Applying novel methods for assessing individual- and neighborhood-level social and psychosocial environment interactions with genetic factors in the prediction of depressive symptoms in the Multi-Ethnic Study of Atherosclerosis. *Behavior Genetics*. Erin B. Ware, Jennifer A. Smith, Bhramar Mukherjee, Seunggeun Lee, Sharon L. R. Kardia, Ana V. Diez-Roux

Corresponding author: Erin Ware[1,2]
Email: ebakshis@umich.edu
[1]Institute for Social Research, University of Michigan, Ann Arbor, MI
[2]Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI

Online Resource 1

Extended Methods

Depressive symptom score

The Center for Epidemiologic Studies Depression scale (CES-D) was developed by the Center for Epidemiologic Studies for use in general population surveys (Comstock & Helsing 1976; Radloff 1977). The CES-D has an excellent internal consistency (Cronbach's alpha = 0.90) (Radloff 1977), and is designed to assess depressive symptoms at a specific period in time (over the past week). Large epidemiologic studies frequently assess depressive symptoms using the CES-D rather than depression as a diagnosis. Depressive symptoms are associated with and predict major depression, with sensitivity and specificity reported anywhere in the range of 64%-90% and 70%-93% respectively when using a cutoff of 16 (Boyd *et al.* 1982; Roberts & Vernon 1983; Breslau 1985; Beekman *et al.* 1997), but the investigation of depressive symptoms as a continuous trait is of interest per se. Few genetic association studies currently exist using a continuous measure of depressive symptoms. The outcome measure for this analysis is a sum of the 20 items, ranging from 0 to 60. If more than 5 items were missing, the CES-D score was not calculated. If 1-5 items were missing, the scores were summed for completed items, dividing the sum by the number of questions answered and then multiplying by 20. There were 5,178 participants with three measures of CES-D, 507 with two measures, and 650 with only baseline CES-D measures, for a total of 6,335 participants with 17,198 observations. The CES-D scores were log-transformed to improve normality.

Anti-depressant use was defined as taking any or multiple of the following medications: Selective Serotonin Reuptake Inhibitors (SSRIs), Monoamine Oxidase Inhibitor (MAOI), Tricyclic anti-depressant, and/or Non-tricyclic anti-depressant other than MAOI. Anti-depressant use was assessed at each exam and corrected CES-D scores were estimated for each exam. A total of 7.6%, 7.9% and 8.1% of persons were on anti-depressant medications at exams 1, 3 and 4, respectively. We corrected for anti-depressant use with methods previously described (Hek *et al.* 2013). Briefly, assuming that the depression score is lower in treated than in untreated participants, and that participants with high depression scores, on average, respond less to their medication than persons with lower depression scores, we used a nonparametric imputation algorithm to adjust for the treatment effect. Separately for men and women and within each ethnicity separately, we replaced CES-D score for a person using anti-depressants with the mean depressive symptom score of all persons using anti-depressants with greater or equal CES-D scores. This method is based on an algorithm previously used to adjust blood pressure for persons on antihypertensive medication (Levy *et al.* 2000). We chose not to exclude participants taking anti-depressant medication as they often are individuals with depression or higher depressive symptom scores and thus add value to genetic studies. To improve normality of the outcomes, averaged measures CES-D (adding one point to all values) were log-transformed after adjustment for anti-depressant use.

Covariates

Age, sex, and study site were assessed at the MESA baseline exam. There were a total of 6,335 MESA participants included in the averaged depressive symptom GWAS and subsequent region-based analyses (AA 25%, EA 40%, CA 12%, HA 23%) (Ware *et al.* 2015). Average age (standard deviation) for the AA, EA, CA, and HA sub-samples was 62.2 (10.1), 62.6 (10.2), 61.4 (10.3), and 62.4 (10.4) years, respectively. Slightly less than half of each ethnicity was male (AA 48.0%, EA 46.4%, HA 49.3%, CA 49.7%) (Ware *et al.* 2015). Ethnicity-specific principal components were used to adjust for population stratification.

