Available online at www.sciencedirect.com
ScienceDirect
Mechanisms of Ageing and Development 128 (2007) 76–82
Mechanisms of Ageing and Development
www.elsevier.com/locate/mechagedev

Susceptibility groups for Alzheimer's disease (OPTIMA cohort): Integration of gene variants and biochemical factors

Elizabeth H. Corder [a,*], Helen Beaumont [b]

[a] *Center for Demographic Studies, Duke University, Durham, NC 27708, USA*
[b] *OPTIMA, Oxford University, Oxford, United Kingdom*

Available online 20 November 2006

Abstract

Information on gene variants and blood levels (*APOE*, *BCHE-K*, *TF-C2*, *HFE-D*, *HFE-Y*, *ACE I/D*, *AR1*; homocysteine, folate and vitamin B_{12}) is available for participants in the Oxford Project to Investigate Memory and Ageing (OPTIMA) cohort ($n = 575$). This information identified four risk sets for Alzheimer's disease (AD) using grade of membership analysis (GoM). Graded membership scores that relate individuals to each set are automatically generated. Sets I and III had low intrinsic risk. Set II had high intrinsic risk associated with multiple gene variants, e.g., *APOE 44*/*34*. Set IV also had high intrinsic risk demonstrating low folate and B_{12} levels. Membership in the high intrinsic risk sets was summed, coded as either close versus not close ($\geq$0.80 versus $<$0.80) and input into logistic models to predict relative risk: close resemblance multiplied risk 80-fold for possible AD before age 65 and 55-fold for probable or definite AD at ages 65–74. These findings implicate both biochemical and genetic factors in the risk for AD and further support dietary supplementation with folate and vitamin B_{12} as a potential means to delay the onset of AD and/or its rate of progression.

Keywords: Grade-of-membership analysis; Alzheimer's disease; Gene variants; Folic acid; Vitamin B_{12}

1. Introduction

Age-related disorders such as Alzheimer's disease (AD) have multiple risk factors. The usual approach is to consider individual factors without consideration of the larger context of other putative or established risk factors. This approach makes replication of any one factor difficult (Lehmann et al., 2001; Prince et al., 2001) as replicate samples differ in frequencies of important unmeasured risk factors, and the frequency at which they are found together with the factor of interest. This approach also limits estimation of risk for individuals as the odds of disease are contrasted between the putative risk factor versus average risk. Considering AD, the association between apolipoprotein E ε4 allele (*APOE4*) (Strittmatter et al., 1993; Saunders et al., 1993; Corder et al., 1993; Corder et al., 1994) has been replicated and found to multiply risk about 3-fold.

We applied an integrative approach to identify high and low intrinsic risk sets for AD employing grade of membership analysis or GoM (Woodbury and Clive, 1974; Woodbury et al., 1978; Clive et al., 1983; Corder and Woodbury, 1993; Woodbury et al., 1994). This approach uses maximum likelihood to identify extreme pure type groups, here, intrinsic risk sets. Graded membership scores are automatically generated that relate individuals to the risk sets. Risk for individuals can be estimated in logistic models where membership in high intrinsic risk sets predicts disease status. No genetic model is specified. Non-genetic information on, e.g., age or symptoms, can be included in the model. This allows a contrast between persons who carry many risk factors and those who carry few or none, and a gradient of intermediate risk. Thus relative risks tend to be higher than found for, e.g., *APOE4* as a single factor (Corder et al., 2006; Licastro et al., 2006; Corder and Hefler, 2006; Corder and Mellick, 2006).

The identified risk sets are, evidently, biologically based often associated with certain age at the time when symptoms manifest. Thus, different samples from the same population, and having comparable age and sex distributions, would be expected to identify very similar risk sets. The role of chance

* Corresponding author at: 4500 19th St #314, Boulder, CO 80304, USA. Tel.: +1 303 449 2491; fax: +1 303 449 2491.
E-mail address: elizabethcorder@hotmail.com (E.H. Corder).

0047-6374/$ – see front matter
doi:10.1016/j.mad.2006.11.014

would be largely confined to the membership distributions of individuals in the risk sets. Presumably, this aspect should improve identification and replication of relevant risk factors, and common important combinations of risk factors. The approach is inclusive in that the dimensions of the problem are better defined by the inclusion of more items of information on individuals. Risk sets can be identified from demographic, symptom, biochemical, pathologic and/or gene data.

In this study, GoM integrates genetic and biochemical data for subjects enrolled in the Oxford Project to Investigate Memory and Ageing (OPTIMA) (Clarke et al., 1998; Budge et al., 2002; Lehmann et al., 1997; Lehmann et al., 2003; Lehmann et al., 2004; Lehmann et al., 2005; Hogervorst et al., 2001; Hogervorst et al., 2002; Robson et al., 2004). Gene variants and biochemical factors were selected that had previously been found to be associated with risk of AD in the sample. *APOE4* is easily the best established risk allele for AD (Saunders et al., 1993; Corder et al., 1993; Farrer et al., 1997). The K variant of butyrylcholinesterase (*BCHE-K*) has a point mutation at nucleotide 1615 (GCA → ACA), which changes Ala539 to threonine and is associated with less plasma activity, due to fewer circulating enzyme molecules (Bartels et al., 1992). A stratified meta-analysis (Lehmann et al., 2001) showed that *BCHE-K* is associated with AD only in men after age 75, particularly in those also positive for *APOE4*. One of the most studied candidate gene variants is the insertion (I)/deletion (D) polymorphism for angiotensin 1-converting enzyme (*ACE I/D*). A recent large meta-analysis (Lehmann et al., 2005) established this polymorphism as a marker of AD, probably in linkage disequilibrium with the true risk factor. D homozygotes were at reduced risk of AD in three ethnic groups studied: North Europeans, South Caucasians (Mediterranean and Middle East) and East Asians. In North Europeans, there was an increased risk for heterozygotes compared to either homozygosity. In a recent study of two variant alleles of iron metabolism (Robson et al., 2004), transferrin C2 (*TF-C2*) and haemochromatosis C282Y (*HFE-Y*), neither allele alone was associated with AD, but the combination raised the risk 5 times. It was suggested that the combination results in an excess of free iron and the generation of free radicals in neurones. The *HFE* H63D variant may also contribute to the problem.

