Dissertation Discussion Outline

**Part 1: EWAS and confounder adjustment methods**

1. Summary of EWAS findings & difficulties of confounder adjustment in EWAS (short paragraph)

1. we did not find evidence of associations between maternal mid-pregnancy vitamin d sufficiency and offspring DNA methylation at any measured site in cord blood.
2. this EWAS has less difficulties than others, because can be reasonably confident that are not simply finding reverse causation
3. However, difficult to determine confounding structure, because determinants of methylation are not fully known, and like other pregnancy outcomes, may be affected by an individual’s own tendency towards health promoting behaviors and stress levels.
4. Some researchers have proposed use of MR as an alternative means the avoid this issue (CITE Relton 2012 - 2 step MR paper, Felix & Cecil 2019 - Population DNA methylation studies in the Developmental Origins of Health and Disease (DOHaD) framework)

**Part 2 : MR advantages, limitations, falsification and bounding approaches**

1. Transition: Brief recap of advantages of MR

1. allows for valid estimation of causal effects even in the presence of unmeasured confounding
2. especially compelling for prenatal contexts, as confounders are often complex, difficult to measure, or (for EWAS), potentially unknown
3. prenatal period can more reasonably be considered a fixed intervention, as offspring are theoretically directly exposed to maternal genotype only for 9 months
4. However, prenatal MR is a somewhat unique context, and may be subject to unique biases

2. MR Lit Review Summary of Findings

1. We found that researchers rarely discuss issues specific to the prenatal MR context, including selection on pregnancy, pleiotropy via offspring genotype, of pleiotropy via pre- or postconceptional exposure status.
2. Although the majority of prenatal MR studies present point estimates, the additional assumptions necessary for point estimation are rarely discussed. This is especially concerning in the context of prenatal MR, where certain estimands are defined differently, and there is biologic evidence that the point estimating conditions cannot hold for certain exposures of interest.
3. No studies applied the instrumental inequalities, falsification in general was rarely applied

3. Instrumental Inequalities Summary of Findings

* 1. brief reminder of what instrumental inequalities are/what they can do
  2. We found that, within our sample, the MR assumptions were violated for at least half of the 4 SNPs proposed as instruments. This provides a clear example of how, in the context of multiple proposed instruments, the instrumental inequalities can be used to detect violations of the instrumental conditions.
  3. Our simulations suggest that the inequalities will be increasingly violated as the magnitude of violations grow, and are more sensitive as larger numbers of variables are proposed as joint instruments.

4. Transition: The instrumental inequalities are able to falsify the first 3 assumptions alone. However, in order to obtain a point estimate, MR studies must make one of a set of possible homogeneity assumptions. In the context of prenatal MR, these homogeneity assumptions are often implausible. However, although historically unpopular, bounding approaches may allow for identification of the direction of effect and information about the magnitude of the effect under the 3 primary MR conditions alone. We set out to investigate whether bounds can be usefully applied in the context of MR with multiple proposed instruments.

5. Summary of MR Bounds paper

1. We found that, when single SNPs were proposed as instruments, bounds on the average causal effect barely improved on the assumption free bounds. However, when larger numbers of SNPs were proposed as joint variants, the bounds narrowed and were sometimes able to identify the direction of effect (though bounds from different subsets of SNPs identified different directions of effect).
2. In addition, the variation we observed across bounds calculated using different sets of proposed instruments highlights the utility of the bounds for comparing and critically evaluating assumption sets used in MR.

i. in our results, bounds computed using different sets of proposed joint instruments sometimes identified opposite directions of effect

6. Transition: Within our own study, we conducted analyses within two different cohorts. Whenever data from multiple cohorts are available, it would be useful to combine information across cohorts. However, typical meta-analytic techniques do not translate easily to the context of bounds.

7. Summary of Pooled Bounds Paper

1. We show that bounds computed within different study populations can be pooled using the set intersection methods described by Manski 2020. However, in addition to the MR conditions, these pooled bounds require an assumption of consistency of potential outcomes across study populations, and an assumption of no effect modification by study population.
2. Bounds can be computed assuming that the identifiability conditions hold within each study, or assuming that the identifiability conditions hold in at least some number of studies (though this second approach may only be plausible in very specific settings).
3. While these assumptions are strong, they are consistent with the assumptions required in meta-analyses that are explicitly focused on estimating an effect within a population of individuals, rather than a population of studies

**Part 3: General Synthesis**

***Broader Implications***

1. Greater attention needs to be paid to the validity of the MR conditions in prenatal MR studies, especially for offspring behavioral outcomes
2. prenatal MR is subject to more forms of bias than previous articles using this method have made clear (MR lit review)
3. the fact that there were violations of the instrumental inequalities for SNPs whose relationship to the exposure is relatively well understood in multiple cohorts, suggests that either pleiotropic effects of these snps are stronger than was previously understood, or that the prenatal MR model is being affected by more severe structural biases (selection on pregnancy, pre- or postnatal exposure status, assortative mating)

i. Note limitation: inequalities are study specific and confidence intervals/statistical inference have not been developed outside of the all binary case. So presence of violations within a single cohort is not necessarily damning for all cases. However, violations were detected for multiple exposures, many SNPs, in multiple cohorts. Suggests we were extremely unlucky or there are more issues at play.