Adult socioeconomic position (ASEP) was included as an additional covariate to assess any residual confounding over the adjustment for ancestry through the inclusion of principal components. Since several measures of ASEP were available (measuring different dimensions of socioeconomic position), indicators were summarized into an ASEP score. The methods are based on previous work and combine information on income, education, and wealth (ownership of a home, car, land/property or investments).(Pollitt *et al.* 2005; Lemelin *et al.* 2009) Income was defined in four categories (<$25,000, $25,000–39,999, $40,000–74,999, or +$75,000). At the baseline examination, highest level of education completed was reported and for these analyses operationalized into four categories (completed high school or less, some college but no degree/technical school certificate, associate or bachelor's degree, or graduate/professional degree). The four wealth indexes included whether the participant: (1) had investments such as stocks, bonds, mutual funds, retirement investments, or other investments (yes/no), (2) owned their home (yes/no) (3) owned a car (yes/no) and (4) owned land or another property that was not their primary residence (yes/no). To create the summary score for ASEP, the individual measures for income, education and wealth were summed (income variable (0 – 3, low to high), education (0 – 3, low to high), and for each affirmative wealth indicator, a single point was added). The ASEP score ranged from 0 – 10, with higher scores indicating greater ASEP.

Environment

CB was measured at two exams in MESA (exams one and three) and is based off of the chronic burden scale developed for the Healthy Women Study.(Bromberger & Matthews 1996) It is an index of affirmative responses to five individual burdens including health (self and others), finances, employment, and relationships that were ongoing for more than six months. Within each exam if a component score was missing, the overall CB for that exam was set to missing. CB was averaged across the two exams for each individual. If either exam was missing, CB was created from the existing measure. If both exams were missing, CB was set to missing. CB was centered at the overall mean.

Emotional social support is available at exams one and three of MESA and is based on a scale from the Enhancing Recovery in Coronary Heart Disease study.(Enrichd Investigators 2001) It is an index rating six questions on a five-point Likert scale (1 = "none of the time" to 5 = "all of the time"). These questions included asking if someone was available to listen, give advice, show love and affection, help with daily chores, provide emotional support, and confide in. Within each exam if a component score was missing, the overall SS for that exam was set to missing. If either exam was missing, SS was created from the existing measure. If both exams were missing, SS was set to missing. SS was centered at the overall mean.

Neighborhood social environment is summarized into a neighborhood index score (NIS) composed of three dimensions: aesthetic quality (AQ), safety (SF), and social cohesion (SC) measured with a 1-mile radius as the definition of neighborhood. The respondent's own answer was not included in the crude mean estimates for the neighborhood, allowing for more objective neighborhood measures than using the MESA participant's perception of neighborhood dimensions alone. The neighborhood level data was linked to the participant's addresses within a 1-mile buffer by matching each participant of the survey within 1 mile based on the latitude/longitude of the address. Responses of "Don't Know" or "Refused" values were set to missing for each of the original variables in each of the surveys. Several questions were reverse coded so that questions reflected better social outcomes with increasing scores.

If any one of the nine variables (AQ exam one, three, and four; SF exam one, three, and four; or SC exam one, three, and four) was missing, then NIS was set to missing. The index score was then mean-centered by the combined-ethnicity mean to aid interpretability. The index scores range from -1.30 to 0.95, with a mean of 0 and a standard deviation of 0.33 in the combined sample (AA, EA, CA, HA).

