

1 Title: Mendelian randomization study of diabetes and dementia in the Million Veteran Program

2
3 Elizabeth M Litkowski (0000-0002-7079-696X),^{1,2,3} Mark W Logue,^{4,5} Rui Zhang,⁴ Brian R Charest,⁶ Ethan
4 M Lange,² John E Hokanson,³ Julie A Lynch,^{7,8} Marijana Vujkovic,^{9,10} Lawrence S Phillips,^{11,12} Richard L
5 Hauger,^{13,14} *Leslie A Lange,^{2,3} *Sridharan Raghavan,^{1,2} on behalf of the VA Million Veteran Program
6 (MVP)

7
8 Corresponding Author: Elizabeth Litkowski
9 e-mail: elizabeth.litkowski@cuanschutz.edu

10 University of Colorado, School of Medicine
11 Colorado Center for Personalized Medicine
12 1890 N Revere Ct Mail Stop F563
13 Aurora, CO 80045 USA
14 303-588-9411

15
16 Author Affiliations

17 ¹VA Eastern Colorado Healthcare System, Aurora, CO, 80045 USA

18 ²Department of Biomedical Informatics, University of Colorado Anschutz Medical Campus, Aurora, CO,
19 80045, USA

20 ³Department of Epidemiology, University of Colorado Anschutz Medical Campus, Aurora, CO, 80045,
21 USA

22 ⁴National Center for PTSD, Behavioral Sciences Division, VA Boston Healthcare System, Boston, MA,
23 02301, USA

24 ⁵Boston University Schools of Medicine and Public Health, Boston, MA, 02118, USA

25 ⁶VA Boston Healthcare System, Boston, MA, 02118, USA

26 ⁷Salt Lake City VA, VA Informatics & Computing Infrastructure, Salt Lake City, UT, 84148, USA

27 ⁸University of Utah, School of Medicine, Salt Lake City, UT, 84132, USA

28 ⁹Corporal Michael J. Crescenz VA Medical Center, Philadelphia, PA, 19104, USA

29 ¹⁰University of Pennsylvania, Philadelphia, PA, 19104, USA

30 ¹¹Atlanta VA Health Care System, Decatur, GA, 30033, USA

31 ¹²Division of Endocrinology and Metabolism, Department of Medicine, Emory University School of
32 Medicine, Atlanta, GA, 30307, USA

33 ¹³Center of Excellence for Stress and Mental Health, VA San Diego Healthcare System, San Diego, CA,
34 92161, USA

35 ¹⁴Center for Behavior Genetics of Aging, School of Medicine, University of California, San Diego, La Jolla,
36 CA, 92093, USA

37 *Co-Senior Authors

38
39 Abstract: 150 out of 150 words

40 Research in Context: 141 out of 150 words

41 Manuscript Word Count: 3231 out of 3500 words

42 Tables/Figures: 2 tables, 4 figures (out of 6)

43

44

45 **Structured Abstract**

46 INTRODUCTION: Diabetes and dementia are diseases of high healthcare burden worldwide. Individuals
47 with diabetes have 1.4 to 2.2 times higher risk of dementia. Our objective was to evaluate evidence of
48 causality between these two common diseases.

49 METHODS: We conducted a one-sample Mendelian randomization (MR) analysis in the U.S. Department
50 of Veterans Affairs Million Veteran program. The study included 334,672 participants ≥ 65 years of age
51 with type 2 diabetes and dementia case-control status and genotype data.

52 RESULTS: For each standard deviation increase in genetically-predicted diabetes, we found increased
53 odds of three dementia diagnoses in non-Hispanic White participants (all-cause: OR=1.07[1.05-
54 1.08], $P=3.40E-18$; vascular: OR=1.11[1.07-1.15], $P=3.63E-09$, Alzheimer's: OR=1.06[1.02-1.09], $P=6.84E-$
55 04) and non-Hispanic Black participants (all-cause: OR=1.06[1.02-1.10], $P=3.66E-03$, vascular:
56 OR=1.11[1.04-1.19], $P=2.20E-03$, Alzheimer's: OR=1.12 [1.02-1.23], $P=1.60E-02$) but not in Hispanic
57 participants (all $P>.05$).

58
59 DISCUSSION: We found evidence of causality between diabetes and dementia using a one-sample MR
60 study, with access to individual level data, overcoming limitations of prior studies utilizing two-sample
61 MR techniques.

62

63

64

65

66

67

68

69

70

71

72

73

74 **1.0 Background**

75 Diabetes and dementia each impose a high healthcare burden on the U.S. population. As of March 2020,
76 the CDC estimated diabetes prevalence at 14.7%[1] of the adult U.S. population. Separately, a 2019
77 study reported that 11.5% of adults at least 65 years of age had a diagnosis of Alzheimer’s disease (AD)
78 or related dementias (ADRD), representing 1.6% of the U.S. population, a rate projected to double by
79 2060.[2] Observational studies have demonstrated a clear association between diabetes and
80 dementia.[3–7] Individuals with diabetes have a 1.4-2.2 greater relative risk of dementia than those
81 without diabetes,[3,7] with the level of increased risk varying based on the specific dementia diagnosis
82 evaluated.[8,9] Given the projection of 17.9% prevalence of diabetes in the U.S. by 2060,[10] it is critical
83 to gain additional understanding regarding the relationship between diabetes and dementia. The
84 uncertainty is whether the association is due to shared genetic susceptibility, pathophysiology that
85 independently leads to these two common diseases, or if diabetes – which typically occurs at younger
86 ages than dementia – triggers metabolic and/or neurological changes leading to dementia.[11–13] Our
87 previous work in the Million Veteran Program (MVP) demonstrated that a genetic risk score (GRS) for
88 type 2 diabetes (T2D) is associated with all-cause dementia and clinically diagnosed vascular dementia
89 (VaD), and less strongly with clinically diagnosed AD.[14] Like prior observational studies, however, our
90 study strictly assessed an association between T2D and all-cause, VaD, and/or AD without evaluating a
91 causal relationship.