1. this second possibility is a concerning indictment of prenatal MR more broadly, as these researchers often describe these studies as being superior to other observational designs, and sometimes encourage their use in policy determinations
2. further research is needed to evaluate the impact of some biases in applied prenatal MR (selection on pregnancy, assortative mating)
3. researchers need to be made aware of the unique issues of prenatal MR, and should discuss these limitations in their work
4. The instrumental inequalities and IV bounding approaches should be more broadly applied in the IV studies, especially MR with multiple proposed instruments
5. these studies provide multiple concrete examples of cases where the instrumental inequalities were able to provide meaningful information (either clearly falsifying a model or identifying a direction of effect), without additional homogeneity assumptions or other potentially implausible assumptions (e.g., the inside assumption)
6. the bounds in particular also provide an opportunity, in both single studies and meta-analyses, to contextualize point estimates
   1. could give readers a clearer sense of how much information is the data alone
7. on a more meta level, these results argue for the importance of considering existing methods based on minimal assumptions, and trying not to reinvent the wheel
   * 1. recent years have seen an explosion in sensitivity analyses and robust methods for MR
     2. however, these methods typically take some form of homogeneity for granted, although biologic knowledge suggests this homogeneity is implausible for many exposures of interest and proposed instruments
     3. neither the bounds nor the instrumental inequalities require such homogeneity assumptions, and were originally described long before MR became broadly popular
     4. while innovation is vital to science, these results suggest that useful methods incorporating minimal assumptions sometimes are dropped by the wayside

3. Heterogeneity of effects of prenatal exposures is too often overlooked. The presence of such heterogeneity complicates the use of MR, but is also critically important to interpreting results, translating those results into recommended practice, and designing studies that answer the questions that are most important to us.

* 1. definition is sometimes more complex (MR lit review)
  2. conclusions of MR may be strongly reliant on homogeneity (bounds paper), though there are many cases where this is known not to hold because of biological knowledge, and is often implausible
  3. effects may also be heterogenous across populations – as most prenatal MR studies are conducted within European ancestry populations, it is highly likely that such effects (and potentially the validity of the MR model itself) may be heterogenous across populations (issues with both pooled bounds and conduct of MR when trying to avoid population stratification)
  4. ignoring heterogeneity also limits the extent to which the results of these types of analyses can truly be generalized – results from this type of MR study may not translate easily to other populations. Basing universal recommendations on these results could be harmful, and ignores important potential differences between populations.
     1. in the context of the use of MR to understand possible biologic relationships, it also limits our understanding of the mechanisms at play. If effects are stronger/weaker/only present in certain subgroups, this could indicate that particular mechanisms of action are more important.
     2. In the context of basing policy recommendations or medical recommendations on this approach, it is potentially damaging. Effects could be highly heterogenous across subsets of the population. Further, because of the nature of prospective cohorts recruited based on the presence of a pregnancy and strategies used to deal with potential bias related to population stratification, the populations to which MR estimates from these studies apply may be very different from the population at large. Thus effects will not necessarily translate to other populations easily.
  5. more attention should be paid to this issue, both when constructing cohorts, conducting analyses, and discussing results.

***Future Directions***

1. Intro paragraph: Overall, these findings have both substantive and methodologic implications, and suggest a need for further research in several areas. In particular, the results shown in this dissertation indicate a need for further applied research on the relationship between maternal genetic variants and offspring psychiatric outcomes, and how these might bias MR estimates. Our findings also highlight the need for further methods development in several areas of MR. Beyond these research directions, a critical aspect of future work in this area will be making the instrumental inequalities and instrumental bounds accessible to a broader audience.

