Is Hearing Loss a Cause of Dementia? The Use of Mendelian Randomisation in Assessing Causality

Word Count: 3897

Rouse Research Essay Award

2026

Abstract

This paper investigates hearing loss as a potentially modifiable risk factor for dementia. Due to the ageing population, it is important to identify ways to prevent dementia, a disease which primarily affects the elderly population and is therefore a growing public health issue. Whilst hearing loss is strongly associated with increased likelihood of dementia, and several mechanisms for hearing loss causing cognitive decline have been theorised, a causal relationship has not been proven. Using two-sample Mendelian randomisation, I assess the likelihood of causality, harnessing genetic data in order to overcome the issue of confounding factors. My analyses show no significant causal relationship between hearing loss and dementia, suggesting the observed association is not due to causation, and is instead a result of confounding factors and/or reverse causation. Therefore, this study suggests future research into dementia prevention should focus on other identified risk factors such as smoking, diet and education.

Acknowledgements

I would like to thank Dr Dipender Gill and the team at Sequoia Genetics for introducing me to Mendelian randomisation and providing me with work experience which was helpful for this project. Also, thank you to Professor Gibran Hemani at the MRC Integrative Epidemiology Unit, University of Bristol, for giving advice on how to write the code to perform the MR analysis. Finally, thank you to my supervisor Mrs Pepona for all her support throughout the year, including her feedback on my initial draft.

Introduction

Dementia is a growing public health issue, primarily amongst the elderly, presenting serious challenges for healthcare systems globally. Currently, dementia affects over 50 million people and the prevalence is likely to continue to increase substantially, as a result of an ageing population.[1] Pharmaceutical treatments focus on symptom modification or slowing disease progression, but are currently unable to reverse neuropathology.[2, 3] Therefore, identifying and addressing modifiable risk factors is a more promising way to combat this challenge, with studies estimating more than a third of dementia cases could be prevented by taking precautionary measures.[1]

Various studies have shown correlation between hearing loss and dementia.[4,5] Hearing loss is also a prevalent issue amongst the elderly and is a known risk factor for dementia.[1] Interventions such as hearing aids can address hearing loss and could therefore make hearing loss a potentially modifiable risk factor for dementia. If we were certain hearing loss was causal for dementia, society would be able to invest in better screening programmes for hearing loss and patients would be more motivated to comply with use of hearing loss interventions like hearing aids. For example, after human papillomavirus (HPV) infection was proven to cause cervical cancer, the UK implemented an HPV vaccine programme in teenage girls. Over an 11-year period, this programme has already prevented 450 cervical cancers and roughly 17,200 cases of precancerous conditions, demonstrating the public health benefits of treating risk factors to avoid more severe diseases.[6]

However, a causal relationship between hearing loss and dementia has not been proven. Mendelian randomisation (MR) is a technique which uses genetics to address the problem of correlation versus causation and can enable one to infer causality. In this project, I use MR to determine the likelihood of a causal relationship between hearing loss and dementia.

Hearing Loss and Dementia

A systematic review and meta-analysis of cohort studies found that adult-onset hearing loss is linked to an increased risk of mild to severe cognitive problems.[7] The study assessed the likelihood that the observed association is due to a causal relationship according to the Bradford Hill criteria. This is set of nine principles, outlined below, which evaluate the epidemiological evidence of a presumed causal relationship.[8, 9]

1. Strength/effect size – the stronger the association, the more likely it is causal
2. Consistency – reproducible findings strengthen the likelihood of a causal effect
3. Specificity – specific associations, for example a specific population and a specific disease and no other likely explanation, is more likely to be causal
4. Temporality – the effect must occur after the cause
5. Dose-response – greater exposure should lead to greater incidence of the effect

6. Biological plausibility – a plausible mechanism explaining the causal pathway provides helpful evidence for a causal relationship
7. Coherence – a cause-and-effect interpretation of data should not conflict with the current knowledge available
8. Experiment – studies involving manipulation of the exposure may provide the strongest evidence for a causal effect
9. Analogy – if a causal relationship is already established, the standards of evidence for a similar second causal agent can be lowered

One main finding of this study was that the risk of dementia increases by 16% per each decibel worsening of hearing, evidence of the dose response criterion.

Considering the criterion of biological plausibility, there are a number of mechanisms through which hearing loss could cause dementia.[10] Firstly, the decreased stimulation of cognitive processing caused by reduced speech and language input in an impoverished acoustic environment could negatively affect brain structure and function, known as disuse atrophy. This could be a result of reduced access to emotional and verbal information in speech – effectively the brain degenerates due to less 'practice' using some functions. Another potential mechanism is cognitive overload, a mechanism which suggests changes in brain activity are responsible for cognitive deficits. Hearing impairments cause an increased need for cognitive resources to be allocated to listening, therefore these resources are unavailable for other cognitive tasks such as attention, working memory and language processing. However, this mechanism may explain general cognitive deficits in elderly people due to hearing loss (which would be reversible) instead of dementia.

Correlation vs Causation

Despite the strong correlation, a causal relationship has not been proven. There are multiple explanations which could explain the correlation between hearing loss and dementia without a clear cause-and-effect relationship.[11] This idea is captured by the phrase "Correlation does not imply causation." Firstly, there could be reverse causation – for example, the pathological process of dementia damaging the auditory areas of the brain. This would make hearing loss an early indication of dementia, rather than a modifiable risk factor. Alternatively, there could be a third variable, known as a

confounding factor, which affects both the exposure and the outcome, creating the illusion of a causal relationship, see Figure 1.

