ralph Moeller Trane (UW-Madison, US)
Nonparametric Bounds in Two-Sample Summary-Data Mendelian Randomization: Some Cautionary Tales for Practice

Recently, in genetic epidemiology, Mendelian randomization (MR) has become a popular approach to estimate causal exposure effects by using single nucleotide polymorphisms from genome-wide association studies (GWAS) as instruments. The most popular type of MR study, a two-sample summary-data MR study, relies on having summary statistics from two independent GWAS and using parametric methods for estimation. However, little is understood about using a nonparametric bound-based analysis, a popular approach in traditional instrumental variables frameworks, to study causal effects in two-sample MR. In this work, we explore using a nonparametric, bound-based analysis in two-sample MR studies, focusing primarily on implications for practice. We also propose a framework to assess how likely one can obtain more informative bounds if we used a different MR design, notably a one-sample MR design. We conclude by demonstrating our findings through two real data analyses concerning the causal effect of smoking on lung cancer and the causal effect of high cholesterol on heart attacks. Overall, our results suggest that while a bound-based analysis may be appealing due to its nonparametric nature, it is far more conservative in two-sample settings than in one-sample settings to get informative bounds on the causal exposure effect.

Linbo Wang (Toronto, Canada)
Fighting Noise with Noise: Mendelian Randomization with Pseudo Variables

Mendelian randomization (MR) is a method by which genetic variants are leveraged as instrumental variables (IV) to investigate causal relationships between a modifiable exposure and a clinically relevant outcome. The application of standard IV methods in MR analysis faces two main challenges. First, most genetic variants are not relevant for studying a particular exposure of interest. Second, not all relevant genetic variants may serve as valid instruments due to pleiotropy, a phenomenon that one genetic variant may influence many seemingly unrelated traits. In this article, we propose a data-driven method for Mendelian randomization that addresses these two challenges simultaneously. A key component of our proposal is a novel resampling method that construct pseudo variables to address false positives. Such
issue commonly arises from the application of standard MR methods to ultra-high dimensional genetic datasets. Theoretical and synthetic data analyses show that the proposed method performs favorably compared to existing MR methods. We illustrate our approach through estimating the causal effect of obesity on health-related quality of life using data from the Wisconsin Longitudinal Study.