



Contents lists available at ScienceDirect

Neurobiology of Aging

journal homepage: www.elsevier.com/locate/neuaging

Brief communication

Polygenic score prediction captures nearly all common genetic risk for Alzheimer's disease

Valentina Escott-Price^{a,*}, Maryam Shoai^b, Richard Pither^c, Julie Williams^a, John Hardy^{b,**}^a Institute of Psychological Medicine and Clinical Neurosciences, MRC Centre for Neuropsychiatric Genetics & Genomics, Cardiff University, Cardiff, UK^b Institute of Neurology, University College London, London, UK^c Cytox Ltd, UK, Oxford, UK

ARTICLE INFO

Article history:

Received 3 May 2016

Received in revised form 14 July 2016

Accepted 15 July 2016

Keywords:

AUC

Polygenic score

Heritability

ABSTRACT

We estimate the maximum prediction accuracy for the risk of Alzheimer's disease based on disease prevalence and heritability of liability. We demonstrate that the recently reported AUC values for predicting of Alzheimer's disease using polygenic scores reach about 90% of the estimated maximum accuracy that can be achieved by predictors of genetic risk based on genomic profiles.

© 2016 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

With increasing longevity due to better medical care, the prevalence of Alzheimer's disease (AD) is increasing. Hope for finding preventative and clinical therapies lies in the ability to gain a better understanding of the underlying biology of the disease, and genetics can provide a valuable starting point for advancement. Despite enormous efforts across the research community, and the successful identification of dozens of risk loci identified genes that associate with AD, the impact on therapies or prevention is negligible.

Rare monogenic forms of AD, the majority of which are attributable to mutations in 1 of 3 genes, *APP*, *PSEN1*, and *PSEN2*, exist, but common, late-onset AD is genetically complex. Along with the *APOE* polymorphism (Corder et al., 1993), 20 common susceptibility loci (*ABCA7*, *BIN1*, *CD33*, *CLU*, *CR1*, *CD2AP*, *EPHA1*, *MS4A4*, *PICALM*, *HLA-DRB5/DRB1*, *SORL1*, *PTK2B*, *SLC24A4/RIN3*, *ZCWPW1*, *SELF1*, *NME8*, *FERMT2*, *CASS4*, *INPP5D*, *MEF2C*, *CD33*) have been identified to be associated with AD (Harold et al., 2009; Hollingworth et al., 2011; Lambert et al., 2009, 2013; Naj et al.,

2011; Seshadri et al., 2010). Two additional genes *TP53INP1* and *IGHV1-67* were reported as gene-based genome-wide significant, as well as *ZNF3* gene and a cluster of genes on chromosome 11 (*SPI1-MTCH2*), shown as gene-based genome-wide significant association with AD which are in proximity with those detected by genome-wide significant single nucleotide polymorphisms (SNPs; Escott-Price et al., 2014). Another moderately rare variant in *TREM2* has also shown evidence for association (Guerreiro et al., 2013). However, rare risk-increasing variants may not be tagged by single SNPs, as for example, the case for *CLU* in which significant enrichment of rare variants in cases was observed independent of the single locus genome-wide association signal (Bettens et al., 2012). Disease risk may reflect the coaction of several loci but the number of loci involved at the individual or the population levels are unknown, as is the spectrum of allele frequencies and effect sizes (Risch, 1990).

Recently, it has been reported that there is a large polygenic contribution to the overall heritable risk of AD (Escott-Price et al., 2015), implying that the genetic architecture of AD includes many common variants of small effect that is likely to reflect a large number of susceptibility genes and a complex set of biological pathways related to disease. The approach using the polygenic risk score (PRS), introduced by the International Schizophrenia Consortium (International Schizophrenia et al., 2009), has facilitated the creation of genomic profiles which combine the effects of many associated genetic variants to predict risk of disease.

* Corresponding author at: Institute of Psychological Medicine and Clinical Neurosciences, MRC Centre for Neuropsychiatric Genetics & Genomics, Cardiff University, Cardiff CF24 4HQ, UK. Tel.: +442920688429; fax: +442920687068.