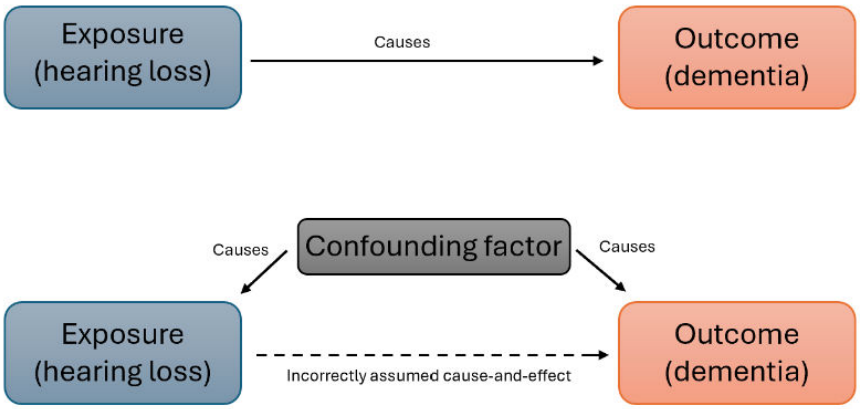


Figure 1. Diagram representation of confounding factors.

Thirdly, the observed association could be due to misdiagnosis of pseudodementia as dementia. Pseudodementia is a clinical term used to describe reversible symptoms of cognitive decline which mimic dementia (including memory deficits, reduced attention, decreased speech fluency and processing issues).[12] It is thought to be caused by mood-related conditions such as depression. Due to the social impact of hearing loss in the elderly population, it could be an explanation for the correlation in which hearing loss does not cause dementia, but instead causes pseudodementia through the intermediate of social isolation and depression.

Finally, one study has suggested hearing loss could be a 'catalyst' for dementia diagnosis. Instead of directly affecting brain structure, it could promote certain factors such as impaired communication, amplifying and unmasking the effects of neurodegenerative pathology and enabling clinical dementia to be diagnosable at lower levels of neurodegenerative pathology.[13]

To truly prove causation, an interventional study involving the manipulation of an independent variable (the exposure) would be required.[14] The most obvious and reliable example in clinical research would be a randomised controlled trial (RCT). Unfortunately, this is not always possible due to ethical reasons or practicalities. For hearing loss and dementia, an interventional study would likely involve recruiting people with hearing loss and assigning them to either the test group (hearing aids provided) or the placebo group (no treatment – e.g. faulty hearing aids). Then, cognitive function would be measured over time. Clearly, the study would be very long and time-consuming. Also, including a placebo group (which is fundamental to the study) could be considered unethical because there are known risks to not treating hearing loss, such as social isolation and depression.[15] Thus, an interventional study is not viable.

Mendelian Randomisation

Determining the causes of disease, known as disease aetiology, is essential in medicine for informing public policy, advising healthy lifestyles, developing new medicines, assessing the impacts of treatments and prioritising healthcare resources, such as in this case when considering the importance of screening for hearing loss.[14] Whilst 'classic epidemiology' is often unable to overcome the issue of confounding factors, the rise of genetic epidemiology offers new solutions.

Genetic epidemiology uses genetic factors in the study of patterns of health and disease at the population level.[14] Recent advances in genomics and technology have provided more opportunities for the use of genetics in clinical research. Firstly, the Human Genome Project produced the publication of the entire human genetic code in 2004.[16] Reductions in the cost of both DNA sequencing and data storage means it is now economically possible to collect and store genetic information for a large number of individuals. This enables the discovery of genetic variants linked to specific diseases, often though genome-wide association studies (GWAS).[14] Finding these genetic variants may lead to more routine use of the widely debated gene editing which could allow us to eradicate hereditary diseases. However, currently, one of the main uses of genetic data is a technique called Mendelian randomisation (MR).

Mendelian randomisation is the use of genetic variants as instrumental variables (IVs) to make causal inferences about the effect of an 'exposure' on an 'outcome'.[14] 'Exposure' refers to the presumed cause which is being investigated, whilst 'outcome' refers to the clinical situation of interest. MR relies on the instrumental variable meeting three core assumptions (see Figure 2):

a. The variant is associated with the exposure
b. The variant is not associated with any confounding factor
c. The variant must not be directly associated with the outcome, except via the potential hypothesised causal pathway through the exposure

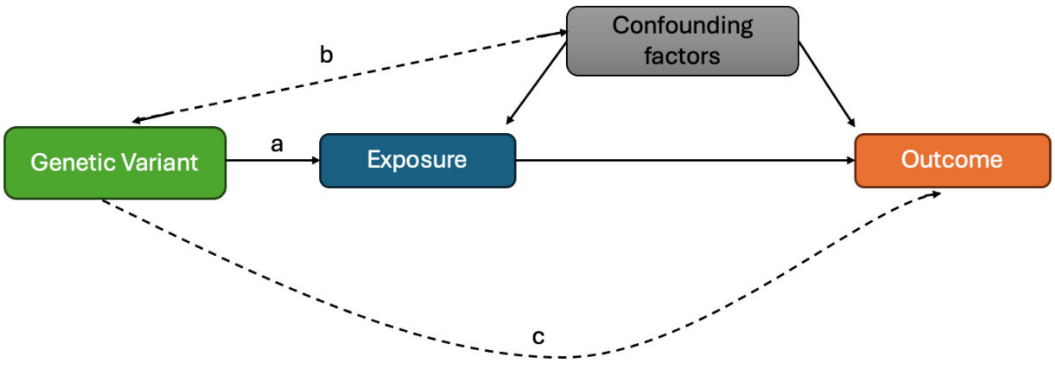


Figure 2. Visual representation of the role of the genetic variant.

