Review Article
http://mjiri.iums.ac.ir
Medical Journal of the Islamic Republic of Iran (MJIRI)

Med J Islam Repub Iran. 2024 (7 May);38.51. https://doi.org/10.47176/mjiri.38.51

Pleiotropic Bias and Study Design Considerations in Genetic Association Studies

Sana Eybpoosh[1]*, Seyyed Amir Yasin Ahmadi[2]

Received: 30 Mar 2023 Published: 7 May 2024

Abstract

Background: Case-control studies are efficient designs for investigating gene-disease associations. A discovery of genome-wide association studies (GWAS) is that many genetic variants are associated with multiple health outcomes and diseases, a phenomenon known as pleiotropy. We aimed to discuss about pleiotropic bias in genetic association studies.

Methods: The opinions of the researchers on the basis of the literature were presented as a critical review.

Results: Pleiotropic effect can bias the results of gene-disease association studies if they use individuals with pre-existing diseases as the control group, while the disease in cases and controls have shared genetic markers. The idea supports the conclusion that when the exposure of interest in a case-control study is a genetic marker, the use of controls from diseased cases that share similar genetic markers may increase the risk of pleiotropic effect. However, not manifesting the disease symptoms among controls at the time of recruitment does not guarantee that the individual will not develop the disease of interest in the future. Age-matched disease-free controls may be a better solution in similar situations. Different analytical techniques are also available that can be used to identify pleiotropic effects. Known pleiotropic effects can be searched from various online databases.

Conclusion: Pleiotropic effects may result in bias in genetic association studies. Suggestions consist of selecting healthy yet age-matched controls and considering diseases with independent genetic architecture. Checking the related databases is recommended before designing a study.

Keywords: Pleiotropy, Research Design, Observational Study, Genetic Epidemiology, Case-Control Studies, Genetic Association Studies

Conflicts of Interest: *None declared*
Funding: *None*

This work has been published under CC BY-NC-SA 1.0 license.
Copyright© Iran University of Medical Sciences

Cite this article as: Eybpoosh S, Yasin Ahmadi SA. Pleiotropic Bias and Study Design Considerations in Genetic Association Studies. *Med J Islam Repub Iran.* 2024 (7 May);38:51. https://doi.org/10.47176/mjiri.38.51

Introduction

Genome-wide association studies (GWAS) have revealed thousands of genome-wide significant associations with hundreds of complex human traits. The highly polygenic architecture of most diseases implies that the genetic part of the diseases is largely mediated through complex biological networks (1, 2). On the other hand, a notable discovery is that many genetic regions appear to contain variants that are associated with several seemingly unrelated traits. This phenomenon is called "pleiotropy". It refers to

Corresponding author: *Dr Sana Eybpoosh, s_eybpoosh@pasteur.ac.ir*

1. Research Centre for emerging and reemerging infectious diseases, Department of Epidemiology and Biostatistics, Pasteur institute of Iran, Tehran, Iran
2. Preventive Medicine and Public Health Research Center, Psychosocial Health Research Institute, Iran University of Medical Sciences, Tehran, Iran

↑What is "already known" in this topic:
Genome-wide association studies have unveiled numerous significant genetic associations with complex human traits, indicating their polygenic nature. Pleiotropy, where a single gene influences multiple traits or diseases, is evident. This phenomenon aids in comprehending disease pathways, potentially revolutionizing medical understanding.

→What this article adds:
Pleiotropic effects can introduce bias in gene-disease association studies, especially if the controls have pre-existing diseases. Awareness and strategies to mitigate this bias are crucial. Suggestions include selecting healthy yet age-matched controls and considering diseases with independent genetic architecture.