92
93 Mendelian randomization (MR)[15] is an analysis approach that utilizes genetics to assess evidence of
94 causality between an exposure and an outcome. In observational studies, it is difficult to establish
95 causality because many environmental factors can influence both the exposure and outcome. The MR
96 methodology was developed to mitigate these limitations by using genetic variation as an instrumental
97 variable randomized at conception, therefore making the analysis less susceptible to environmental

98 confounders. Given the widespread availability of summary statistics from genome-wide association
99 studies (GWAS), two sample techniques[16] have become popular options for conducting MR analysis.
100 In two-sample MR studies, summary statistics from a GWAS of the exposure are analyzed against
101 summary statistics from a GWAS of the outcome. These techniques are efficient but do not allow for
102 sensitivity tests requiring individual-level data. Two-sample MR studies assessing the causal relationship
103 of T2D and other glycemc traits with the risk of AD or reduced cognitive function have found little
104 evidence that such a relationship exists.[5,17–20] However, the opportunity to apply similar methods to
105 assess causal relationships with outcomes such as vascular dementia has been limited by the lack of
106 published GWAS for this specific diagnosis.[5] One-sample MR is an alternative approach that can take
107 advantage of individual level genotype, exposure, and outcome data in a single study, as long as the
108 sample has enough power to conduct such analysis. Large clinical biobanks which have physician
109 diagnoses and other clinical data are well-suited to this approach.

110

111 Our objective in this study was to conduct a one-sample MR analysis in MVP, the largest clinical biobank
112 in the U.S., to evaluate if diabetes (our exposure) causes dementia (our outcome).

113

114 **2.0 Methods**

115 2.1 Population

116 We conducted our study in MVP,[21] a biobank linking clinical data from the Veterans Affairs Healthcare
117 System (VA) with genotype data on over 650,000 racially/ethnically diverse individuals.[22] The VA
118 Central Institutional Review Board provided approval for the study protocol in accordance with the
119 principles of the Declaration of Helsinki.[23] We analyzed MVP participants who were ≥ 65 years of age
120 at the time of the release of their genotype data, stratified by harmonized ancestry and race/ethnicity
121 classifications (HARE)[24] of Non-Hispanic White (EUR), Non-Hispanic Black (AFR), and Hispanic (HIS).

122 HARE classifications are determined through a self-report of race/ethnicity validated with genetic data.

123

124 2.2 Exposure: Instruments Using Genotype Data

125 We utilized the study sample (N = 334,672) of participants with a coefficient of kinship ≤ 0.088 , whose

126 imputed genotype data was sourced from Release 4 of MVP 1.0 as previously described.[14,25] Our

127 instruments were 331 variants demonstrated to be statistically significant in a T2D genome wide

128 association analysis from the Diabetes Meta-analysis of Trans-ethnic Association Studies (DIAMANTE)

129 Consortium[26] with European effect sizes from Mahajan et al., 2018[27] used in our previous work.[14]

130 We chose not to use a more recent GWAS of type 2 diabetes[23] to avoid overfitting, as detailed in our

131 previous paper.[14] The variants were not in linkage disequilibrium ($R^2 \leq 0.5$),[26] and for our initial

132 steps, we retained the rs429358 variant known to be associated with both diabetes and dementia.[14]

133 2.3 Outcome: Case-Control Definitions of Dementia

134 Our outcomes were three clinical diagnoses of dementia (Figure S1 in the Supplement): all-cause

135 dementia, vascular dementia (VaD), and Alzheimer's Disease (AD), based on International Classification

136 of Diseases 9th (ICD9) or 10th (ICD10) Revision[28] diagnosis codes (Table S1 in the Supplement). In the

137 VA electronic health record, these codes are logged by presiding physicians during routine clinical care.

138 A case was defined as a participant with at least two ICD 9 or 10 codes corresponding to one or more

139 outcomes, as defined by the grouping in Table S1. Note that individuals meeting the case definition of

140 AD and VaD were included as cases for both – that is, AD and VaD were not mutually exclusive. Controls

141 were those participants without a single dementia ICD code in VA clinical records.

142 2.4 Statistical Analysis

143 We compared individuals with and without prevalent T2D using chi-square tests for categorical variables

144 and t-tests for continuous variables. Prevalent T2D cases were defined as individuals whose first

145 diagnosis occurred prior to MVP enrollment. We used MR[15] analysis to evaluate evidence of causality

146 between T2D and clinical diagnoses of dementia. The use of MR requires that three assumptions are
147 satisfied (Figure 1): A. The genetic variants must serve as strong instruments for the exposure
148 (relevance); B. the instrumental variable (IV) must be related to the outcome only through the exposure
149 (exclusion restriction, i.e., no pleiotropy); and C. the IV must not be associated with any confounders
150 between the exposure and outcome (exchangeability).

151 2.5 Two Stage Least Squares Method

152 We used a two stage least squares (2SLS) MR[29] approach to estimate the causal associations between
153 T2D and three clinical diagnoses of dementia. The 2SLS method employs two regression equations

$$\text{Stage 1: } x_1 = \gamma_1 + \pi_1 G_1 + v_1$$

$$\text{Stage 2: } y = \alpha + \beta_1 \hat{x}_1 + u$$

154 where stage 1 evaluates the genetic risk with the exposure and stage 2 estimates the prediction of the
155 exposure with the outcome. For our genetic risk, we regressed a standardized weighted score of valid
156 genetic instruments and effect sizes from published GWAS[30] against the log odds of T2D in
157 the MVP biobank, adjusting for 10 genetic principal components. In stage 2, we evaluated the
158 association of the standardized genetically predicted probability of T2D (referred to as ‘genetically
159 predicted T2D risk’ going forward) from stage 1 with our three outcomes of dementia using logistic
160 regression adjusting for age, and self-reported biological sex. To account for the uncertainty inherent in
161 the first stage of the analysis, we utilized the ‘ivtools’ methodology introduced in Sjölander and
162 Martinussen[31] to provide estimates for nonlinear models. In more precise terms, we specified the ‘ts’
163 estimation method in the R function ‘ivglm’ supplying the regression model output from the two stages
164 and reported the estimates resulting from this function. The beta coefficient estimates can be
165 interpreted as the log odds of dementia per standard deviation increase in genetically predicted T2D

166 risk. All models were stratified by EUR, AFR, and HIS, with a significance threshold of 0.017 (0.05/3) to
167 account for multiple testing.