If all three assumptions are fulfilled, then any association of the genetic variant with the outcome must come via the association between the variant and the exposure, therefore implying a causal relationship between exposure and outcome.[14]

Applications of Mendelian Randomisation

One main use of MR, such as in this study, is to assess observed associations for causality to find modifiable risk factors which can help when deciding public health policies and advising on lifestyle choices.

However, there are other useful applications of MR such as finding drug targets to improve pharmaceutical research and development (R&D) efficiency. Drug R&D efficiency refers to the number of drugs successfully approved by regulators such as the FDA and brought to market, compared to R&D spending.[17] Unfortunately, despite technological advancements increasing the rates of identifying targets and discovering new drugs against them, efficiency of this industry has declined since the 1950s. New drugs have an average cost of over $1-2 billion, not due to productions costs but instead as a result of needing to absorb the costs of the many failed attempts.[18, 19]

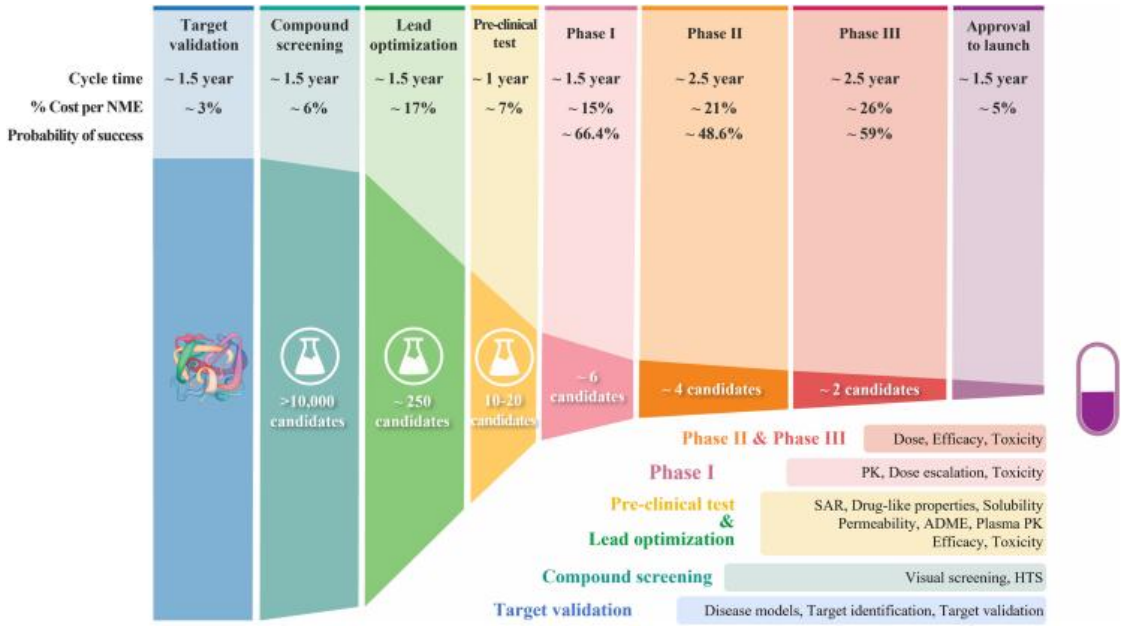


Figure 3. The drug discovery pipeline.[18]

Figure 3 shows the pathway that new drug candidates follow and how the chance of reaching the 'Approval to launch' stage is low.[18] One study on the success rates of clinical development between 2011 and 2020 calculated the overall likelihood of approval (LOA) from Phase I for all developmental candidates to be 7.9%.[20] The main reasons for drug failure are lack of commercial need (10%), poor drug-like properties such as issues with absorption (10-15%), safety concerns over the toxicity (30%) and majorly, lack of clinical efficacy (40-50%)[21, 22] The fact that the highest cause for drug failure at the clinical stage is that the drug is not actually effective is potentially surprising. However, this means

more focused work into finding drug targets with high causal effects on disease is required, in order to ensure drugs developed hit the right molecule to produce an effective treatment. One study has estimated that selecting genetically supported targets could double the success rate in clinical development.[23] Thus, Mendelian randomisation is a promising technique for improving drug R&D productivity by providing insight into which targets to focus on and thus optimise use of resources on only the most promising drugs.[24]

Assessing the Relationship between Hearing Loss and Dementia

Two previous MR studies have inferred a causal relationship between hearing loss and dementia.[25, 26] The first, 'Relationship between hearing impairment and dementia and cognitive function: a Mendelian randomization study', found evidence supporting a causal relationship between hearing impairment and increased risk of dementia, in addition to poorer cognitive function.[25] Whilst the causal effect estimates were statistically significant, they used a relaxed P-value threshold for one of the exposure datasets, potentially leading to the presence of weak instrument bias (see appendix I for definition). The second study, 'A comprehensive evaluation on the associations between hearing and vision impairments and risk of all-cause and cause-specific dementia: results from cohort study, meta-analysis and Mendelian randomization study', concluded there was compelling evidence for a causal relationship between hearing impairments and dementia.[26] However, the auditory and nervous systems are very complex which raises concerns over the issue of pleiotropy (see appendix I for definition). Therefore, I performed a Mendelian randomisation study using larger and more comprehensive datasets in an attempt to more definitively assess the association between hearing loss and dementia.

