Characterizing social environment's association with neurocognition using census and crime data linked to the Philadelphia Neurodevelopmental Cohort

T. M. Moore[1],[*], I. K. Martin[1], O. M. Gur[2], C. T. Jackson[1], J. C. Scott[1], M. E. Calkins[1], K. Ruparel[1], A. M. Port[1], I. Nivar[1], H. D. Krinsky[1], R. E. Gur[1], R. C. Gur[1]

[1]Department of Psychiatry, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

[2]Department of Criminal Justice, Pennsylvania State University, Abington College, Abington, PA, USA

Abstract

Background.—The contribution of 'environment' has been investigated across diverse and multiple domains related to health. However, in the context of large-scale genomic studies the focus has been on obtaining individual-level endophenotypes with environment left for future decomposition. Geo-social research has indicated that environment-level variables can be reduced, and these composites can then be used with other variables as intuitive, precise representations of environment in research.

Method.—Using a large community sample ($N = 9498$) from the Philadelphia area, participant addresses were linked to 2010 census and crime data. These were then factor analyzed (exploratory factor analysis; EFA) to arrive at social and criminal dimensions of participants' environments. These were used to calculate environment-level scores, which were merged with individual-level variables. We estimated an exploratory multilevel structural equation model (MSEM) exploring associations among environment- and individual-level variables in diverse communities.

Results.—The EFAs revealed that census data was best represented by two factors, one socioeconomic status and one household/language. Crime data was best represented by a single crime factor. The MSEM variables had good fit (e.g. comparative fit index = 0.98), and revealed that environment had the largest association with neurocognitive performance ($\beta = 0.41$, $p < 0.0005$), followed by parent education ($\beta = 0.23$, $p < 0.0005$).

Conclusions.—Environment-level variables can be combined to create factor scores or composites for use in larger statistical models. Our results are consistent with literature indicating that individual-level socio-demographic characteristics (e.g. race and gender) and aspects of

[*]Address for correspondence: T. M. Moore, Ph.D., M.Sc., University of Pennsylvania, Philadelphia, Pennsylvania, USA. (tymoore@upenn.edu).

familial social capital (e.g. parental education) have statistical relationships with neurocognitive performance.

Keywords

Introduction

The contribution of 'environment' has been investigated across diverse and multiple domains related to health and cognition (Yen & Syme, 1999; McEwen, 2012; Berkman *et al.* 2014; Krabbendam *et al.* 2014). However, in the context of large-scale genomic studies the focus has been on obtaining individual-level biomarkers or endophenotypes and environment is considered as a monolithic component that is left for future decomposition (e.g. Gur *et al.* 2007; Greenwood *et al.* 2013), even though the importance of environmental factors has been recognized and demonstrated (e.g. Mezuk *et al.* 2015; Smoller, 2015). Some limited amount of information on the home environment is typically collected from research participants, such as parental education and household income, and demographic characteristics of their communities such as median age, sex ratios, average education level and ethnicity proportions can be extrapolated based on their home address. It is generally assumed that the environment affects the outcome measures in multiple ways, but there is limited time to collect such information, and the emphasis on measuring complex biomarkers precludes deep environmental phenotyping.

The increased availability of large-scale public databases with detailed information on environmental factors now enables probing of environmental effects on research participants after completion of data collection, provided that information has been collected on their residence. Notably, prospective cohort designs are particularly useful for examining environmental influences on disease risk (Manolio *et al.* 2006). Here we will illustrate a methodology for accomplishing such integration in a large prospective cohort and show how this paradigm can help elucidate some environmental associations with biomarkers. The analytic objective was to not only capture the statistical associations of the census-derived social environment characteristics, but to begin to characterize the complex social dynamics that they represent, using the underlying structure of correlation between those characteristics. The Philadelphia Neurodevelopmental Cohort (PNC) participants present an opportunity to harness the robust social diversity of the Philadelphia area and apply appropriate quantitative methodology to examine the association between social environment and neurocognitive performance.

When confronting large complex datasets, as is common in the efforts to dissect environment, exploratory factor and principal components analysis (PCA) of spatially linked, 'census-like' data have been applied. As in the present case, the goal is often to reduce a larger number of available variables to a more manageable number of summary variables. These summary variables are then used in larger substantive analyses involving, for example, hypertension-related mortality rates (James & Kleinbaum, 1976), neighborhood

change (Temkin & Rohe, 1998), quality of life (Lo & Faber, 1997; Li & Weng, 2007), child maltreatment (Ernst, 2001), chronic pain (Fuentes *et al.* 2007[1]†), economic development (Roberts & McBee, 1968), multiple sclerosis (Lauer, 1994), body mass (Wang *et al.* 2007), and use of mental health services (Tello *et al.* 2005).