168

169 2.6 Quartiles of Genetic Risk

170 To evaluate the impact of increasing genetic risk of diabetes on the three dementia diagnoses, we also
171 assessed the difference in the odds of dementia by quartiles of diabetes genetic risk. That is, we divided
172 the genetically predicted T2D risk from the first stage of our analysis into quartiles and evaluated the top
173 three quartiles of risk against the reference quartile, the one with the lowest genetic risk of T2D. In this
174 evaluation, the quartile was treated as a factor input to the second stage of the model and again, we
175 used the ‘ivglm’ method to account for uncertainty in the first stage. Additionally, we assessed potential
176 trends in the quartiles of progressive genetic risk using a Cochran Armitage trend test.[32,33]

177

178 2.7 Summarized Data and Pleiotropy Assessment

179 We also assessed the causal association using MR methods built to analyze effects when conducting
180 two-sample MR. We calculated estimates using the inverse-weighted (IVW), median, and MR Egger[15]
181 methods recommended as part of the MendelianRandomization[34] R package (v0.3.0). These methods
182 provide effect estimates and standard errors based on summary measures of the relationship with the
183 exposure versus the relationship with the outcome. The methods also assess evidence of pleiotropy.

184 We additionally used MR-PRESSO[35] to test for outliers in the list of valid genetic instruments (R
185 package version 1.0).[35] Details of these methods can be found in the Supplement (Methods S1).

186

187 2.8 Addressing Mendelian Randomization Assumptions

188 We used linear regression to evaluate the weighted instrument association strength with diabetes status
189 by race/ethnicity population to test assumption A. An F-statistic in the linear relationship between the
190 instrument and the exposure >10 is conventionally accepted as an indicator of a strong instrument[36].
191 To evaluate pleiotropy (assumption B), we used the MR-Egger method from the
192 MendelianRandomization[15,37,38] package in R. MR-Egger constructs the best fit line in the
193 relationship between the exposure effects and outcome effects. Pleiotropy is concluded to occur if the
194 intercept of the slope of this relationship differs from zero. We also used MR_PRESSO[35] to check for
195 pleiotropy. MR_PRESSO uses a leave-one out strategy to test for the impact of any single variant being
196 considered as an instrument. If we found that any variants violated the no pleiotropy assumption, we
197 removed those variants from the list of valid instruments for this analysis. To address assumption C, we
198 used the recommendation by Vanderweele, et. al., [39,40] to calculate E-Values. An E-Value puts a
199 boundary on the potential bias introduced by unmeasured confounders. The E-Value is the effect size of
200 an unmeasured confounder, with both the exposure and the outcome, required to explain away the
201 exposure-outcome effect.

202 **3.0 Results**

203 We had 82,980 participants with a T2D diagnosis and 251,692 without for a sample size of 334,672 that
204 met our study inclusion criteria (Table 1). Compared to those without T2D, those with T2D were younger
205 (74.0 versus 74.7 years), less likely to be female (2.5% versus 3.2%), and more likely to be AFR (17.4%
206 versus 11.4%) or HIS (7.5% versus 4.7%). All-cause dementia and VaD cases were more prevalent in
207 those with T2D than in those without: 8.7% vs 6.7%, and 2.0% vs 1.0%, respectively. Prevalence of AD
208 was similar between those with/without T2D (1.4% vs 1.3%).

209

210 **3.1 Two Stage Least Squares Results**

211 From the 331 initial variants, we identified 330 that met our criteria for valid instruments. For the
212 remainder of our discussion, we will call the standardized weighted combination of these variants
213 ‘GRS330’. In stage 1 of the 2SLS model, the odds of diabetes were significantly associated with GRS330
214 for all three HARE groups (OR of T2D per standard deviation increase in GRS330 in EUR: 1.50, $P=4.6e-$
215 308 ; AFR: 1.28, $P=5.2e-125$; HIS: 1.46, $P=3.2e-118$). The standardized predicted probability of T2D risk
216 from stage 1 (we will call this value ‘GRS330_Predicted’) was calculated for each individual and the
217 distribution of this value differed by T2D status (Figure 2).

218 GRS330_Predicted was associated with all dementia diagnoses (Table 2) in EUR. For each standard
219 deviation increase in GRS330_Predicted, the odds of all-cause dementia increased by 1.07 (95% CI: 1.05-
220 1.08, $P=3.40E-18$). We found similar results for VaD (OR=1.11, 95% CI: 1.07-1.15, $P=3.63E-09$) and AD
221 (OR=1.06, 95% CI: 1.02-1.09, $P=6.84E-04$). In AFR, GRS_Predicted was associated (Table 2) with all-cause
222 dementia (OR=1.06, 95% CI: 1.02-1.10, $P=3.66E-03$), VaD (OR=1.11, 95% CI: 1.04-1.19, $P=2.20E-03$) and
223 AD (OR = 1.12, 95%CI: 1.02-1.23, $P=1.60E-02$). For HIS, there were no statistically significant associations
224 (Table 2) between GRS330_Predicted and any dementia diagnosis (all $P>.05$).

225

226 3.2 Quartile Analysis Results

227 To demonstrate the impact of higher genetic liability for type 2 diabetes, we have also shown the risk of
228 dementia diagnosis by quartile of GRS330_Predicted (Figure 3). In EUR, the 4th quartile of
229 GRS330_Predicted increased the odds of all dementia diagnoses as compared to the 1st quartile (Figure
230 3a): all-cause dementia (OR=1.18, 95% CI: 1.14-1.24, $P=3.82E-15$), VaD (OR=1.33, 95% CI: 1.20-1.47,
231 $P=3.08E-08$), and AD (OR=1.14, 95% CI: 1.03-1.25, $P=8.13E-03$). We observed similar results for the
232 comparison of the 4th quartile of genetically-predicted T2D risk to the 1st in AFR (Figure 3b) for all-cause-
233 dementia (OR=1.18, 95% CI: 1.06-1.32, $P=3.33E-03$) and VaD (OR=1.42, 95% CI: 1.16-1.74, $P=6.63E-03$),

234 but not for AD ($P=.07$). In HIS, none of the quartile associations with dementia diagnoses were
235 statistically significant (all $P>.18$).

236
237 The Cochran-Armitage[32,41] test (Table S2) for the trend in the relationship between
238 GRS330_Predicted quartiles and dementia diagnoses was significant for all-cause dementia ($P=3.6e-05$)
239 and VaD ($P=3.6e-05$) but not AD ($P=.82$) in EUR. The trend tests for GRS330_Predicted quartiles and
240 dementia were not significant for any diagnosis in AFR (all $P>.05$) or in HIS.