Region analyses

SKAT

SKAT assumes the following genetic main effect model:

$$\bar{y}_{i\cdot} = \alpha_0 + \boldsymbol{\alpha}'\boldsymbol{X_i} + \boldsymbol{\beta}'\boldsymbol{G_i} + \epsilon_i,$$

where $\bar{y}_{i\cdot}$ is the log-transformed, averaged depressive symptom score corresponding to individual *i (i=1,…,n)*, α_0 is an intercept term, $\boldsymbol{X_i}$ is a vector of non-genetic covariates (age, sex, study site, PC1 – 4, and ASEP), $\boldsymbol{G_i} = (g_{i1}, \ldots, g_{ip})'$ is a vector of best-call genotypes (0 = no copies of the coded allele, 1 = one copy of the coded allele, 2 = two copies of the coded allele). The coded allele is the same for all ethnicities. Residual error ϵ_i follows a normal distribution with mean zero and variance σ^2. The vector of regression coefficients for the covariates is represented by $\boldsymbol{\alpha}$, and $\boldsymbol{\beta}$ is a vector of regression coefficients for the *p* observed genetic variants in the region. A primary assumption of SKAT is that each β_j, j = 1, …, p follows an arbitrary distribution with mean zero and variance $w_j^2\tau$. The weights, w_j, are specified based on MAF. Testing H_0: $\tau = 0$ is equivalent to testing H_0: $\boldsymbol{\beta} = 0$. The SKAT test statistic is

$$Q_{SKAT} = (\overline{\boldsymbol{y}} - \hat{\mu})'GWG'(\overline{\boldsymbol{y}} - \hat{\mu})$$

where $\widehat{\boldsymbol{\mu}} = (\hat{\mu}_1, \ldots, \hat{\mu}_n)'$ is the estimated mean of $\overline{\boldsymbol{y}}_{\boldsymbol{i}} = (\bar{y}_{1\cdot}, \ldots, \bar{y}_{n\cdot})'$ under the null model of no genetic effects. The Q_{SKAT} also can be written as

$$Q_{SKAT} = \sum_{j=1}^{p} w_j^2 S_j^2 ,$$

where $S_j = \sum_{j=1}^{n} g_{ij}(\bar{y}_{i\cdot} - \hat{\mu}_i)$ is a score statistic of single variant j. The Q_{SKAT} follows a mixture of chi-squared distributions under the null hypotheses that can be evaluated explicitly and used as a reference distribution to compute the p-values. Results with significant p-values indicate that there is at least one non-zero $\beta_j$in the region. Since this analysis is only concerned with the effects of common SNP variants, not the effects of rare variants, the analysis is implemented without a MAF based weighting scheme, which indicates that $w_j = 1$ for all j=1, …, p.

MetaSKAT

MetaSKAT allows for the meta-analysis of SNP set-level results across cohorts or, in this case, ethnicities.(Lee *et al.* 2013) To allow for heterogeneity across ethnicities, MetaSKAT assumes effect sizes of genetic variants in different ethnic groups are independent and follow a common distribution. Suppose S_{kj} is a score statistic of the j^{th} variant ($j=1, \ldots, p$) in the k^{th} ethnic group ($k=1, \ldots, K$). The meta-analysis SKAT test statistic with assuming the heterogeneous genetic effect is:

$$Q_{Het-meta-SKAT} = \sum_{j=1}^{p} \sum_{k=1}^{K} w_{kj}^2 S_{kj}^2,$$

where w_{kj}^2 is the ethnic-specific weight for variant j. Individual-level genotype data were used to construct tests statistics.

References

Beekman A.T., Deeg D.J., Van Limbeek J., Braam A.W., De Vries M.Z. & Van Tilburg W. (1997) Criterion validity of the Center for Epidemiologic Studies Depression scale (CES-D): results from a community-based sample of older subjects in The Netherlands. *Psychol Med* **27**, 231-5.

Boyd J.H., Weissman M.M., Thompson W.D. & Myers J.K. (1982) Screening for depression in a community sample. Understanding the discrepancies between depression symptom and diagnostic scales. *Arch Gen Psychiatry* **39**, 1195-200.

Breslau N. (1985) Depressive symptoms, major depression, and generalized anxiety: a comparison of self-reports on CES-D and results from diagnostic interviews. *Psychiatry research* **15**, 219-29.

