Introduction to Genetic Epidemiology

Shea J Andrews, PhD

3/14/24

Table of contents

Part I

Introduction

This is a Quarto book.

To learn more about Quarto books visit https://quarto.org/docs/books.

Schedule

Welcome to this BMI Mini-Course on an Methods in Genetic Epidemiology. The course structure is divided into three key methodologies: genome-wide association studies, polygenic risk scores, and Mendelian randomization. Each trainee will apply these methods to a trait of their choosing and, at the course's conclusion, present their findings to the class.

Week 1 04/22 - 04/26

Monday 04/22

Alzheimer's disease

In this mini-course, we will frequently reference Alzheimer's disease to illustrate various genetic epidemiology methods. This session aims to introduce Alzheimer's disease and explore significant findings related to its genetic architecture.

Readings

- Knopman, D. S. et al. Alzheimer disease. Nat Rev Dis Primers 7, 33 (2021).
- Alzheimer's Association. 2023 Alzheimer's disease facts and figures. Alzheimer's Dementia (2023).
- Kornblith, E. et al. Association of Race and Ethnicity With Incidence of Dementia Among Older Adults. Jama 327, 1488–1495 (2022).
- Andrews, S. J. et al. The complex genetic architecture of Alzheimer's disease: novel insights and future directions. eBioMedicine 90, 104511 (2023).
- Andrews, S. J., Fulton-Howard, B. & Goate, A. Interpretation of risk loci from genome-wide association studies of Alzheimer's disease. Lancet Neurology 19, 326–335 (2020).

Wensday 04/24

Genome-Wide Association Studies

Genome-Wide Association Studies (GWAS) are foundational to various genetic analysis methodologies. In this session, we will delve into what a GWAS entails, the process of conducting one, and then engage in a hands-on exercise to carry out our own GWAS.

Readings

- Uffelmann, E. et al. Genome-wide association studies. Nat Rev Methods Primers 1, 59 (2021).
- Abdellaoui, A., Yengo, L., Verweij, K. J. H. & Visscher, P. M. 15 years of GWAS discovery: Realizing the promise. Am J Hum Genetics (2023)
- Marees, A. T. et al. A tutorial on conducting genome-wide association studies: Quality control and statistical analysis. Int J Method Psych 27, e1608 (2018).
- MacArthur, J. A. L. et al. Workshop proceedings: GWAS summary statistics standards and sharing. Cell Genom 1, 100004 (2021).

Tools

- PLINK[1]
- BCFTools[2]
- MungeSumStats[3]

Friday 04/26

Genetic Ancestry

Genetic ancestry explores the lineage and heritage inferred from our DNA, providing insights into population history and individual heritage. This session will introduce the concepts and methodologies used in determining genetic ancestry, emphasizing their importance in genetic epidemiology research.

Readings

- Lewis, A. C. F. et al. Getting genetic ancestry right for science and society. Science 376, 250–252 (2022).
- National Academies of Sciences, Engineering, and Medicine. 2023. Using Population Descriptors in Genetics and Genomics Research: A New Framework for an Evolving Field. Washington, DC: The National Academies Press.

Tools

- ADMIXTURE[4]
- RFmix[5]

Week 2 04/29 - 05/03

Monday 04/29

Heritability & Genetic Correlations

Heritability quantifies the proportion of phenotype variance attributable to genetic factors, whereas genetic correlations assess the extent of shared genetic architecture between traits. In this session, we will concentrate on the tools utilized to estimate these metrics from GWAS summary statistics.

Readings

- Rheenen, W. van, Peyrot, W. J., Schork, A. J., Lee, S. H. & Wray, N. R. Genetic correlations of polygenic disease traits: from theory to practice. Nat Rev Genetics 20, 567–581 (2019).
- Barry, C.-J. S. et al. How to estimate heritability: a guide for genetic epidemiologists. Int J Epidemiol (2022)

Tools

- LDSC[6]
- HDL[7]
- GenomicSEM[8]

Wensday 05/01

Polygenic Risk Scores I Polygenic risk scores (PRS) measure an invididueals total genetic liability for a trait. This session will cover the process of constructing a PRS and assessing its performance in predicting the trait.

Readings

- Choi, S. W., Mak, T. S.-H. & O'Reilly, P. F. Tutorial: a guide to performing polygenic risk score analyses. Nat Protoc 15, 2759–2772 (2020).
- Wand, H. et al. Improving reporting standards for polygenic scores in risk prediction studies. Nature 591, 211–219 (2021).
- Lennon, N. J. et al. Selection, optimization and validation of ten chronic disease polygenic risk scores for clinical implementation in diverse US populations. Nat. Med. 1–8 (2024)

Tools

- PRSice2[9]
- PRSet[10]

Friday 05/03

Polygenic Risk Scores II

The accuracy of polygenic risk scores (PRS) diminishes as the genetic distance from the training population increases. This session will explore cross-ancestry PRS methods designed to enhance PRS accuracy across diverse populations.