241
242 3.3 Summarized Data Results
243 When utilizing the two sample MR approaches, we used the weighted combination of the instruments,
244 GRS330, for the exposure. All methods for assessing the association between GRS330 and all-cause
245 dementia in EUR provided consistent directions of effect with the primary 2SLS analyses and significant
246 associations (Figure 4a, Table S3 in the Supplement). The results for AFR (Figure 4b) and HIS (Figure 4c)
247 were directionally consistent with those of EUR except for the MR Egger approach in AFR (Figure 4b),
248 though the associations were not always statistically significant (Table S3 in the Supplement). Similarly,
249 the two sample methods provided consistent results with the primary one-sample analysis for the
250 dementia subtypes, though not always achieving statistical significance (See Results S1 in the
251 Supplement for full details).

252
253 3.4 Assessment of Mendelian Randomization Assumptions
254 Assumption A. Strong Instrument (relevance): The combination of the final 330 valid instruments
255 (GRS330) demonstrated a significant association with diabetes. The F-statistics in the linear associations
256 between the instruments and diabetes were 61.8 for HIS, 57.3 for AFR, and 740.4 for EUR.

257 Assumption B. Pleiotropy (exclusion restriction): The MR Egger test with GRS330 did not exhibit
258 pleiotropy ($P > 0.05$, Table S4 in the Supplement). As expected, however, when additionally including the
259 rs429358 variant known to be associated with both T2D and dementia, heterogeneity was detected
260 (Table S4 in the Supplement, all $P < .017$) except in the case of HIS with VaD. This heterogeneity was
261 attenuated when removing rs429358 from the test (all $P > .01$). The MR_PRESSO test (Table S5 in the
262 Supplement) characterized rs429358 as an outlier in the relationship between the exposure effect and
263 outcome effect for all HARE groups and all clinical dementia diagnoses except HIS with VaD. Given the
264 results of the pleiotropy and heterogeneity assessment, we excluded rs429358 and used GRS330 as our
265 valid instrument.

266 Assumption C. Unmeasured Confounders (exchangeability): The sensitivity E-Value calculations were
267 based on our previous work[14] in which we calculated odds ratios for the association of GRS330 with
268 each clinical diagnosis of dementia stratified by HARE group. We used the risk ratio formula
269 recommended by VanderWeele and Ding[40] for the scenario in which disease prevalence is less than
270 15%. As the odds ratios ranged from 1.02 to 1.15, the E-Value ranged from 1.16 to 1.57 with the E-Value
271 of the lower bound of the 95% confidence interval ranging from 1.0 to 1.4. (Table S6 in the Supplement).

272 **4.0 Discussion**

273 We found evidence of causality between T2D and all-cause dementia as well as clinically diagnosed VaD
274 in EUR, with similar results in AFR. The effect estimates in HIS were similar to EUR and AFR but without
275 statistical significance. The relationship between T2D and clinically diagnosed AD had a reduced effect
276 estimate in EUR in comparison to all-cause dementia but an increased effect estimate in AFR, indicating
277 a need for further research to elucidate the differences by HARE group.

278

279 Given the wealth of observational data linking T2D and dementia, researchers have hypothesized a
280 causal relationship between the two but have found little evidence to support such a conclusion. In

281 work focused on AD as an endpoint, several studies found no evidence of a relationship with genetically
282 predicted T2D risk: a Danish study by Thomassen et. al.,[5] the International Genomics of Alzheimer’s
283 Project by Østergaard et. al.,[18] the Health and Retirement Study (HRS) by Walter et al.,[17] and
284 DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) and Meta-Analyses of Glucose and Insulin-
285 related traits Consortium (MAGIC) by Pan et. al.[42] When focused on general cognitive impairment as
286 the outcome, Ware et. al.[20] (HRS) and Garfield[19] (UK Biobank) had similar results. These results
287 could be explained by the lack of broad genetics studies for other dementia diagnoses such as vascular
288 dementia, as noted by Thomassen et.al.,[5]. Our study, which did identify evidence of causality, was able
289 to assess multiple diagnoses of dementia using recently reported T2D genome-wide significant variants.

290
291 While the previous studies linking the broad spectrum of T2D with AD did not report evidence
292 supporting causality, studies deciphering mechanisms related to diabetes have been able to offer
293 further insight. Walter et. al.,[17] discovered that genetically-predicted insulin sensitivity was causally
294 associated with AD in HRS. In a like manner, Pan et. al.,[42] detected a causal relationship when
295 assessing genetic instruments for higher fasting glucose and lower HOMA- β -cell function with AD using a
296 two-sample MR approach across the DIAGRAM and MAGIC consortia. Tschritter et al.,[12] demonstrated
297 a cortical activity response difference to insulin infusion between carriers and non-carriers of the 972Arg
298 variant of the *IRS-1* gene known to be associated with type 2 diabetes[43], as well as a similar response
299 difference between lean (insulin-sensitive) and obese (insulin-resistant) individuals. These studies
300 suggest future research directions that delve into the potential metabolic pathways through which
301 diabetes might lead to dementia.

302
303 Previous studies assessing the relationship between T2D and AD were focused on populations with
304 European ancestry. The MVP biobank is well powered to assess associations in an African American

305 population. In particular, we discovered evidence of causality between T2D and clinically diagnosed VaD
306 in AFR to have a similar magnitude of effect to EUR even though the sample size was smaller. The
307 higher effect sizes for clinically diagnosed AD (Table 2) and the outlier effect observed with all-cause
308 dementia (Figure 4b) in this population suggest an area for focused follow-up in future research.

309
310 We had several limitations in our study. The small number of female participants does not allow an
311 assessment of distinctive results by biological sex. Given the small proportion of females, our results are
312 largely driven by male participants, and so cannot be generalized to both sexes. Our outcomes were
313 determined based on a physician's diagnosis which may cause some amount of imprecision. Our T2D
314 weighting structure was based on a previous study of European ancestry that might overlook important
315 genetic architecture specific to the African American and Hispanic populations.