Turning to the biochemical factors, it is now well established that higher levels of serum homocysteine are associated with increased risk of AD and other dementias (Clarke et al., 1998; Seshadri et al., 2002; Morris, 2003). Raised homocysteine is also associated with increased risk of cognitive impairment in the elderly (Lehmann et al., 2000; Budge et al., 2002; Kado et al., 2005; Nurk et al., 2005; Elias et al., 2006). Conversely, lower serum levels of the B vitamins, folate and vitamin B_{12}, have been associated with higher serum homocysteine (Jacques et al., 2001) and with increased risk both of AD (Clarke et al., 1998) and of cognitive impairment in the elderly (Kado et al., 2005; Nurk et al., 2005; Elias et al., 2006).

Four risk sets were identified from the OPTIMA cohort data labelled as I, II, III, and IV. Sets I and III had low intrinsic risk while sets II and IV demonstrated high intrinsic risk. Set II carried multiple high-risk gene variants. Set IV demonstrated elevated plasma levels of homocysteine, and low levels of folate and vitamin B_{12}. Close resemblance to the high intrinsic risk sets multiplied the risk of AD up to 80-fold.

2. Methods

2.1. Study subjects

OPTIMA is a prospective, longitudinal, clinicopathological study of ageing, in both 'control' elderly and memory-impaired subjects. Controls are volunteers without cognitive dysfunction and with CAMCOG > 80 (Roth et al., 1988). Control subjects in this study had no neuropsychiatric diagnosis at the time of death or last examination ($n = 260$). Case subjects are referred to the project with varying degrees and types of mental deterioration (Jobst et al., 1992). Affected subjects were diagnosed with possible AD ($n = 90$) or probable AD ($n = 80$) according to NINCDS-ADRDA criteria (McKhann et al., 1984) or with definite AD ($n = 145$) according to CERAD criteria (Mirra et al., 1991). The subjects are considered according to diagnosis (probable or definite AD, possible AD, unaffected) and age group (<age 65, 65 to 75, 76 or older).

2.2. Laboratory determinations

Total plasma homocysteine (tHcy), vitamin B_{12}, serum folate, and red blood cell (RBC) folate were measured as in Budge et al. (2002). These continuous values are represented by quintiles found for the sample from 1 = lowest quintile to 5 = highest quintile.

2.3. Genetic determinations

APOE genotype, i.e. inherited combinations of ε2, ε3, and ε4 alleles, was determined according to the method of Wenham et al. (1991). *BCHE-K* genotyping was by the method of Jensen et al. (1996), *TF-C2* genotyping by the method of Namekata et al. (1997), as modified by Robson et al. (2004). *HFE* genotyping was by the method of Merryweather-Clarke et al. (1997), with a modified primer for C282Y (Merryweather-Clarke et al., 1999). *ACE-I/D* genotyping was as described (Rigat et al., 1992), with 5% DMSO included in the PCR, as recommended (Fogarty et al., 1994), and all 135 DNA samples typed as DD, plus two weak examples of DI, confirmed by an insertion-specific second amplification, as proposed (Shanmugam et al., 1993). There are two triplet repeats in exon 1 of the androgen receptor (*AR1*) gene; the CAG repeat has been implicated in both low testosterone level and AD for men (only), interacting with *APOE* ε4—with longer repeats conferring decreased androgen sensitivity (Yaffe et al., 2003).

2.4. Data analytic approach

The statistical machinery to accomplish integration of the data is termed grade-of-membership analysis. GoM has been in existence since the early 1970's (Woodbury and Clive, 1974; Woodbury et al., 1978; Clive et al., 1983; Woodbury et al., 1994).

The GoM model likelihood can be described after first identifying four indices. One is the number of subjects I ($i = 1, 2, \ldots, I$). Here, $I = 575$ subjects were identified. The second index is the number of variables J ($j = 1, 2, \ldots, J$). There are $J = 11$ variables. Our third index is Lj: the set of response levels for the Jth variable. This leads to the definition of the basic GoM model where the probability that the ith subject has the Ljth level of the Jth variable is defined by a binary variable (i.e., $y_{ijl} = 0, 1$). The model with these definitions is

$$\text{Prob}(y_{ijl} = 1.0) = \sum_k g_{ik}\lambda_{kjl}, \quad (1)$$

where the g_{ik} are convexly constrained scores (i.e., $0.0 \le g_{ik} \le 1.0$; $\sum_k g_{ik} = 1.0$) for subjects and the λ_{kjl} are probabilities that, for the Kth latent group, the Ljth level is found for the Jth variable. The procedure thus uses this expression to identify K profiles representing the pattern of $J \times Lj$ responses found for I subjects.