Mendelian Randomisation in this Investigation

For the practical element of this project, I use two-sample MR, which is a method of Mendelian randomisation that uses summary level statistics to derive the causal estimates. It takes the genetic IVs for exposure and outcomes from two different, non-overlapping samples (meaning there is no overlap of participants). Genetic IVs are usually single nucleotide polymorphisms (SNPs, see glossary for definition). The summary level statistics include the effect of each identified SNP on the trait of interest (either the exposure or outcome). The effect is calculated by linear regression and represents the strength of the association between the SNP and the trait.

In two-sample MR, SNPs which are common to both the exposure and outcome datasets are extracted and then a Wald ratio for each SNP is calculated. The Wald ratio is the estimated causal effect of the exposure on the outcome, calculated by dividing the effect of the variant on the outcome (β_{out}) by its effect on the exposure (β_{exp}), as shown in Figure 4.[27]

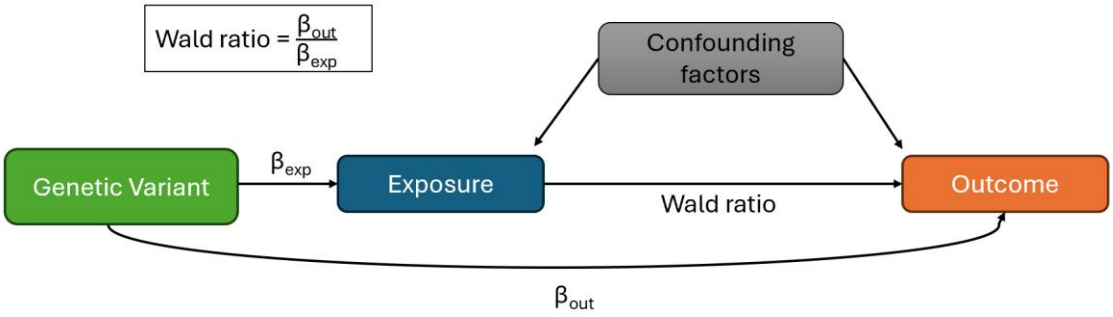


Figure 4. Wald ratio calculation.

A Wald ratio close to 1 implies a causal relationship, a ratio close to -1 implies an inverse relationship and a ratio close to 0 suggests the null hypothesis (the exposure has no causal effect on the outcome) is true. The Wald ratios for different SNPs can be combined using techniques such as Inverse Variance Weighted (IVW) regression.[28] IVW regression calculates a weighted average of the ratios, where variants with smaller standard errors are given more weight. MR Egger is another similar method, however uses a regression approach which allows for detection of bias by not forcing the line through 0, and instead uses the intercept to detect pleiotropy.

Methodology

In order to perform two-sample MR analysis, I followed a series of steps outlined in the flowchart shown in Figure 5. Firstly, I identified two genome-wide association studies, for hearing loss and for dementia, which would be suitable to provide summary statistics for the exposure and outcomes in the MR analysis. When selecting appropriate GWAS, I focused on studies with larger sample sizes and clear case definitions (for example, clinically diagnosed dementia as opposed to self-reported dementia). After selection of GWAS, I wrote the code in R (using R studio) which would be used to perform the MR analysis using the Two sample MR package for R.[29] The R code is attached in Appendix II. I followed the workflow explained by Gibran Hemani in his introduction to two sample Mendelian randomisation.[30]

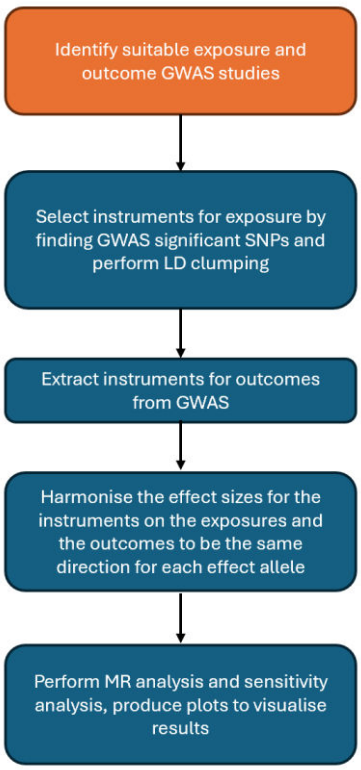


Figure 5. MR analysis, blue = R code steps.

For the exposure dataset, I used the risk variants associated with hearing loss identified by Natalia Trpchevska et al. in [31]. This GWAS is considered a strong research paper due to its large sample size and its analysis of gene expression. The authors mapped the variants to nearby genes and determined which cells expressed these genes. The finding that the genes were more frequently expressed in cells which form the cochlea (a key component of the ear) strengthened the association between these variants and hearing loss. Therefore, this study provides a valuable dataset to use for the exposure as the genetic variants are highly associated with hearing loss and may reduce the potential for pleiotropy.

For the outcome dataset, I used the genetic variants associated with Alzheimer's and related dementias identified by Céline Bellenguez et al. in [32] because this GWAS had a very large sample size (over 788,000 participants) and therefore strong statistical power and increased reliability of results.

When writing the code, the first step was selecting the genome-wide significant SNPs to be used as genetic instruments for the exposure. I used the standard threshold of p-value $< 5x10^{-8}$ in order to meet the strong association assumption (a) described on page 5. LD clumping is a technique which is used to overcome the issue of linkage disequilibrium (LD). LD refers to the positive association of alleles in a population, often due to them being located close to each other on a chromosome, leading to them being inherited together.[33] LD clumping selects only the most significant SNP from each locus and removes the SNPs with higher p-values that are located within the same region.[34]

All SNPs associated with the outcome in the outcome GWAS were extracted, and then harmonisation was performed. Harmonisation aligns the direction of the effect alleles in both the exposure and outcome datasets (see glossary).