** Corresponding author at: Institute of Neurology, University College London, UK. Tel.: +442034484722; fax: +442078331016.

E-mail addresses: EscottPriceV@cardiff.ac.uk (V. Escott-Price), j.hardy@ucl.ac.uk (J. Hardy).

2. Material and methods

ROC curve analysis and area under the curve (AUC) provides the most widely accepted test of prediction accuracy in correctly classifying diseased and unaffected individuals. We were interested in investigating the maximal possible predictive power (AUC_{\max}) of genomic profiles based on disease prevalence and heritability of AD. AUC_{\max} is defined as the maximum AUC that could be achieved for a disease when the test classifier is a perfect predictor of genetic risk. To evaluate the maximal prediction accuracy of the genetic variability, we used the approach introduced by (Wray et al., 2010). This approach requires estimates of the heritability of liability (i.e., genetic variances on the liability scale) using genome-wide SNPs and disease prevalence. It has been reported that for AD the proportion of total variation tagged by all SNPs is 0.24 (SE 0.03; Lee et al., 2013). This variance was estimated using a restricted maximum likelihood (Yang et al., 2011) algorithm fitting all SNPs simultaneously in a data set of 3290 cases and 3849 controls (Genetic and Environmental Risk for Alzheimer's disease [GERAD1] consortium [Harold et al., 2009]), using a lifetime AD prevalence of 2% to transform the heritability on the observed scale to that on the liability scale (Brookmeyer et al., 1998).

The disease prevalence (K) is the other parameter, which has to be specified for the estimation of AUC_{\max} . The prevalence of AD in the population depends very much on age. Recent estimates show a 3%, 17%, and 33% prevalence in the 65–74, 75–84, and 85+ age groups, respectively (Hebert et al., 2013). For the AUC_{\max} calculations, we used the online calculator (Wray et al., 2010) and performed those for AD lifetime prevalence 2% and as well as prevalences for 3 age groups above.

Then we compared the theoretical AUC_{\max} with the empirical AUC estimates provided in (Escott-Price et al., 2015). In that report, the empirical AUCs were calculated in a subsample of GERAD1 participants, for whom *APOE* genotype data were available (3049 cases and 1554 controls, mean age was 73 years [SD = 9.6]). Since the genetic architecture of AD may be different for early, medium, and late onset, therefore the heritability and, consequently, the prediction accuracy may also vary with age. Thus to make the comparison of theoretical and empirical estimates fair, we first matched the prevalence parameter used for heritability of liability estimation and AUC_{\max} , and second, we compared the obtained maximum prediction accuracies with the empirical ones in the corresponding age groups. The heritability of liability h_l^2 for the disease prevalences (3%, 17%, and 33%) was re-calculated according to equation 23 in (Lee et al., 2011), $h_l^2 = h_0^2 \frac{K(1-K)}{z^2} \frac{K(1-K)}{P(1-P)}$, where P is the proportion of cases in the whole GERAD1 data set ($P = 30\%$), K is disease prevalence, h_0^2 is the estimate of heritability on the observed scale, and z^2 is the height of the standard normal probability density function at the truncation threshold K .

3. Results

The results are summarized in Table 1. Our analysis revealed that the maximum prediction accuracy is $AUC_{\max} = 82\%$ (95%

CI: 78%–85%) for a lifetime AD prevalence 2%. The best prediction accuracy of the PRS reported for that data set (Escott-Price et al., 2015) is $AUC = 74.5\%$ (95% CI = 0.73–0.76, see Fig. 1), calculated based on prediction results of logistic regression when *APOE* e4, e2, and PRS for SNPs with AD association p -values < 0.5 were included in the model (SNPs with $r^2 > 0.2$ to the best associated SNP in the region of 1 MB were pruned out). Note that the PRS there included proxies to the 20 GWAS SNPs reported in (Lambert et al., 2013) and were based on the full IGAP data set, which includes the GERAD data set. To exclude the potential bias introduced by this fact, we recalculated the prediction accuracy based on SNPs selected only from the independent data set (IGAP-noGERAD), and the AUC was quite similar 73.7 (95% CI: 0.72–0.75). The 74% AUC accounts for about 90% (74% out of 82%) of the prediction accuracy, which can be reached using genetic data as a predictor.