316
317 Despite these limitations, we conclude from our one-sample MR study in a large US biobank that there is
318 evidence of a causal association of diabetes with dementia. Moreover, the difference in the strength of
319 effect in the association of vascular dementia versus AD in the European population is suggestive of
320 different mechanisms. The strength of association for both vascular dementia and AD in the African
321 population requires further investigation. Establishing a causal association represents a first step
322 toward examining the potential impact of the expanding prevalence of diabetes on dementia incidence
323 and whether diabetes prevention and/or treatment can mitigate dementia risk.

324

325 References

- 326 1. Prevalence of Both Diagnosed and Undiagnosed Diabetes | Diabetes | CDC [Internet]. 2020 [cited
327 2020 Oct 13]. Available from: [https://www.cdc.gov/diabetes/data/statistics-report/diagnosed-](https://www.cdc.gov/diabetes/data/statistics-report/diagnosed-undiagnosed-diabetes.html)
328 [undiagnosed-diabetes.html](https://www.cdc.gov/diabetes/data/statistics-report/diagnosed-undiagnosed-diabetes.html)
- 329 2. Matthews KA, Xu W, Gaglioti AH, Holt JB, Croft JB, Mack D, et al. Racial and ethnic estimates of
330 Alzheimer's disease and related dementias in the United States (2015–2060) in adults aged ≥65
331 years. *Alzheimers Dement J Alzheimers Assoc.* 2019 Jan;15(1):17–24.
- 332 3. Biessels GJ, Staekenborg S, Brunner E, Brayne C, Scheltens P. Risk of dementia in diabetes mellitus:
333 a systematic review. *Lancet Neurol.* 2006 Jan;5(1):64–74.
- 334 4. Xue M, Xu W, Ou YN, Cao XP, Tan MS, Tan L, et al. Diabetes mellitus and risks of cognitive
335 impairment and dementia: A systematic review and meta-analysis of 144 prospective studies.
336 *Ageing Res Rev.* 2019 Nov;55:100944.
- 337 5. Thomassen JQ, Tolstrup JS, Benn M, Frikke-Schmidt R. Type-2 diabetes and risk of dementia:
338 observational and Mendelian randomisation studies in 1 million individuals. *Epidemiol Psychiatr*
339 *Sci.* 2020 Apr 24;29:e118.
- 340 6. Exalto LG, Biessels GJ, Karter AJ, Huang ES, Katon WJ, Minkoff JR, et al. Risk score for prediction of
341 10 year dementia risk in individuals with type 2 diabetes: a cohort study. *Lancet Diabetes*
342 *Endocrinol.* 2013 Nov;1(3):183–90.
- 343 7. Luchsinger JA. Diabetes, related conditions, and dementia. *J Neurol Sci.* 2010 Dec 15;299(1–2):35–
344 8.
- 345 8. Hadley G, Zhang J, Harris-Skillman E, Alexopoulou Z, DeLuca GC, Pendlebury ST. Cognitive decline
346 and diabetes: a systematic review of the neuropathological correlates accounting for cognition at
347 death. *J Neurol Neurosurg Psychiatry.* 2022 Mar;93(3):246–53.
- 348 9. Biessels GJ, Despa F. Cognitive decline and dementia in diabetes: mechanisms and clinical
349 implications. *Nat Rev Endocrinol.* 2018 Oct;14(10):591–604.
- 350 10. Lin J, Thompson TJ, Cheng YJ, Zhuo X, Zhang P, Gregg E, et al. Projection of the future diabetes
351 burden in the United States through 2060. *Popul Health Metr.* 2018 Jun 15;16(1):9.
- 352 11. Bello-Chavolla OY, Antonio-Villa NE, Vargas-Vázquez A, Ávila-Funes JA, Aguilar-Salinas CA.
353 Pathophysiological Mechanisms Linking Type 2 Diabetes and Dementia: Review of Evidence from
354 Clinical, Translational and Epidemiological Research. *Curr Diabetes Rev.* 2019;15(6):456–70.
- 355 12. Tschritter O, Preissl H, Hennige AM, Stumvoll M, Porubská K, Frost R, et al. The cerebrocortical
356 response to hyperinsulinemia is reduced in overweight humans: a magnetoencephalographic
357 study. *Proc Natl Acad Sci U S A.* 2006 Aug 8;103(32):12103–8.
- 358 13. Yoo DY, Yim HS, Jung HY, Nam SM, Kim JW, Choi JH, et al. Chronic type 2 diabetes reduces the
359 integrity of the blood-brain barrier by reducing tight junction proteins in the hippocampus. *J Vet*
360 *Med Sci.* 2016 Jul 1;78(6):957–62.

- 361 14. Litkowski EM, Logue MW, Zhang R, Charest BR, Lange EM, Hokanson JE, et al. A Diabetes Genetic
362 Risk Score Is Associated With All-Cause Dementia and Clinically Diagnosed Vascular Dementia in
363 the Million Veteran Program. *Diabetes Care*. 2022 Aug 30;dc220105.
- 364 15. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect
365 estimation and bias detection through Egger regression. *Int J Epidemiol*. 2015 Apr;44(2):512–25.
- 366 16. Burgess S, Scott RA, Timpson NJ, Davey Smith G, Thompson SG, EPIC- InterAct Consortium. Using
367 published data in Mendelian randomization: a blueprint for efficient identification of causal risk
368 factors. *Eur J Epidemiol*. 2015 Jul;30(7):543–52.
- 369 17. Walter S, Marden JR, Kubzansky LD, Mayeda ER, Crane PK, Chang SC, et al. Diabetic Phenotypes
370 and Late-Life Dementia Risk: A Mechanism-specific Mendelian Randomization Study. *Alzheimer Dis
371 Assoc Disord*. 2016 Mar;30(1):15–20.
- 372 18. Østergaard SD, Mukherjee S, Sharp SJ, Proitsi P, Lotta LA, Day F, et al. Associations between
373 Potentially Modifiable Risk Factors and Alzheimer Disease: A Mendelian Randomization Study.
374 *PLoS Med*. 2015 Jun;12(6):e1001841; discussion e1001841.
- 375 19. Garfield V, Farmaki AE, Fatemifar G, Eastwood SV, Mathur R, Rentsch CT, et al. Relationship
376 Between Glycemia and Cognitive Function, Structural Brain Outcomes, and Dementia: A Mendelian
377 Randomization Study in the UK Biobank. *Diabetes*. 2021 Oct 1;70(10):2313–21.
- 378 20. Ware EB, Morataya C, Fu M, Bakulski KM. Type 2 Diabetes and Cognitive Status in the Health and
379 Retirement Study: A Mendelian Randomization Approach. *Front Genet*. 2021;12:634767.
- 380 21. Gaziano JM, Concato J, Brophy M, Fiore L, Pyarajan S, Breeling J, et al. Million Veteran Program: A
381 mega-biobank to study genetic influences on health and disease. *J Clin Epidemiol*. 2016
382 Feb;70:214–23.
- 383 22. Million Veteran Program (MVP) [Internet]. [cited 2020 Jun 11]. Available from:
384 <https://www.research.va.gov/mvp/>
- 385 23. Vujkovic M, Keaton JM, Lynch JA, Miller DR, Zhou J, Tcheandjieu C, et al. Discovery of 318 new risk
386 loci for type 2 diabetes and related vascular outcomes among 1.4 million participants in a multi-
387 ancestry meta-analysis. *Nat Genet*. 2020 Jun 15;
- 388 24. Fang H, Hui Q, Lynch J, Honerlaw J, Assimes TL, Huang J, et al. Harmonizing Genetic Ancestry and
389 Self-identified Race/Ethnicity in Genome-wide Association Studies. *Am J Hum Genet*. 2019
390 03;105(4):763–72.
- 391 25. Hunter-Zinck H, Shi Y, Li M, Gorman BR, Ji SG, Sun N, et al. Genotyping Array Design and Data
392 Quality Control in the Million Veteran Program. *Am J Hum Genet*. 2020 Apr 2;106(4):535–48.
- 393 26. Mahajan A, Spracklen CN, Zhang W, Ng MCY, Petty LE, Kitajima H, et al. Multi-ancestry genetic
394 study of type 2 diabetes highlights the power of diverse populations for discovery and translation.
395 *Nat Genet*. 2022 May;54(5):560–72.