The two-sample MR analysis package was then run to provide causal estimates. The two-sample MR package provides five methods, however I will focus on two methods (MR Egger and Inverse Variance Weighted) because I have greater understanding of how these calculate causal estimates. In addition to this, pleiotropy and heterogeneity tests were performed in order to assess the reliability of the results.

Results

The Mendelian randomisation analyses using 41 SNPs found no causal relationship between hearing loss and dementia. All methods resulted in statistically insignificant causal effect estimates with p-values all greater than 0.05, see Table 1 and Figure 6.

Table 1. MR analysis results

Method	N(SNP)	Beta-value	Standard Error	P-Value
MR Egger	41	-0.77	0.76	0.32
Weighted median	41	-0.40	0.29	0.18
Inverse variance weighted	41	-0.08	0.25	0.75
Simple mode	41	-0.48	0.57	0.41
Weighted mode	41	0.50	0.66	0.45

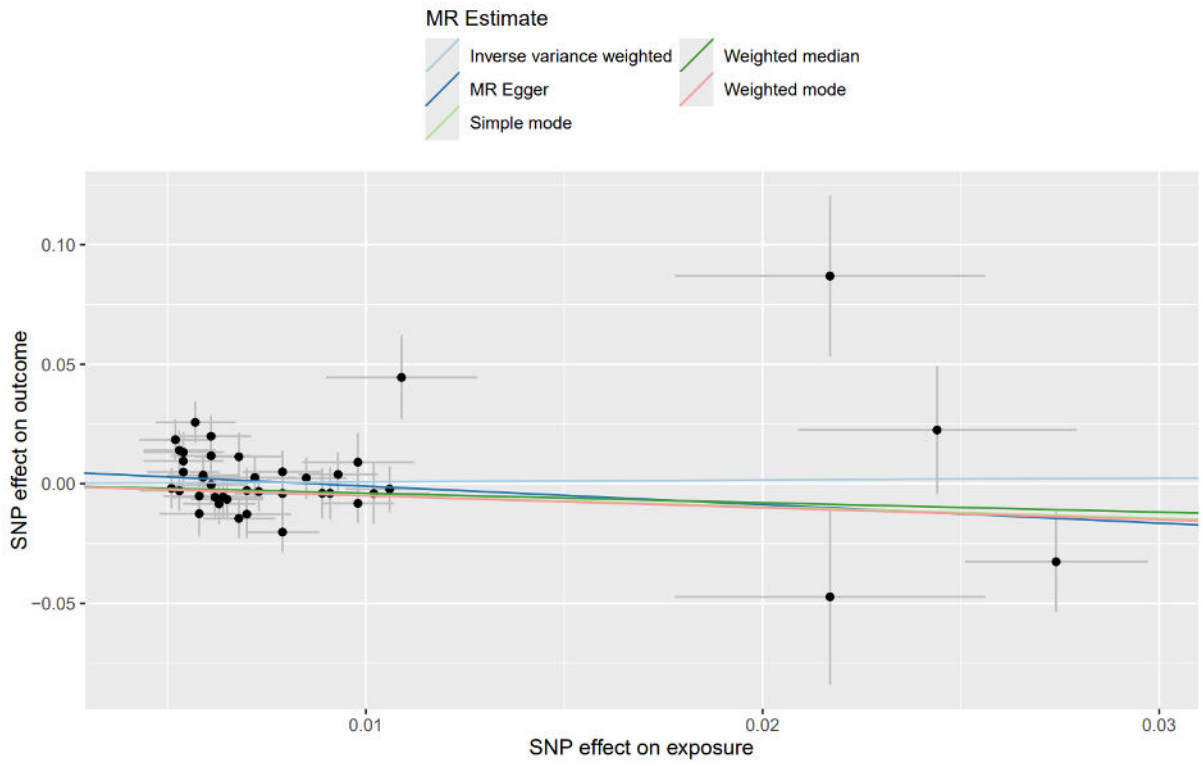


Figure 6. Scatterplot of the MR analysis results.

The inverse variance weighted beta value is close to zero (β = -0.08) which indicates no causal effect of the exposure on the outcome. The p-value (p = 0.75) is greater than 0.05 and therefore not statistically significant. Additionally, the standard error is large (SE = 0.25), suggesting there is great uncertainty in the estimate.

The MR Egger method suggested an inverse relationship between hearing loss and dementia (β = -0.77) however the result is not significant (p = 0.32) and the standard error (SE = 0.76) implies high uncertainty with a value almost as large as the causal estimate.

The heterogeneity tests indicated significant variability between SNP-specific causal effect estimates (MR Egger Q = 64.8, p = 0.0058; IVW Q = 67.1, p = 0.0046) suggesting assumption (c) may have been violated by pleiotropy, see table 2. However, the MR Egger intercept test in table 3, which tests for pleiotropy, was insignificant (p = 0.25). Therefore, there is no strong evidence for horizontal pleiotropy in these analyses.

Table 2. Heterogeneity test results

Method	Q	P-value
MR Egger	64.8	0.0058
Inverse variance weighted	67.1	0.0046

Table 3. Pleiotropy test results

Method	Egger intercept	Standard Error	P-value
MR Egger	0.0067	0.0057	0.25

Discussion

The results of the Mendelian randomisation analyses showed no significant causal relationship between hearing loss and dementia. This may indicate that the observed association is due to confounding factors or reverse causation.