In Table 1, we also present genetic variance estimates for *APOE* locus alone and for *APOE* and 20 genome-wide association study (GWAS) significant loci. Our estimates were similar to (Lee et al., 2013; Ridge et al., 2013), which were obtained using the same and an independent Alzheimer's Disease Genetics Consortium data sets, respectively. In addition, our results show the maximum AUC for *APOE* and *APOE* + GWAS loci, $AUC_{\max} = 65\%$ (95% CI: 0.62–0.67) and 66% (95% CI: 0.64–0.67), respectively. A possible explanation is that in the (Escott-Price et al., 2015) article, the authors used directly genotyped e4 and e2 alleles, whereas the genetic variance estimates here were obtained with their proxies.

In the age group 65–74 (disease prevalence $K = 3\%$), the AUC_{\max} came to 82% (see Table 1). This is very close to the actual $AUC = 79.2\%$ calculated in the younger group of subjects (Supplemental Table 5 in [Escott-Price et al., 2015]). Correcting for the bias in empirical AUC estimation due to not fully independent top significant SNPs, we get the AUC in this age group is 75% (95% CI: 73%–77%) which accounts for 91.5% of the AUC_{\max} in this age group.

4. Conclusions

In summary, we have shown first that the pattern of the polygenic risk score prediction accuracy was similar to theoretical estimates of the accuracy based on heritability and AD prevalence. Second, the actual AUC values achieved for polygenic risk score when only ~87,600 SNPs (Escott-Price et al., 2015) were included are quite close to the upper limit (AUC_{\max}) that could be achieved from that data set given the genetic epidemiology of the disease, namely disease prevalence and heritability. Thus, suggesting that while as yet unknown, the majority of the remaining common variant susceptibility loci are captured, either directly or indirectly within our polygenic risk score model for this data set.

A possible limitation of our study is that the predictive modeling (Escott-Price et al., 2015) and the estimates of the heritability of liability (Lee et al., 2013) were performed mostly on the same subjects (GERAD1 sample). However, we believe that this fact does not affect our results since the SNPs selection and the individual

Table 1
The maximum area under the ROC curve (AUC) possible, based upon the AD prevalence and heritability estimates (Lee et al., 2013)

	2% lifetime prevalence			3% prevalence (age 65–74)	17% prevalence (age 75–84)	33% prevalence (age 85+)
	All SNPs	<i>APOE</i> region ^a	<i>APOE</i> & GWAS loci ^b			
h_l^2	0.24 [0.18–0.30]	0.048 [0.032–0.063]	0.051 [0.039–0.063]	0.27 [0.20–0.33]	0.45 [0.34–0.57]	0.55 [0.41–0.68]
AUC_{\max}	0.82 [0.78–0.85]	0.65 [0.62–0.67]	0.66 [0.64–0.67]	0.82 [0.78–0.86]	0.84 [0.80–0.88]	0.86 [0.81–0.90]

h_l^2 [95%CI], estimated genetic variance proportional to the total variance on the liability scale with 95% confidence intervals; AUC_{\max} [95%CI], the maximum value of AUC based on the genomic profile on heritability and disease prevalence with 95% confidence intervals.

Key: AUC, area under the curve; SNPs, single nucleotide polymorphisms.

^a The *APOE* region is defined between 44,400KB–46,500KB on chromosome 19 for the *APOE* locus.

^b GWAS loci are defined as ± 500 KB of both sides of the GWAS significant SNPs (Lambert et al., 2013). *APOE* region and GWAS loci were pruned with $r^2 = 0.2$.

Download English Version:

<https://daneshyari.com/en/article/4932884>

Download Persian Version:

<https://daneshyari.com/article/4932884>

[Daneshyari.com](https://daneshyari.com)