[Downloaded from mjiri.iums.ac.ir on 2026-11-10]

[DOI: 10.47176/mjiri.38.51]

circumstances in which a single gene or genetic marker influences two or more distinct traits or diseases (3). Therefore, a mutation in a pleiotropic gene may impact several traits simultaneously. Genes/variants that are linked with the marker of interest, also introduce the pleiotropic effect, because they are co-inherited (are in linkage disequilibrium [LD]) (4, 5). Pleiotropy may be horizontal or vertical. In horizontal pleiotropy, a single nucleotide polymorphism (SNP) has multiple phenotypes independently of the exposure of interest. In vertical pleiotropy, a SNP has multiple phenotypes on a causal pathway to the outcome. Besides knowledge about its associated diseases, pleiotropy can help in understanding the underlying biological pathways involved in disease onset and progression (6).

Pleiotropic effects on human diseases are very common. For example, several studies have reported shared genetic markers among immune-mediated diseases (7, 8), such as asthma, allergy, rheumatoid and psoriatic arthritis, and Crohn's disease. While some of these shared associations make immediate, intuitive sense (like diabetes-hypothyroidism (9)), others are more subtle. For example, multiple sclerosis has been identified to have shared genetic markers with the severity of SARS-CoV-2 infection (10) and type 1 diabetes mellitus (11). A human leukocyte antigen (HLA) variant was found to be associated with the slow advancement of human immunodeficiency virus (HIV) infection towards acquired immunodeficiency syndrome (AIDS) (12, 13), and interferon inducible transmembrane (IFITM) genes restrict the replication of several highly pathogenic human viruses, coronavirus, Marburg and Ebola viruses, influenza A viruses, dengue virus, and HIV-1 (14, 15). TRIM5α and TRIM22 are also suggested to be associated with the clinical course of HIV and COVID-19 (16, 17).

The concept of pleiotropy has important implications in epidemiological study design. Solovieff, et al (3), have provided a comprehensive review of the concept of pleiotropy. The authors have also outlined the analytical techniques utilized to identify the pleiotropic effect. Considering a gap in the literature regarding the design of genetic association studies in the presence of pleiotropic effects, the present review has examined how the selection of a control group from hospitals or diseased individuals may impact the results of these types of studies. Study design recommendations have been discussed to address this issue.

Pleiotropic Bias and Study Design Considerations

In case-control studies of gene-disease associations where hospital-based or individuals with pre-existing diseases are used as the control group, pleiotropic effect can bias the gene-disease association if the disease among cases and controls have shared genetic markers. This specific type of pleiotropic bias can be considered as a type of overmatching on genotype, and hence, may distort the magnitude of effect and shift the direction of association toward the null hypothesis. While some genetic associations may have large effect sizes (18), many others have small sizes (19, 20). So, such misclassifications may completely mask small-size associations. For example, the genetic overlap of type-one diabetes mellitus (T1DM) with some autoimmune disease and celiac disease can be mentioned. Therefore, it is important to remove the bias of autoimmune disease (21). For instance, celiac predisposing haplotypes of HLA (including DQ2 and DQ8) are prevalent in children with T1DM (22). Lack of considering celiac and autoimmune disorders that may be occurred in older ages in the control group, may result in smaller magnitude of effect between T1DM and the genetic markers.

Another scenario occurs when the genetic marker predisposes individuals to the disease in cases but has a protective effect on the disease in controls. Known as antagonistic pleiotropy, the direction of bias will be skewed away from the null hypothesis. For example, chromosomal regions 6p22-p24 are risk factors for schizophrenia and protective factors for higher relative fertility, or TNFRSF11B gene polymorphisms are risk factors for cancers and protective factors for bone density in females (23).

It is advisable for researchers to undertake a comprehensive assessment and acquire a thorough understanding of the contemporary pleiotropic effects for the study design and interpretation. However, the authors should also remain mindful of pleiotropic effects even in the absence of documented evidence, especially if they encounter effect sizes larger than expected or effect sizes that diverge from extant data.

One potential strategy to mitigate the risk of pleiotropic bias in gene-disease association studies is to select controls among either healthy individuals or individuals with disease(s) that are assumed, to the best of contemporary knowledge, to have independent genetic architecture compared to disease of interest in the case group. There are various resources that compile information on the pleiotropic effects of genes and mutations. Some examples are provided in Box 1, with a guide on how to find pleiotropic effects in each database.