- 396 27. Mahajan A, Taliun D, Thurner M, Robertson NR, Torres JM, Rayner NW, et al. Fine-mapping type 2
397 diabetes loci to single-variant resolution using high-density imputation and islet-specific
398 epigenome maps. *Nat Genet.* 2018;50(11):1505–13.
- 399 28. International Classification of Diseases (ICD) [Internet]. [cited 2021 Dec 13]. Available from:
400 <https://www.who.int/standards/classifications/classification-of-diseases>
- 401 29. Sanderson E, Davey Smith G, Windmeijer F, Bowden J. An examination of multivariable Mendelian
402 randomization in the single-sample and two-sample summary data settings. *Int J Epidemiol.* 2019
403 Jun 1;48(3):713–27.
- 404 30. Mahajan A, Taliun D, Thurner M, Robertson NR, Torres JM, Rayner NW, et al. Fine-mapping type 2
405 diabetes loci to single-variant resolution using high-density imputation and islet-specific
406 epigenome maps. *Nat Genet.* 2018 Nov;50(11):1505–13.
- 407 31. Sjolander A, Martinussen T. Instrumental Variable Estimation with the R Package ivtools. *Epidemiol
408 Methods* [Internet]. 2019 Dec 1 [cited 2023 Feb 7];8(1). Available from:
409 <https://www.degruyter.com/document/doi/10.1515/em-2018-0024/html?lang=en>
- 410 32. Cochran WG. Some Methods for Strengthening the Common χ^2 Tests. *Biometrics.* 1954;10(4):417–
411 51.
- 412 33. McKenzie JE, Salanti G, Lewis SC, Altman DG. Meta-analysis and The Cochrane Collaboration: 20
413 years of the Cochrane Statistical Methods Group. *Syst Rev.* 2013 Nov 26;2:80.
- 414 34. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent Estimation in Mendelian
415 Randomization with Some Invalid Instruments Using a Weighted Median Estimator. *Genet
416 Epidemiol.* 2016 May;40(4):304–14.
- 417 35. Verbanck M, Chen CY, Neale B, Do R. Detection of widespread horizontal pleiotropy in causal
418 relationships inferred from Mendelian randomization between complex traits and diseases. *Nat
419 Genet.* 2018 May;50(5):693–8.
- 420 36. Staiger D, Stock JH. Instrumental Variables Regression with Weak Instruments. *Econometrica.*
421 1997;65(3):557–86.
- 422 37. Burgess S, Small DS, Thompson SG. A review of instrumental variable estimators for Mendelian
423 randomization. *Stat Methods Med Res.* 2017 Oct;26(5):2333–55.
- 424 38. Burgess S, Thompson SG. Interpreting findings from Mendelian randomization using the MR-Egger
425 method. *Eur J Epidemiol.* 2017;32(5):377–89.
- 426 39. Swanson SA, VanderWeele TJ. E-Values for Mendelian Randomization. *Epidemiology.* 2020
427 May;31(3):e23–4.
- 428 40. VanderWeele TJ, Ding P. Sensitivity Analysis in Observational Research: Introducing the E-Value.
429 *Ann Intern Med.* 2017 Aug 15;167(4):268.

- 430 41. Armitage P. Tests for Linear Trends in Proportions and Frequencies. *Biometrics*. 1955;11(3):375–
431 86.
- 432 42. Pan Y, Chen W, Yan H, Wang M, Xiang X. Glycemic traits and Alzheimer’s disease: a Mendelian
433 randomization study. *Aging*. 2020 Nov 16;12(22):22688–99.
- 434 43. Jellema A, Zeegers MPA, Feskens EJM, Dagnelie PC, Mensink RP. Gly972Arg variant in the insulin
435 receptor substrate-1 gene and association with Type 2 diabetes: a meta-analysis of 27 studies.
436 *Diabetologia*. 2003 Jul;46(7):990–5.

437

438 **Figure Legends**

439 Figure 1. Mendelian Randomization Assumptions: The required assumptions to ensure the validity of the
440 Mendelian Randomization Analysis Approach

441

442 Figure 2. Predicted Probability: Distribution of the standardized predicted probability of GRS330
443 (excluding the rs429358 variant) in those with and without diabetes stratified by race/ethnicity
444 classification.