Readings

- Kachuri, L. et al. Principles and methods for transferring polygenic risk scores across global populations. Nat. Rev. Genet. 1–18 (2023) doi:10.1038/s41576-023-00637-2.
- Ding, Y. et al. Polygenic scoring accuracy varies across the genetic ancestry continuum. Nature 618, 774–781 (2023).

Tools

- PRS-CSx[11]

Week 3 05/06 - 05/10

Monday 05/06

Mendelian Randomization I

Mendelian Randomization (MR) is a method employed to identify causal risk factors for diseases. This session will cover the fundamentals of MR and demonstrate how to execute a two-sample MR analysis.

Readings

- Sanderson, E. et al. Mendelian randomization. Nat Rev Methods Primers 2, 6 (2022).
- Davies, N. M., Holmes, M. V. & Smith, G. D. Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. BMJ 362, k601 (2017).
- Hemani, G. et al. The MR-Base platform supports systematic causal inference across the human phenome. Elife 7, e34408 (2018).

Tools

- TwoSampleMR[12]

Wensday 05/08

Mendelian Randomization II

A crucial aspect of Mendelian Randomization (MR) studies is assessing whether the causal associations derived from MR analyses remain valid despite potential violations of MR's underlying assumptions. This session will focus on diagnostic and sensitivity analyses in MR, along with guidance on effectively reporting MR findings.

Readings

- Skrivankova, V. W. et al. Strengthening the Reporting of Observational Studies in Epidemiology Using Mendelian Randomization. JAMA 326, 1614–1621 (2021).
- Skrivankova, V. W. et al. Strengthening the reporting of observational studies in epidemiology using mendelian randomisation (STROBE-MR): explanation and elaboration. BMJ 375, n2233 (2021).

Friday 05/10

Trainee Presentations

Upon concluding this mini-course, trainees will showcase the results of their analyses.

Part II

Genome-wide Association Studies

This is a book created from markdown and executable code.

See[13] for additional discussion of literate programming.

1 Genome-wide Association Studies

This is a book created from markdown and executable code.

See[13] for additional discussion of literate programming.

Part III

Polygenic Risk Scores

This is a book created from markdown and executable code.

See[13] for additional discussion of literate programming.

2 Polygenic Risk Scores

This is a book created from markdown and executable code.

See[13] for additional discussion of literate programming.

Part IV

Mendelian Randomization

This is a book created from markdown and executable code.

See[13] for additional discussion of literate programming.

3 Mendelian Randomization

This is a book created from markdown and executable code.

See[13] for additional discussion of literate programming.

References

1. Purcell S, Neale B, Todd-Brown K, et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *The American Journal of Human Genetics*. 2007;81(3):559-575.
2. Danecek P, Bonfield JK, Liddle J, et al. Twelve years of SAMtools and BCFtools. *GigaScience*. 2021;10(2):giab008.
3. Murphy AE, Schilder BM, Skene NG. MungeSumstats: A Bioconductor package for the standardisation and quality control of many GWAS summary statistics. *Bioinformatics*. 2021;37(23):btab665-.
4. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*. 2009;19(9):1655-1664.
5. Maples BK, Gravel S, Kenny EE, Bustamante CD. RFMix: A Discriminative Modeling Approach for Rapid and Robust Local-Ancestry Inference. *The American Journal of Human Genetics*. 2013;93(2):278-288.
6. Bulik-Sullivan B, Finucane HK, Anttila V, et al. An atlas of genetic correlations across human diseases and traits. *Nature Genetics*. 2015;47(11):1236-1241.
7. Ning Z, Pawitan Y, Shen X. High-definition likelihood inference of genetic correlations across human complex traits. *Nature Genetics*. 2020;52(8):859-864.
8. Grotzinger AD, Rhemtulla M, Vlaming R de, et al. Genomic structural equation modelling provides insights into the multivariate genetic architecture of complex traits. *Nature human behaviour*. 2019;3(5):513-525.
9. Choi SW, O'Reilly PF. PRSice-2: Polygenic Risk Score software for biobank-scale data. *GigaScience*. 2019;8(7).
10. Choi SW, García-González J, Ruan Y, et al. PRSet: Pathway-based polygenic risk score analyses and software. *PLOS Genetics*. 2023;19(2):e1010624.
11. Ruan Y, Lin YF, Feng YCA, et al. Improving polygenic prediction in ancestrally diverse populations. *Nature Genetics*. 2022;54(5):573-580.
12. Hemani G, Zheng J, Elsworth B, et al. The MR-Base platform supports systematic causal inference across the human phenome. *eLife*. 2018;7:e34408.
13. Knuth DE. Literate programming. *Comput J*. 1984;27(2):97-111.