When selecting the controls, it should be noted that not manifesting the disease symptoms at the time of recruitment does not guarantee that the individual will not develop the disease of interest. To account for this, age-matched disease-free individuals may be a better choice for genetic association studies because this better controls for pleiotropic effects.

Conclusions and future directions

In conclusion, pleiotropic effects can introduce bias in gene-disease association studies, especially when individuals with pre-existing diseases are selected as controls, such as in hospital-based case-control studies. The bias can lead to a shift in the magnitude and direction of the association between the genetic marker and the disease. Researchers need to be informed about the potential for pleiotropic bias and take steps to mitigate the risk of this bias during study design and interpretation. One potential recommendation would be the selection of healthy controls, especially among individuals that, based on their age, have had enough time to manifest the target disease in controls. Age-matched controls would be a remedy in this regard. Another solution might be the selection of controls among those with diseases that have independent genetic architecture compared to the disease of interest. No need to say that making such assumptions about underlying biological

[Downloaded from mjiri.iums.ac.ir on 2026-11-10]

[DOI: 10.47176/mjiri.38.51]

Box 1. Guide to available resources for identification of diseases with shared genetic markers (pleiotropic effect)

1. Online Mendelian Inheritance in Man (OMIM) database: OMIM is a comprehensive database that catalogs human genes and genetic disorders. By March 2023, the database contains all known mendelian disorders and more than 16,000 genes. One of the key features of the database is its annotation with information on the pleiotropic effects of genes and their mutations. To identify pleiotropic effects in the OMIM database, the specific genetic disorder or gene of interest should be searched for. The OMIM database will provide detailed information on the disorders, including any known pleiotropic effects. This information can be used to identify other symptoms or traits that may be associated with the disorder or gene of interest. Additionally, the OMIM database provides links to other resources, such as PubMed, NCBI Genes, MedlinePlus, GeneCard, and GeneReviews, which can provide further information on the underlying mechanism of pleiotropy in genetic disorders (24).
2. GeneCards database: GeneCards is an important resource for researchers and provide a wealth of information about genes' description and function, protein domains and structures, pathways, interactions, expression patterns and associated diseases and disorders. To identify pleiotropic effects in the GeneCards database, the gene of interest should be searched for. In the "function" section, the different biological pathways and molecular functions can be evaluated. Under the "pathway" section, the involvement of the gene in multiple pathways, as well as the tissues and cell types involved, and the role of the gene in different diseases can be identified. The "interaction" section provides information on how the gene interacts with other genes or proteins that are involved in different diseases. The database provides links to relevant publications and resources, such as PubMed, Ensembl, OMIM, and UniProt (25, 26).
3. The Human Phenotype Ontology (HPO) database: HPO is a standardized vocabulary of human phenotypic abnormalities and their related genes. It includes information on the pleiotropic effects of genes and their mutations.
4. Subject-specific online resources and databases: Depending on the research topic, there are other resources and databases that compile information on pleiotropic effects of genes and genetic variants in specific disease contexts. The researchers should search for, and use such resources, to complete their search. Some examples include DisGeNET (27), Databases of Genotypes and Phenotypes (dbGaP) (28), Genome Browser of the University of California Santa Cruz (Ucsc) (29, 30), etc.

Note: it is recommended to include a review of published literature in addition to the previously mentioned resources and database to complete the contemporary information on the pleiotropic effects of genes or variants.

mechanisms of diseases is not a trivial task, as pleiotropy is a complex phenomenon, and identifying the underlying genetic and biological mechanisms can be challenging.

The resources presented here for the identification of pleiotropic effects, may not provide a comprehensive list of all known effects, as new research is constantly uncovering new pleiotropic effects of genes and variants.