Bromberger J.T. & Matthews K.A. (1996) A longitudinal study of the effects of pessimism, trait anxiety, and life stress on depressive symptoms in middle-aged women. *Psychol Aging* **11**, 207-13.

Comstock G.W. & Helsing K.J. (1976) Symptoms of depression in two communities. *Psychol Med* **6**, 551-63.

Enrichd Investigators (2001) Enhancing Recovery in Coronary Heart Disease (ENRICHD) study intervention: rationale and design. *Psychosom Med* **63**, 747-55.

Hek K., Demirkan A., Lahti J., Terracciano A., Teumer A., Cornelis M.C., Amin N., Bakshis E., Baumert J., Ding J., Liu Y., Marciante K., Meirelles O., Nalls M.A., Sun Y.V., Vogelzangs N., Yu L., Bandinelli S., Benjamin E.J., Bennett D.A., Boomsma D., Cannas A., Coker L.H., de Geus E., De Jager P.L., Diez-Roux A.V., Purcell S., Hu F.B., Rimm E.B., Hunter D.J., Jensen M.K., Curhan G., Rice K., Penman A.D., Rotter J.I., Sotoodehnia N., Emeny R., Eriksson J.G., Evans D.A., Ferrucci L., Fornage M., Gudnason V., Hofman A., Illig T., Kardia S., Kelly-Hayes M., Koenen K., Kraft P., Kuningas M., Massaro J.M., Melzer D., Mulas A., Mulder C.L., Murray A., Oostra B.A., Palotie A., Penninx B., Petersmann A., Pilling L.C., Psaty B., Rawal R., Reiman E.M., Schulz A., Shulman J.M., Singleton A.B., Smith A.V., Sutin A.R., Uitterlinden A.G., Volzke H., Widen E., Yaffe K., Zonderman A.B., Cucca F., Harris T., Ladwig K.H., Llewellyn D.J., Raikkonen K., Tanaka T., van Duijn C.M., Grabe H.J., Launer L.J., Lunetta K.L., Mosley T.H., Jr., Newman A.B., Tiemeier H. & Murabito J. (2013) A Genome-Wide Association Study of Depressive Symptoms. *Biol Psychiatry*.

Lee S., Teslovich T.M., Boehnke M. & Lin X. (2013) General framework for meta-analysis of rare variants in sequencing association studies. *Am J Hum Genet* **93**, 42-53.

Lemelin E.T., Diez Roux A.V., Franklin T.G., Carnethon M., Lutsey P.L., Ni H., O'Meara E. & Shrager S. (2009) Life-course socioeconomic positions and subclinical atherosclerosis in the multi-ethnic study of atherosclerosis. *Soc Sci Med* **68**, 444-51.

Levy D., DeStefano A.L., Larson M.G., O'Donnell C.J., Lifton R.P., Gavras H., Cupples L.A. & Myers R.H. (2000) Evidence for a gene influencing blood pressure on chromosome 17. Genome scan linkage results for longitudinal blood pressure phenotypes in subjects from the framingham heart study. *Hypertension* **36**, 477-83.

Pollitt R.A., Rose K.M. & Kaufman J.S. (2005) Evaluating the evidence for models of life course socioeconomic factors and cardiovascular outcomes: a systematic review. *BMC public health* **5**, 7.

Radloff L. (1977) The CES-D scale: A self-report depression scale for research in the general population. *Applied psychological measurement* **1**, 385-401.

Roberts R.E. & Vernon S.W. (1983) The Center for Epidemiologic Studies Depression Scale: its use in a community sample. *Am J Psychiatry* **140**, 41-6.

Ware E.B., Mukherjee B., Sun Y.V., Diez Roux A.V., Kardia S.L. & Smith J.A. (2015) Comparative genome-wide association studies of a depressive symptom phenotype in a repeated measures setting by race/ethnicity in the Multi-Ethnic Study of Atherosclerosis. *BMC Genetics* **Under Review**.