Given these findings, it is probable that attempts to modify hearing loss through the use of interventions such as hearing aids may not reduce the incidence or prevalence of dementia. Thus, my results suggest that investment in screening and treatment programs for hearing loss would not be an effective effort to combat the public health issue of dementia. Investigations should focus on other potentially modifiable risk factors such as diet, smoking, and education.[35]

Despite this, there are still important benefits of widespread hearing loss treatment in an ageing population. Hearing loss has been linked to social isolation and loneliness in older adults.[36] These issues can lead to more severe mental health issues and have also been linked to increased mortality rates and higher healthcare spending. One example would be pseudodementia, where social isolation caused by hearing loss may lead to depression, then cognitive decline which puts increased pressure on the healthcare

system. This emphasises the importance of hearing loss interventions in order to improve quality of life of those with hearing loss and relieve pressure on healthcare systems by improving the general health of the elderly.

This study has several strengths, including the large sample sizes, strong instrument selection and use of multiple MR analyses. The use of large GWAS datasets improved the reliability of the findings. To avoid weak instrument bias, I selected SNPs at genome-wide significance ($p < 0.5e^{-8}$), ensuring SNPs are strongly associated with the exposure. This was especially important due to the genetic complexity of hearing loss. Performing LD clumping prevents bias from correlated SNPs. Additionally, the application of multiple MR methods allows for validation of evidence and makes use of the advantages of each method. For example, the IVW method has high statistical power, however assumes all SNPs are valid and is susceptible to bias in the presence of pleiotropy. Conversely, the MR Egger method is more robust to invalid SNPs though has lower power and thus often produces non-significant results even if the effect exists.

However, these results should be interpreted with caution because my study has various limitations that could have led to a false negative result. Firstly, the heterogeneity tests suggested there was pleiotropy present due to the high variability between causal effects of individual SNPs. Despite the MR Egger test for pleiotropy being insignificant, there could still be pleiotropy present due to the low statistical power of the test. As a result of the SNPs giving inconsistent signals, the true effect of the exposure on the outcome could be 'blurred', causing the overall estimate to move towards the null hypothesis.

Secondly, the auditory system is genetically complex, meaning hearing loss has many genes which affect it. Therefore, it is likely that each SNP has a small effect on the exposure. This results in the causal effect estimates being biased towards the null hypothesis, which is called weak instrument bias. The insignificant p-values and large standard errors in my results are indicators of low precision and suggest weak instrument bias was present.

A third limitation is the age-dependent nature of both hearing loss and dementia. MR assumes that genetic variants serve as a proxy for lifelong exposure, however hearing loss typically develops in later life and may only influence dementia risk during specific critical periods.[37] Although I used exposure data with a phenotype of age-related hearing impairment, the genetic instruments used may not accurately reflect the timing, duration, or severity of the hearing loss which might cause dementia, leading to dilution of any true causal effect. Consequently, the null findings may reflect limitations in identifying genes associated with the time-dependent traits rather than a true absence of causality. It could be useful to conduct an investigation which differentiates between different onset ages of hearing loss and their effect on dementia.

Conclusion

In conclusion, this study finds hearing loss does not cause dementia. Despite the strong association between hearing loss and dementia, as well as the evidence for the relationship meeting some of the Bradford Hill criteria for causation (including biological plausibility and dose response), the results of my Mendelian randomisation analyses suggest the relationship is not causal.[7,10] This highlights the value in using genetics to overcome the issues of confounding factors and reverse causation when assessing the likelihood of causality for an observed correlation. In light of these findings, projects which aim to combat the growing public health concern of dementia should potentially focus on alternative risk factors in order to identify new ways to prevent this disease. I also stress the importance of detecting and treating hearing loss (e.g. through screening and the use of hearing aids) for protection against other associated social and health issues.[36] Due to the limitations of my research, more studies investigating the relationship between age-related hearing loss and dementia could be beneficial, perhaps distinguishing between different onset ages for hearing loss to determine if the age-dependent nature of these traits has affected this study.

References

[1] Livingston, Gill et al. “Dementia prevention, intervention, and care: 2020 report of the Lancet Commission.” *Lancet (London, England)* vol. 396,10248 (2020): 413-446. doi:10.1016/S0140-6736(20)30367-6

[2] Kabir, Md Tanvir et al. “Combination Drug Therapy for the Management of Alzheimer's Disease.” *International journal of molecular sciences* vol. 21,9 3272. 5 May. 2020, doi:10.3390/ijms21093272

[3] NHS. *What are the treatments for dementia?* 24 July 2023. 19 February 2026. <https://www.nhs.uk/conditions/dementia/about-dementia/treatment/#:~:text=Medicines%20to%20treat%20dementia,Memantine>.

[4] Lin, Vincent Y. W.; Black, Sandra E.. Linking Deafness and Dementia: Challenges and Opportunities. Otology & Neurotology 38(8):p e237-e239, September 2017. | DOI: 10.1097/MAO.0000000000001408

[5] Loughrey, David G et al. “Association of Age-Related Hearing Loss With Cognitive Function, Cognitive Impairment, and Dementia: A Systematic Review and Meta-analysis.” *JAMA otolaryngology-- head & neck surgery* vol. 144,2 (2018): 115-126. doi:10.1001/jamaoto.2017.2513

[6] Matson, L. (2021, November 3). “The power of science”: HPV vaccine proven to dramatically reduce cervical cancer. Retrieved from Cancer Research UK: https://news.cancerresearchuk.org/2021/11/03/the-power-of-science-hpv-vaccine-proven-to-dramatically-reduce-cervical-cancer/