To further identify and characterize pleiotropic effects, the development of new analytical tools and methods that identify and account for pleiotropy is recommended. The integration of data sources and technologies should also be improved. Additionally, efforts should be made to expand and optimize the resources available for identifying and characterizing the pleiotropic effects, and to ensure that these resources are kept up-to-date. The user interface of these resources should be friendly to allow for a better listing of pleiotropic effects, especially by epidemiologists and researchers without a genetic background.

Authors Contributions

SE: designed and conceptualized the study, collected and interpreted the evidence, wrote and edited the manuscript draft; SAYA: collected and interpreted the evidence, wrote and edited the manuscript draft. Both authors approved the final version of the manuscript.

Conflict of Interests

The authors declare that they have no competing interests.

References

1. Uffelmann E, Huang QQ, Munung NS, De Vries J, Okada Y, Martin AR, et al. Genome-wide association studies. Nat Rev Methods Primers. 2021;1(1):59.
2. Eybpoosh S, Haghdoost AA, Mostafavi E, Bahrampour A, Azadmanesh K, Zolala F. Molecular epidemiology of infectious diseases. Electron Physician. 2017;9(8):5149.
3. Solovieff N, Cotsapas C, Lee PH, Purcell SM, Smoller JW. Pleiotropy in complex traits: challenges and strategies. Nat Rev Genet. 2013;14(7):483-95.
4. Eybpoosh S. Hardy Weinberg equilibrium testing and interpretation: focus on infection. J Med Microbiol Infect Dis. 2018;6(1):35-6.
5. Reich DE, Cargill M, Bolk S, Ireland J, Sabeti PC, Richter DJ, et al. Linkage disequilibrium in the human genome. Nature. 2001;411(6834):199-204.
6. Liu Y, Elsworth B, Erola P, Haberland V, Hemani G, Lyon M, et al. EpiGraphDB: a database and data mining platform for health data science. Bioinformatics. 2021;37(9):1304-11.
7. Cotsapas C, Hafler DA. Immune-mediated disease genetics: the shared basis of pathogenesis. Trend Immunol. 2013;34(1):22-6.
8. Ricano-Ponce I, Wijmenga C. Mapping of immune-mediated disease genes. Ann Rev Genom Hum Genet. 2013;14:325-53.
9. Tabasi M, Eybpoosh S, Heravi FS, Siadat SD, Mousavian G, Elyasinia F, et al. Comparison of Gut Microbiota and Serum Biomarkers in Obese Patients Diagnosed with Diabetes and Hypothyroid Disorder. 2020.
10. Baranova A, Cao H, Teng S, Su KP, Zhang F. Shared genetics and causal associations between COVID-19 and multiple sclerosis. J Med Virol. 2023;95(1):e28431.
11. Maier LM, Lowe CE, Cooper J, Downes K, Anderson DE, Severson C, et al. IL2RA genetic heterogeneity in multiple sclerosis and type 1 diabetes susceptibility and soluble interleukin-2 receptor production. PLoS Genet. 2009;5(1):e1000322.
12. Ramirez de Arellano E, Diez-Fuertes F, Aguilar F, de la Torre Tarazona HE, Sánchez-Lara S, Lao Y, et al. Novel association of five HLA alleles with HIV-1 progression in Spanish long-term non progressor patients. PLoS One. 2019;14(8):e0220459.
13. Teixeira S, De Sá N, Campos D, Coelho A, Guimarães M, Leite T, et al. Association of the HLA-B* 52 allele with non-progression to AIDS in Brazilian HIV-1-infected individuals. Gene Immunit. 2014;15(4):256-62.
14. Mehrbod P, Eybpoosh S, Farahmand B, Fotouhi F, Khanzadeh Alishahi M. Association of the host genetic factors, hypercholesterolemia and diabetes with mild influenza in an Iranian population. Virol J. 2021;18:1-11.
15. Mehrbod P, Eybpoosh S, Fotouhi F, Shokouhi Targhi H, Mazaheri V, Farahmand B. Association of IFITM3 rs12252 polymorphisms, BMI, diabetes, and hypercholesterolemia with mild flu in an Iranian population. Virol J. 2017;14(1):1-8.

