Alzheimer's & Dementia 13 (2017) 933-939

Featured Article

Impact of home visit capacity on genetic association studies of late-onset Alzheimer's disease

David W. Fardo[a], Laura E. Gibbons[b], Shubhabrata Mukherjee[b], M. Maria Glymour[c], Wayne McCormick[b], Susan M. McCurry[d], James D. Bowen[e], Eric B. Larson[f], Paul K. Crane[b,*]

[a]*Department of Biostatistics and Sanders-Brown Center on Aging, University of Kentucky, Lexington, KY, USA*
[b]*Department of Medicine, University of Washington, Seattle, WA, USA*
[c]*Department of Epidemiology, University of California, San Francisco, San Francisco, CA, USA*
[d]*Department of Psychosocial and Community Health, University of Washington, Seattle, WA, USA*
[e]*Department of Neurology, Swedish Medical Center, Seattle, WA, USA*
[f]*Group Health Research Institute, Group Health, Seattle, WA, USA*

Abstract

Introduction: Findings for genetic correlates of late-onset Alzheimer's disease (LOAD) in studies that rely solely on clinic visits may differ from those with capacity to follow participants unable to attend clinic visits.

Methods: We evaluated previously identified LOAD-risk single nucleotide variants in the prospective Adult Changes in Thought study, comparing hazard ratios (HRs) estimated using the full data set of both in-home and clinic visits (n = 1697) to HRs estimated using only data that were obtained from clinic visits (n = 1308). Models were adjusted for age, sex, principal components to account for ancestry, and additional health indicators.

Results: LOAD associations nominally differed for 4 of 21 variants; *CR1* and *APOE* variants were significant after Bonferroni correction.

Discussion: Estimates of genetic associations may differ for studies limited to clinic-only designs. Home visit capacity should be explored as a possible source of heterogeneity and potential bias in genetic studies.

Keywords: Home research study visits; Research clinic study visits; Missing data; Bias; Prospective studies; Cohort studies; Longitudinal studies; Inference; Late-onset Alzheimer's disease; Genetics; Genome-wide studies; Genome-wide association studies; Ascertainment bias; Selection bias; Population-based studies

1. Introduction

Study design is underemphasized in planning or interpretation of many genome-wide association studies (GWASs) and sequencing projects [1]. In many settings, issues such as sampling, recruitment, and data collection strategies are assumed to be of secondary importance, although the potential for bias is well established. Many GWAS analysis projects amass participants from cohorts with varying recruitment strategies and phenotyping protocols. In resulting articles, these details are often relegated to supplementary information or not described at all.

Even subtle differences in subject ascertainment between studies could produce result heterogeneity, and such heterogeneity may be due to true differences in the relevance of genetic variants across subgroups or due to bias induced by selection processes. Many GWAS statistical models include few covariates, so subgroup effect heterogeneity is not explored, and there is little hope of correcting selection bias. If gene-environment interactions exist, or if the genetic effect occurs only in a subgroup of people, success of the GWAS framework may be especially dependent on the

*Corresponding author. Tel.: +1 206 744 1831; Fax: +1 206 897 4688.
E-mail address: pcrane@uw.edu

http://dx.doi.org/10.1016/j.jalz.2017.01.012

sample selection process as it affects the distributions of potentially important subject characteristics. More troubling, if genetic variants and symptoms of incipient AD interact in determining chances of participating in genetic studies, the observed association between the genetic variant and AD may be severely biased when estimated in the participants. Indeed, such a process can create an observed association that does not match the true effect even among study participants. Of course, most studies cannot evaluate this possibility because they have no information on study's nonparticipants. Not accounting for these scenarios can result in bias and/or diluted statistical power [2–6].

We recently published findings from a prospective cohort with a two-stage sample design indicating that genetic associations with late-onset Alzheimer's disease (LOAD) for the *APOE* locus may differ for individuals who attend in-person clinic visits than for the larger community-dwelling population, which includes some people who do not attend study visits in a research clinic [7]. We determined that risk factor associations with LOAD differed for participants who had only in-clinic visits versus the full sample that also included people seen at home. For *APOE* ε4, the estimated hazard ratio (HR) for incident LOAD in the full sample was 1.66 (95% confidence interval [CI] 1.37, 2.01), whereas in the clinic-only data set, *APOE* ε4 had an HR of 2.28 (95% CI 1.57, 3.30). The *P*-value for this difference was .008. *APOE* was the only genetic factor we evaluated.

The notion that study design could be important in the relevance and magnitude of associations with *APOE* has been known previously [8]. Indeed, although *APOE* genotype may be strongly predictive of LOAD status in specialty clinic settings, this association is attenuated in community-based settings [9]. We hypothesized that this pattern could be explained by selection bias due to specialty clinic studies lacking home study visit capacity. This phenomenon may also apply to other genetic variants. It is important to distinguish between bias (which occurs when estimated effects on average are not centered about the true value) and diminished power (where there is less chance of discovering a true effect).

We used genetic and research study data from the Adult Changes in Thought (ACT) study to determine whether home study visit capacity would have an influence on the strength of association with LOAD for single nucleotide variants (SNVs) from the largest LOAD GWAS meta-analysis to date [10].

2. Methods

2.1. Parent study description, ethical considerations, and funding

Methods for ACT have been published [11–13]. The original cohort enrolled between 1994 and 1996 included 2581 randomly selected dementia-free people aged ≥65 years who were members of Group Health, a Washington State health care system. An additional 811 participants were enrolled between 2000 and 2003, and in 2005, we began continuous enrollment. Participants are evaluated at 2-year intervals at a research clinic or in their home at the participant's choice. Other than location (i.e., home vs. clinic), screening research study visits are identical.

Study procedures were approved by the Institutional Review Boards of Group Health and the University of Washington. Participants provided written informed consent.

ACT is supported by the National Institute on Aging, which had no role in study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the article for publication.

2.2. LOAD identification

Participants were assessed at home or in clinic every 2 years with the Cognitive Abilities Screening Instrument, for which scores range from 0 to 100 and higher scores indicate better cognitive functioning [14]. Participants with scores of 85 or less underwent further evaluations, including a clinical examination and a battery of neuropsychological tests; dementia evaluations are in the participant's home regardless of the location of the triggering/screening visit. Results of these evaluations, laboratory testing, and imaging records were reviewed in a consensus conference, where research criteria were used to identify cases of dementia [15] and probable or possible AD [16]. Dementia-free participants continued with scheduled follow-up visits. In this study, we are examining whether associations between genetic variants and LOAD differ for people who participated in biennial screening visits in the clinic compared to all study participants; by design in ACT, the dementia evaluations all occur in the participant's home, so the location of these evaluations is not under study here.

2.3. Genotyping

ACT participants were genotyped in two waves. The bulk of the cohort was genotyped using the Illumina Human 660 Quad chip, and a subsequent genotyping wave used the Illumina OmniExpress chip. Data from European Americans from both samples were imputed to the same CEU Human HapMap reference panels as used in the International Genomics of Alzheimer's Project study from Lambert et al [10]. *APOE* genotype was obtained separately using standard procedures. Of the other 21 SNVs identified as the top hits in Lambert et al. [10], 20 were available in the ACT data either being directly genotyped or successfully imputed; the lone exception was the *DSG2* SNV.

2.4. Statistical analyses

We limited our analyses to European Americans based on principal components analysis (a conventional statistical genetics approach to discern ancestry from GWAS data [17]) and self-reported white race. We constructed two data sets from the subset of the ACT study with at least one follow-up