*Applied Genetic and Prenatal Epidemiology*

1. Further research needed on selection on pregnancy and assortative mating
2. in multiple analyses of different exposure-outcome relationships, in different cohorts, we found strong evidence of violations of the MR conditions for prenatal exposures, including for SNPs with relatively well understood relationships to the exposure of interest
3. maybe look into whether there is existing evidence of pleiotropy through more common sensitivity analyses for these
4. almost no attention has been paid to selection on pregnancy and assortative mating, although these biases have the potential to powerfully impact the results of MR analyses, and are note easily detected or ameliorated through commonly used robust methods
   * 1. selection on pregnancy violates the INSIDE assumption and would affect all SNPs (thus violating the requirements for weighted median regression)
     2. assortative mating would likely similarly affect all SNPs
5. both require somewhat unique samples, in that they require samples that include information on pregnancy exposures, parental genotypes, and offspring outcomes in study populations that were not selected on the basis of pregnancy

*Methods development*

1. Need for development of methods that account for sampling variability/uncertainty (95% confidence intervals and hypothesis testing) for bounds and inequalities
2. Considerable debate on how to generate confidence intervals or conduct statistical inference for bounded parameters
   * 1. briefly summarize swanson 2018/tamer 2010 on this topic, or just reference
3. importance of development of confidence intervals for bounds
   * 1. our samples are not infinite, and the ability to incorporate this source of uncertainty will be critical to using bounding approaches for decisionmaking
     2. development of 95% confidence intervals will also maximize comparability to point estimating approaches
4. complexity of confidence intervals and testing for instrumental inequalities
   * 1. random confounding and random violations of the instrumental inequalities
     2. in their existing state, still work as a sharp test within a sample
     3. however, development of confidence intervals could allow us to better differentiate between random violations and structural violations.
     4. This has some key advantages : first, identification of structural violations in one cohort provides evidence against the same study in similar populations (potentially minimizing replication studies that are doomed to fail because of structural violations of the MR conditions)
     5. Second: when pooling studies or conducting meta-analyses, determining whether a violation is structural or random would allow us to better understand whether an entire analysis should be eliminated (structural violation) or whether a single cohort should be eliminated from the analysis (random violation)

🡪 this pre-assumes that the violation is not a form of selection bias: this will need to be evaluated further

* + 1. third: structural violations of the instrumental inequalities can also provide information about the pleiotropic relationships of a particular SNP to multiple outcomes

🡪 Cai 2008: under the assumption that there are no common causes of the proposed instrument and outcome, the instrumental inequalities can themselves be used as bounds on the controlled direct effect of the proposed instrument on the outcome

🡪 Zaidi 2020: can also be used as bounds on the average causal effect within particular principal strata (under various assumptions)

🡪 Allows for identification of pleiotropic effects and thus clues as to genetic variant to phenotype pathway.

🡪 This may be especially useful when examining traits that have not been subject to large GWAS, and for which it may be difficult to use methods like LD regression to identify pleiotropic relationships (because of lack of secondary data)

1. Development of strategies to deal with heterogeneity of effects when pooling bounds across multiple cohorts
2. pooling bounds requires assumption of no effect modification of effect by study on relevant scale (because this requires that the distribution of effect modifiers is equal or at least balanced out across cohorts, it ends up being close to the assumption of no effect modification in distribution)
3. this assumption is likely unreasonable in most contexts
4. Could potentially use IP weighting/transportability methods developed by Dahabreh 2020 to weight distribution of effect modifiers in population to match either other populations or target population (requires individual level data on sample of target population)
   * 1. has not been extended to bounds context yet
     2. in the context of instrumental variables (especially mendelian randomization) this might be especially complex if the effect modifier is downstream of the proposed instrument
5. Development of methods that allow for time-varying exposures
6. pregnancy is somewhat unique case, where we might plausibly believe an individual is exposed to a constant level of an exposure for a fixed period of 9 months (the MR lit review shows this is not really the case, but deviates less from the truth than typical individual-level MR)
7. However, expansion of bounding and use of the instrumental inequalities into time-varying exposure cases will be key to use of these methods in the broader MR sphere
8. Some expansion of MR point estimation into time-varying case (Joy’s work)
9. Need for development of bounds for time-varying case
   * 1. Finkelstein/Shpitser 2020 developed general bounds that might serve as framework for this expansion.
     2. Either expanding their approach to cover the specific context of a vector of discrete exposures at different time points or as categorical variable with levels for every possible combination of exposure levels at each time point – and generating counterfactual contrasts for those comparisons
     3. Potential benefit of more specifically expanding their approach might be tighter bounds, as in the case of covariate inclusions. It would also allow researchers to evaluate how the bounds change depending on the assumptions about the relationship between genetic variants, exposures at different time points, and the outcome (e.g., if early life exposure only impacts the outcome through changes in later life exposure status, if genetic variants have no effect on early exposure status, but impact variation at later time points).
     4. Finkelstein shpitser 2020 could also allow for extended inequalities in setting of time-varying exposures (more explicit causal structure could provide sharper inequalities)
10. Alternatively, bounds based more on the MR-Genius framework, that allow for bounding of effect at specific time even if there are violations because of time-varying exposure status (tchetgen tchetgen)
11. Application of further inequalities and formalization/formal investigation of the “one bad apple” approach when comparing the instrumental inequalities across different subsets of proposed instruments
    1. The Balke-Pearl inequalities are sharp in the all-binary setting, but are not sharp when larger numbers of SNPs are proposed as joint instruments
    2. Bonet 2001 and Evans 2018 both describe but do not give further constraints implied by models with larger numbers of levels – but these are computationally difficult and do not have closed forms
    3. Finkelstein/Shpitser gives further inequalities for cases with multicategorical proposed instruments, exposures, and outcomes
       1. Further work is needed to explore the application of these additional inequalities to MR studies with multiple proposed instruments
12. Our relatively simple simulations suggested it may be possible to identify patterns in visualizations of the instrumental inequalities that are consistent with a single “bad apple” – one proposed instrument responsible for all the violations of the instrumental inequalities. However, this approach has not been formalized
13. Further work is needed to clarify in what contexts this is actually possible, and what the benefits and limitations of this approach are.
14. It is not fully clear if it is possible to use the values of the instrumental inequalities to compare relative degree of bias across models proposing different numbers of SNPs as joint instruments.
15. Cai 2008: instrumental inequalities can be considered bounds on the controlled direct effect of the proposed instrument on the outcome.
16. However, the Balke-Pearl inequalities are not necessarily sharp outside of the all binary case, and sharpness may differ when different numbers of SNPs are proposed as joint instruments.
17. Further, the current lack of confidence interval estimation procedures makes it more difficult to compare across settings with potentially very different numbers of participants in strata of the proposed joint instrument (give example of the small strata issue – imagine two strata with only one individual in each, and they happen to have bad values of x and y. For a binary outcome and exposure, that would give the maximum value of the inequalities (2), but may not translate to a larger magnitude of bias than “smaller” violations of the inequalities in settings with larger strata).
18. Further simulation studies may help to clarify the relationship of the instrumental inequalities to relative magnitude of bias across proposed joint instruments incorporating different numbers of SNPs.