[Downloaded from mjiri.iums.ac.ir on 2026-11-10]

[DOI: 10.47176/mjiri.38.51]

16. Singh R, Patel V, Mureithi MW, Naranbhai V, Ramsuran D, Tulsi S, et al. TRIM5α and TRIM22 are differentially regulated according to HIV-1 infection phase and compartment. J Virol. 2014;88(8):4291-303.
17. Tavakoli R, Rahimi P, Hamidi-Fard M, Eybpoosh S, Doroud D, Sadeghi SA, et al. Impact of TRIM5α and TRIM22 Genes Expression on the Clinical Course of Coronavirus Disease 2019. Arch Med Res. 2022.
18. Xiang R, van den Berg I, MacLeod IM, Daetwyler HD, Goddard ME. Effect direction meta-analysis of GWAS identifies extreme, prevalent and shared pleiotropy in a large mammal. Communicat Biol. 2020;3(1):88.
19. Thomas DL, Thio CL, Martin MP, Qi Y, Ge D, O'hUigin C, et al. Genetic variation in IL28B and spontaneous clearance of hepatitis C virus. Nature. 2009;461(7265):798-801.
20. Ware JJ, Aveyard P, Broderick P, Houlston RS, Eisen T, Munafò MR. The association of rs1051730 genotype on adherence to and consumption of prescribed nicotine replacement therapy dose during a smoking cessation attempt. Drug and Alcohol Depend. 2015;151:236-40.
21. Márquez A, Martín J, editors. Genetic overlap between type 1 diabetes and other autoimmune diseases. Semin Immunopathol; 2022: Springer.
22. Zubkiewicz-Kucharska A, Jamer T, Chrzanowska J, Akutko K, Pytrus T, Stawarski A, Noczyńska A. Prevalence of haplotype DQ2/DQ8 and celiac disease in children with type 1 diabetes. Diabetol Metabol Synd. 2022;14(1):128.
23. Carter AJ, Nguyen AQ. Antagonistic pleiotropy as a widespread mechanism for the maintenance of polymorphic disease alleles. BMC Med Geneti. 2011;12:1-13.
24. McKusick V. Online Mendelian inheritance in man (OMIM) database [internet]. Bethesda: National Center for Biotechnology Information for the National Institute of Health. 2004.
25. Safran M, Dalah I, Alexander J, Rosen N, Iny Stein T, Shmoish M, et al. GeneCards Version 3: the human gene integrator. Database. 2010;2010.
26. Rebhan M, Chalifa-Caspi V, Prilusky J, Lancet D. GeneCards: a novel functional genomics compendium with automated data mining and query reformulation support. Bioinform (Oxford, England). 1998;14(8):656-64.
27. Piñero J, Saüch J, Sanz F, Furlong LI. The DisGeNET cytoscape app: Exploring and visualizing disease genomics data. Comput Struct Biotechnol J. 2021;19:2960-7.
28. Mailman MD, Feolo M, Jin Y, Kimura M, Tryka K, Bagoutdinov R, et al. The NCBI dbGaP database of genotypes and phenotypes. Nat Genet. 2007;39(10):1181-6.
29. Dreszer TR, Karolchik D, Zweig AS, Hinrichs AS, Raney BJ, Kuhn RM, et al. The UCSC Genome Browser database: extensions and updates 2011. Nucl Acid Res. 2012;40(D1):D918-D23.
30. Meyer LR, Zweig AS, Hinrichs AS, Karolchik D, Kuhn RM, Wong M, et al. The UCSC Genome Browser database: extensions and updates 2013. Nucl Acid Res. 2012;41(D1):D64-D9.

[Downloaded from mjiri.iums.ac.ir on 2026-11-10]

[DOI: 10.47176/mjiri.38.51]