445

446 Figure 3. Risk Quartiles: Quartiles, listed in parentheses, of genetically predicted risk of diabetes
447 (GRS330_Predicted) associated with increased risk of dementia diagnosis for all-cause dementia
448 (orange), clinically diagnosed VaD (blue) and clinically diagnosed AD (green) in a) EUR: European
449 population, b) AFR: African population, c) HIS: Hispanic population

450 Figure 4. Mendelian Randomization Methods: Methods using the Mendelian Randomization package are
451 shown here. Simple estimates have no weighting and weighted estimates use standard errors for
452 weighting factors. a) Associations between GRS330 and all-cause dementia in EUR, b) Associations
453 between GRS330 and all-cause dementia in AFR, and c) Associations between GRS330 and all-cause
454 dementia in HIS.

455 Acknowledgements

456 Author Contributions: EML designed the study, performed the literature review, conducted the analysis,
457 and wrote the manuscript. MWL contributed to the dementia phenotype definitions and edited the
458 manuscript. RZ implemented the dementia phenotypes. BRC implemented the diabetes phenotype.
459 EML(2) provided guidance on statistical methods and edited the document. JEH provided guidance on
460 applicable epidemiology concepts and study design. JAL designed the dementia phenotype groupings.
461 MV provided epidemiology advice on the MR study design. LSP designed the diabetes complications
462 study. LAL provided epidemiology, study structure and writing guidance. RLH provided dementia subject
463 matter expertise. SR offered diabetes subject matter expertise, study design and analysis guidance, and
464 edited the manuscript. EML and SR are the guarantors of this work and, as such, had full access to all the
465 data in this study and take responsibility for the integrity of the data and accuracy of the data analysis.
466 The corresponding author attests that all listed authors meet authorship criteria and that no others
467 meeting the criteria have been omitted.

468 Preliminary results of this study were presented as a poster: E.Litkowski, et al., Evidence of causality
469 between type 2 diabetes and dementia in the Million Veteran Program; Abstract/Program 3896.
470 Presented at the 71st Annual Meeting of The American Society of Human Genetics, October 18, 2021,
471 Virtual.

472 Funding: EML is supported by US National Institutes of Health award P30DK116073, and by funds from
473 the Boettcher Foundation's Webb-Waring Biomedical Research Program. Phenotype development in
474 MVP was supported by US Department of Veterans Affairs award BLR&D 1 I01BX004192 (MWL PI). LSP is
475 supported in part by VA awards CSP #2008, I01 CX001899, I01 CX001737, and I01 BX005831; NIH awards
476 R01 DK127083, R03 AI133172, R21 AI156161, U01 DK098246, UL1 TR002378; and a Cystic Fibrosis
477 Foundation award PHILL12A0. RLH is supported by the Million Veteran Program MVP022 award # I01
478 CX001727, VISN-22 VA Center of Excellence for Stress and Mental Health (CESAMH), and National
479 Institute of Aging RO1 grants AG050595 (*The VETSA Longitudinal Twin Study of Cognition and Aging*
480 *VETSA 4*), AG05064 (*Effects of Androgen Deprivation Therapy on Preclinical Symptoms of Alzheimer's*
481 *Disease*), and AG065385 (*Novel Antagonists of the N-terminal Domain of the CRF Receptor Type 1 for*
482 *Alzheimer's Disease*). SR is supported by US Department of Veterans Affairs award IK2-CX001907, by US
483 National Institutes of Health award P30DK116073, and by funds from the Boettcher Foundation's Webb-
484 Waring Biomedical Research Program. This research is based on data from the Million Veteran Program,
485 Office of Research and Development, Veterans Health Administration, and was supported by awards
486 MVP003, MVP009, MVP015, and MVP022. This publication does not represent the views of the
487 Department of Veteran Affairs or the United States Government.

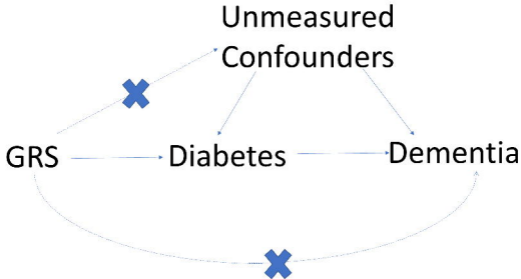
488 Conflicts of Interest: Within the past several years LSP has served on Scientific Advisory Boards for
489 Janssen, and the Profil Institute for Clinical Research, and has or had research support from Merck,
490 Amylin, Eli Lilly, Novo Nordisk, Sanofi, PhaseBio, Roche, Abbvie, Vascular Pharmaceuticals, Janssen,
491 Glaxo SmithKline, Pfizer, and the Cystic Fibrosis Foundation. In the past, LSP was a speaker for Novartis
492 and Merck, but not for the last five years. LSP is also a cofounder and Officer and Board member and
493 stockholder of a company, DIASYST, Inc., which is developing software aimed to help improve diabetes
494 management. SR has previously received research grant funding from the American Heart Association.
495 EML, MWL, RZ, BRC, EML(2), JEH, JAL, MV, LAL, RLH have no conflicts to disclose.

496 **Consent Statement**

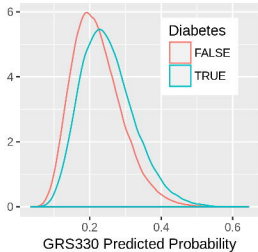
497 MVP participants went through a counseling process before they enrolled and provided consent to have
498 their electronic health records reviewed. The VA Central Institutional Review Board gave approval for
499 the study protocol in accordance with the principles of the Declaration of Helsinki.

500

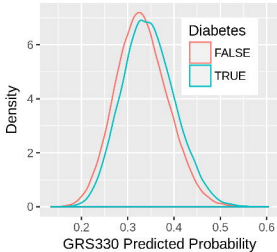
501 **Key Words:** Type 2 Diabetes, Vascular Dementia, Alzheimer’s Disease, Genetics, Mendelian
502 Randomization, Instrumental Variable, All-cause dementia