*Teaching & Software Development for MR*

1. Expansion of teaching examples and educational articles to expand the use of the instrumental inequalities and bounding approaches in mendelian randomization and instrumental variables more broadly
2. both the inequalities and the bounds have existed for a long time
3. however, neither has been broadly used
4. in contrast, methods like MR-Egger have obtained a high degree of use in a relatively short period of time
5. This is likely due to both the accessibility of tools for these methods and the work of the original authors in expanding teaching about MR-Egger and related methods
6. One strength of this dissertation is the inclusion of adaptable R functions for the use of the inequalities and the bounds in the context of MR with multiple proposed instruments
7. further work is needed to improve the efficiency of these functions, especially in the context of large numbers of proposed instruments and large datasets (this will likely involve parallelizing the code and potentially incorporating other methods to improve the processing speed)
8. creation of an R package, creation of stata and sas code to implement the bounds and inequalities
9. While we took efforts to make this dissertation as approachable as possible, it likely isn’t reaching the largest audience of clinician-scientists and applied researchers interested in MR
10. to improve the reach of these methods, we need to increase the number of approachable educational articles on this topic in clinical journals with a wide reach, as well as articles that tend to focus on MR, such as international journal of epidemiology
11. work on expanding the use of these methods could also include the creation of teaching materials and toy datasets to allow for easy integration of these techniques into epidemiologic teaching
12. hopefully, one of the things that will expand the reach of these methods is further research into their properties, and how they behave in real data.
13. further simulation research on the inequalities in particular may help applied researchers feel more confident in a type of analysis they don’t fully understand.

**Part 4: Conclusions**

1. While the use of observational epidemiology is and should be focused on developing interventions to promote human health, it is inherently a risky endeavor, because it is based on unverifiable assumptions, and may give us the wrong answers. This is especially true in MR, an area where we have relatively little understanding of the mechanisms at play, and a weaker conceptualization of how violations of the required conditions translate into bias. Moreover, because of the relative makeup of genotyped cohorts and valid concerns regarding population stratification, MR studies are primarily conducted in white European ancestry populations, and it is unclear how such effects translate to other populations. To limit harm, it is critical that researchers incorporate a quality of humbleness into their work. Wider use of bounding approaches, falsification methods with the inequalities, and a clearer awareness of the impact of heterogeneity on analyses is a key step in this process. Adding falsification and bounding into standard MR practice would formalize that attitude, encouraging researchers to clarify the reliance of their conclusions on the assumptions of their model and to refocus attentions on the plausibility of model assumptions, rather than what findings are simply eye-catching.