[7] Yu, Ruan-Ching et al. “Adult-onset hearing loss and incident cognitive impairment and dementia - A systematic review and meta-analysis of cohort studies.” *Ageing research reviews* vol. 98 (2024): 102346. doi:10.1016/j.arr.2024.102346

[8] Wikipedia. Bradford Hill criteria. 7 November 2025. 19 February 2026. <https://en.wikipedia.org/wiki/Bradford_Hill_criteria>.
[9] Fedak, Kristen M et al. “Applying the Bradford Hill criteria in the 21st century: how data integration has changed causal inference in molecular epidemiology.” *Emerging themes in epidemiology* vol. 12 14. 30 Sep. 2015, doi:10.1186/s12982-015-0037-4
[10] Griffiths, Timothy D et al. “How Can Hearing Loss Cause Dementia?.” *Neuron* vol. 108,3 (2020): 401-412. doi:10.1016/j.neuron.2020.08.003
[11] *Correlation does not imply causation*. 25 January 2026. 21 February 2026. <https://en.wikipedia.org/wiki/Correlation_does_not_imply_causation#Examples_of_illogically_inferring_causation_from_correlation>.
[12] Mouta, Salomé et al. “What do we know about pseudodementia?.” *General psychiatry* vol. 36,4 e100939. 22 Aug. 2023, doi:10.1136/gpsych-2022-100939
[13] Levett, Benjamin A et al. “Hearing impairment and dementia: cause, catalyst or consequence?.” *Journal of neurology* vol. 272,6 402. 16 May. 2025, doi:10.1007/s00415-025-13140-x
[14] Burgess, Stephen and Simon G Thompson. *Mendelian Randomization: Methods for Using Genetic Variants in Causal Estimation*. Taylor & Francis Group, 2015.
[15] John Hopkins Medicine. *The Hidden Risks of Hearing Loss*. n.d. 21 February 2026. <https://www.hopkinsmedicine.org/health/wellness-and-prevention/the-hidden-risks-of-hearing-loss>.
[16] National Human Genome Research Institute. *Human Genome Project Timeline*. 5 July 2022. 22 February 2026.
[17] Scannell, J., Blanckley, A., Boldon, H. *et al.* Diagnosing the decline in pharmaceutical R&D efficiency. *Nat Rev Drug Discov* 11, 191–200 (2012). https://doi.org/10.1038/nrd3681
[18] Sun, Duxin et al. “Why 90% of clinical drug development fails and how to improve it?.” *Acta pharmaceutica Sinica. B* vol. 12,7 (2022): 3049-3062. doi:10.1016/j.apsb.2022.02.002
[19] Torreele, Els. “Why are our medicines so expensive? Spoiler: Not for the reasons you are being told....” *The European journal of general practice* vol. 30,1 (2024): 2308006. doi:10.1080/13814788.2024.2308006
[20] BIO, Informa Pharma Intelligence, QLS. (2021). *Clinical Development Success Rates and Contributing Factors 2011–2020*. Biotechnology Innovation Organization.
[21] Dowden H., Munro J. Trends in clinical success rates and therapeutic focus. Nat Rev Drug ` Discov. 2019;18:495–496. doi: 10.1038/d41573-019-00074-z.
[22] Harrison R.K. Phase II and phase III failures: 2013–2015. Nat Rev Drug Discov. 2016;15:817–818. doi: 10.1038/nrd.2016.184.
[23] Nelson, Matthew R et al. “The support of human genetic evidence for approved drug indications.” Nature genetics vol. 47,8 (2015): 856-60. doi:10.1038/ng.3314
[24] Daghlas, Iyas, and Dipender Gill. “Mendelian randomization as a tool to inform drug development using human genetics.” *Cambridge prisms. Precision medicine* vol. 1 e16. 8 Feb. 2023, doi:10.1017/pcm.2023.5
[25] Jiang, Deming et al. “Relationship between hearing impairment and dementia and cognitive function: a Mendelian randomization study.” *Alzheimer's research & therapy* vol. 16,1 215. 9 Oct. 2024, doi:10.1186/s13195-024-01586-6
[26] Jiang, Fan et al. “A comprehensive evaluation on the associations between hearing and vision impairments and risk of all-cause and cause-specific dementia: results from

cohort study, meta-analysis and Mendelian randomization study." *BMC medicine* vol. 22,1 518. 7 Nov. 2024, doi:10.1186/s12916-024-03748-7

[27] Burgess S, Small DS, Thompson SG. A review of instrumental variable estimators for Mendelian randomization. Stat Methods Med Res 2017; 26: 2333-2355.

[28] Nguyen, Kevin, and Braxton D Mitchell. "A Guide to Understanding Mendelian Randomization Studies." *Arthritis care & research* vol. 76,11 (2024): 1451-1460. doi:10.1002/acr.25400

[29] Posit team (2026). RStudio: Integrated Development Environment for R. Posit Software, PBC, Boston, MA. URL http://www.posit.co/.