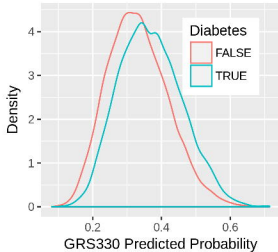
EUR Population



AFR Population



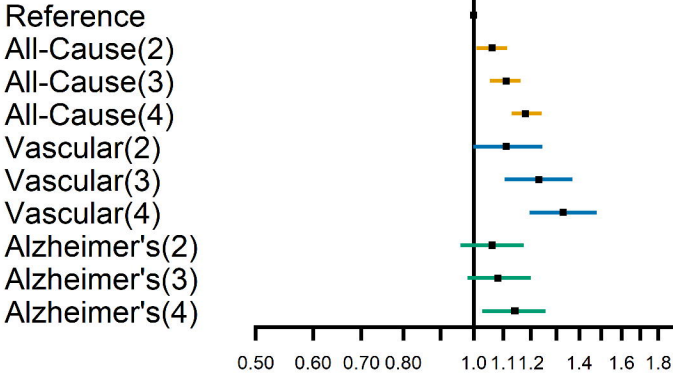
HIS Population



a) EUR Population: Odds of Dementia with GRS Quartiles

Diagnosis

OR [95% CI], p-value

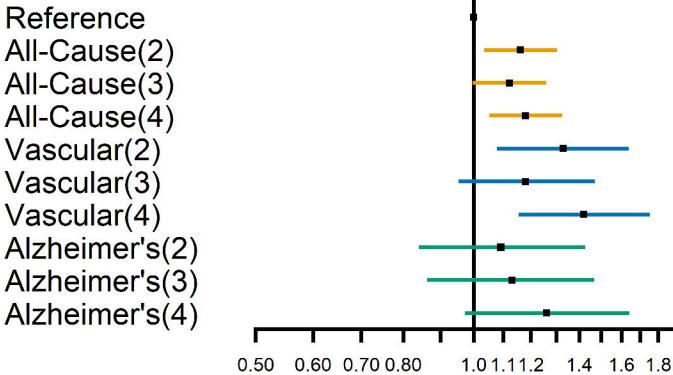


Diagnosis	OR [95% CI]	p-value
Reference	NA	NA
All-Cause(2)	1.06 [1.02, 1.11]	6.98e-03
All-Cause(3)	1.11 [1.06, 1.15]	3.15e-06
All-Cause(4)	1.18 [1.14, 1.24]	3.82e-15
Vascular(2)	1.11 [1.01, 1.24]	3.96e-02
Vascular(3)	1.23 [1.11, 1.36]	7.31e-05
Vascular(4)	1.33 [1.20, 1.47]	3.08e-08
Alzheimer's(2)	1.06 [0.96, 1.17]	2.28e-01
Alzheimer's(3)	1.08 [0.99, 1.19]	9.58e-02
Alzheimer's(4)	1.14 [1.03, 1.25]	8.13e-03

b) AFR Population: Odds of Dementia with GRS Quartiles

Diagnosis

OR [95% CI], p-value

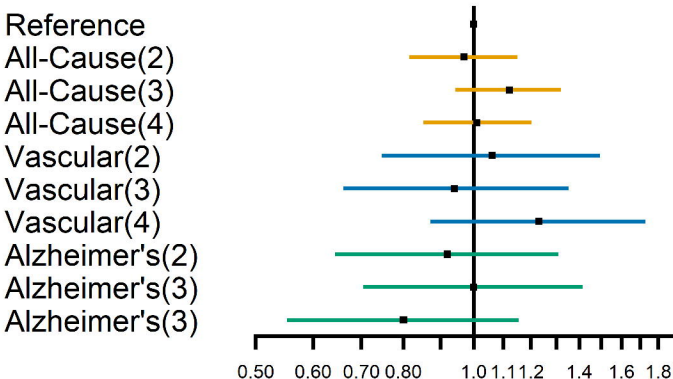


Diagnosis	OR [95% CI]	p-value
Reference	NA	NA
All-Cause(2)	1.16 [1.04, 1.30]	7.95e-03
All-Cause(3)	1.12 [1.00, 1.25]	4.47e-02
All-Cause(4)	1.18 [1.06, 1.32]	3.33e-03
Vascular(2)	1.33 [1.08, 1.63]	6.41e-03
Vascular(3)	1.18 [0.96, 1.46]	1.15e-01
Vascular(4)	1.42 [1.16, 1.74]	6.63e-04
Alzheimer's(2)	1.09 [0.85, 1.42]	4.92e-01
Alzheimer's(3)	1.13 [0.87, 1.46]	3.73e-01
Alzheimer's(4)	1.26 [0.98, 1.63]	7.29e-02

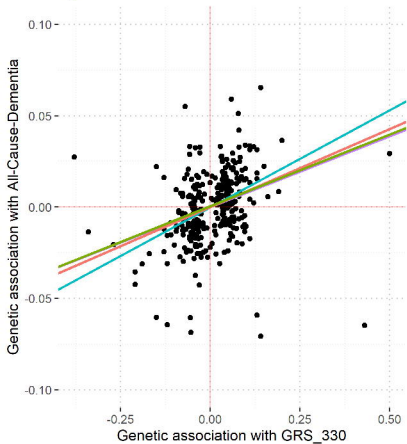
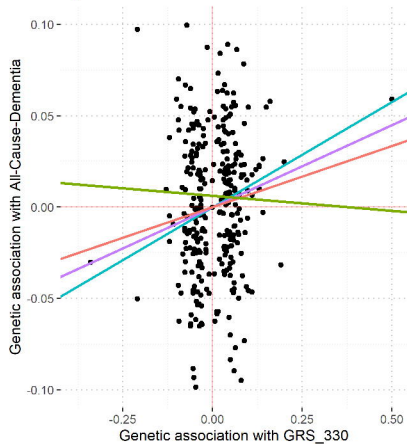
c) HIS Population: Odds of Dementia with GRS Quartiles

Diagnosis

OR [95% CI], p-value



Diagnosis	OR [95% CI]	p-value
Reference	NA	NA
All-Cause(2)	0.97 [0.82, 1.14]	6.96e-01
All-Cause(3)	1.12 [0.95, 1.31]	1.82e-01
All-Cause(4)	1.01 [0.86, 1.19]	8.84e-01
Vascular(2)	1.06 [0.75, 1.49]	7.52e-01
Vascular(3)	0.94 [0.66, 1.34]	7.52e-01
Vascular(4)	1.23 [0.88, 1.72]	2.33e-01
Alzheimer's(2)	0.92 [0.65, 1.30]	6.30e-01
Alzheimer's(3)	1.00 [0.71, 1.40]	9.86e-01
Alzheimer's(3)	0.80 [0.56, 1.15]	2.24e-01

a) EUR**b) AFR****c) HIS**