[30] Hemani, G. (n.d.). Introduction. Retrieved from TwoSampleMR: https://mrcieu.github.io/TwoSampleMR/articles/introduction.html

[31] Trpchevska, Natalia et al. "Genome-wide association meta-analysis identifies 48 risk variants and highlights the role of the stria vascularis in hearing loss." *American journal of human genetics* vol. 109,6 (2022): 1077-1091. doi:10.1016/j.ajhg.2022.04.010

[32] Bellenguez, C., Küçükali, F., Jansen, I.E. *et al.* New insights into the genetic etiology of Alzheimer's disease and related dementias. *Nat Genet* 54, 412–436 (2022). https://doi.org/10.1038/s41588-022-01024-z

[33] Wikipedia. (2026, March 1). Linkage Disequilibrium. Retrieved from Wikipedia: https://en.wikipedia.org/wiki/Linkage_disequilibrium

[34] Buxton J, Turney J. (2007). The Rough Guide to Genes & Cloning

[35] Jones, Aaron et al. "Potentially Modifiable Risk Factors for Dementia and Mild Cognitive Impairment: An Umbrella Review and Meta-Analysis." *Dementia and geriatric cognitive disorders* vol. 53,2 (2024): 91-106. doi:10.1159/000536643

[36] Shukla, Aishwarya et al. "Hearing Loss, Loneliness, and Social Isolation: A Systematic Review." *Otolaryngology--head and neck surgery : official journal of American Academy of Otolaryngology-Head and Neck Surgery* vol. 162,5 (2020): 622-633. doi:10.1177/0194599820910377

[37] Richmond, Rebecca C, and George Davey Smith. "Mendelian Randomization: Concepts and Scope." Cold Spring Harbor perspectives in medicine vol. 12,1 a040501. 4 Jan. 2022, doi:10.1101/cshperspect.a040501

Appendix I - Glossary

Allele – a variant of a specific gene. Different alleles may result in proteins which are more/less effective. It is possible to have 0, 1 or 2 copies of a given allele.

Beta-value – estimated causal effect size.

Effect Allele – the genetic variant associated with the phenotype of interest.

Effect Allele Direction – indicates if the effect allele increases likelihood/severity of phenotype (positive value) or if the allele decreases risk (negative value).

Gene – section of DNA which codes for a specific protein.

Genotype – refers to which alleles an individual has for a specific gene.

Genome-Wide Association Study (GWAS) – large-scale observational studies which analyse genomes of thousands of individuals to identify genetic variations (often SNPs) that are strongly associated with specific diseases or traits.

Harmonisation – changes all the effect allele directions to be in the same direction for both exposures and outcomes.

Heterogeneity – occurs when the same phenotypic trait is caused by different genetic mechanisms, not a single mutation. In Mendelian randomisation, presence of heterogeneity is suggested if different genetic instruments yield different causal estimates. Presence of heterogeneity is an indicator that one of the core assumptions of MR has been violated.

Phenotype – observable characteristics/traits resulting from genotype.

Pleiotropy – a single gene produces multiple apparently unrelated phenotypic traits. In the context of Mendelian randomisation, it could lead to the genetic variant affecting the exposure and the outcome through different biological pathways which violates one of the core assumptions.

P-value – a statistical measure between 0 and 1 which indicates the probability of observing the results, assuming the null hypothesis is true. Typically, a p-value less than 0.05 means the results can be considered statistically significant.

Single Nucleotide Polymorphism (SNP) – a common type of genetic variation amongst the population. It refers to a difference in a single DNA building block (called nucleotides), producing a variant of a gene which must be present in at least 1% of the population to be classified as a SNP.

Statistical power – the probability of a test detecting an effect when one exists, representing the sensitivity of a study to avoid false negatives.

Weak instrument bias – weakly associated variables cause bias (often towards the null).

Appendix II – R Code

```r
library(TwoSampleMR)
library(data.table)
library(dplyr)
library(gwasvcf)
library(gwasglue)
# set working directory
setwd("/Users/Richard.Mason/MR analysis")
gwas <- fread("ARHL.txt") # create table using data from 'ARHL.txt'
gwas <- as.data.frame(gwas) # convert table to standard data frame

# reformat table for use in functions
exposure_dat <- format_data(
  gwas,
  type = "exposure",
  snp_col = "SNP",
  beta_col = "Effect",
  se_col = "StdErr",
  effect_allele_col = "Allele1",
  other_allele_col = "Allele2",
  eaf_col = "Freq1",
  pval_col = "P.value"
)

# find SNPs with significant p-value (less that 5e-8, the standard GWAS level)
exposure_sig <- exposure[exposure$pval.exposure < 5e-8, ]

write.table(
  exposure_sig,
  "exposure_sig.txt",
  sep = "\t",
  row.names = FALSE,
  quote = FALSE
)
# insert API token to perform clumping
Sys.setenv(OPENGWAS_JWT="eyJhbGciOiJSUzI1NiIsImtpZCI6ImFwaS1qd3QiLCJ0eXAiOiJKV1QifQ.eyJpc3MiOiJhcGkub3Blbmd3Y
# perform LD clumping (removes variants that are highly correlated with each other)
exposure_clumped <- clump_data(exposure_sig)

# extract outcome data
outcome <- read.table("outcomes.tsv",
                      header = TRUE,       # first row contains column names
                      sep = "\t",          # tab-separated
                      stringsAsFactors = FALSE)

# reformat table
outcome_dat <- format_data(
  outcome,
  type = "outcome",
  snp_col = "variant_id",
  beta_col = "beta",
  se_col = "standard_error",
  effect_allele_col = "effect_allele",
  other_allele_col = "other_allele",
  eaf_col = "effect_allele_frequency",
  pval_col = "p_value"
)

# harmonise exposure and outcome data, ready for MR analyses
harmonised_data <- harmonise_data(
  exposure_clumped,
  outcome_dat,
  action = 2
)

# run MR analyses
mr_results <- mr(harmonised_data)
mr_results # output table with results

# visualise results
mr_scatter_plot(mr_results, harmonised_data)

# perform heterogeneity and pleiotropy tests
mr_heterogeneity(harmonised_data)
mr_pleiotropy_test(harmonised_data)
```