A second goal of factor analyzing social environment data, especially in the geographic and sociological sciences, is to profile the geo-social characteristics of a particular area. For example, Langlois & Kitchen (2001) examined the PCA structure of economic and social variables in Montreal, and Ray (1971) conducted a very similar analysis across Canada. Comparable analyses with similar purposes have been carried out in Canberra, Australia (Jones, 1965), Manhattan (Carey, 1966) and Great Britain (summary in Herbert, 1968). Sometimes, as in the present case, these local analyses are conducted with the specific purpose of calculating a summary score that can be used in subsequent analyses. For example, Barros & Victora (2005) used PCA to develop a geographic wealth score in Brazil, and Havard *et al.* (2008) used PCA to develop a neighborhood-level index of socioeconomic deprivation in France. However, this methodology is rarely used in genomic studies of neurobehavioral domains, and here we describe the application of such a factor analysis to a study of cognition and brain development in a large population-based cohort of genotyped youths. Such an approach helps interpret individual differences in cognitive performance, illustrating the power of integration with environmental data to elucidate potential causes for variability that can separate genetic from environmental processes.

Method

Participants

The participants and recruiting methods of the PNC have been described in detail (Gur *et al.* 2010; Calkins *et al.* 2014, 2015; Moore *et al.* 2015). The sample included youths (age 8–21) recruited through an NIMH-funded Grand Opportunity study characterizing clinical and neurobehavioral phenotypes in a genotyped prospectively accrued community cohort. All study participants were previously consented for genomic studies when they presented for pediatric services within the Children's Hospital of Philadelphia (CHOP) healthcare network. At that time they provided a blood sample for genetic studies, authorized access to electronic medical records and gave written informed consent/assent to be re-contacted for future studies. Of the 50 540 genotyped subjects, 18 344 met criteria and were randomly selected, with stratification for age, sex and ethnicity.

The sample included ambulatory youths in stable health, proficient in English, physically and cognitively capable of participating in an interview and performing computerized neurocognitive testing. Youths with disorders that impaired motility or cognition (e.g. significant paresis or palsy, intellectual disability) were excluded. Notably, participants were not recruited from psychiatric clinics and the sample is not enriched for individuals who seek psychiatric help. Also, because CHOP services a large area covering the entire greater

---

[1]Technically, these researchers use only three neighborhood-level variables to construct a unidimensional factor, used within a larger confirmatory model; however, the factor itself is still 'exploratory' in the sense that the relative weights (loadings) given to the three variables are determined empirically by the factor method.
†The notes appear after the main text.

Philadelphia region and surrounding counties (including parts of New Jersey and Delaware), the geographical distribution of participants was quite wide. A total of 9498 participants enrolled in the study, the majority between November 2009 and September 2011 and were included in this analysis. Participants provided informed consent/assent after receiving a complete description of the study and the Institutional Review Boards at Penn and CHOP approved the protocol.

Measures

Variables were measured or collected at one of three levels: the individual level (e.g. cognitive test scores, age, medical health ratings), the family level (e.g. number of siblings, mother age at birth, family turbulence score), and the census block group/neighborhood[2] level (neighborhood crime rates, percentage of neighborhood residents who are female, etc.)[3]. Our focus was on the reduction of the plethora of variables in the last category, the neighborhood-level variables. These were collected from the 2010 census-based American Community Survey (ACS) and the 2008 police database on crime rates in the Philadelphia area, which included both violent and non-violent crimes[4]. Examples of census-based variables included median family income, percent of residents who are married, percent of households that are non-family[5], percent of residents with children, percent of residents who speak English, etc. Examples of crime rate variables include aggravated assaults *per capita*[6], theft from automobiles *per capita*, etc. Note that because the census and police databases provide absolute counts, most of these variables had to be converted to percentages by dividing by the total block-group population.

Results from the computerized neurocognitive assessments are integrated in the structural analysis example described below as summary variables computed using methods described elsewhere (Gur *et al.* 2010; Calkins *et al.* 2014, 2015; Moore *et al.* 2015). Specifically,

1. Neurocognitive performance test scores (accuracy and speed) were factor scores generated from a battery of twelve tests designed to probe major neurobehavioral domains. Gur *et al.* (2010) describe the test battery, and details of the factor analyses are in Moore *et al.* (2015).

2. Psychopathology scores (such as 'externalizing' and 'psychosis') were factor scores generated from item-wise analyses of a comprehensive clinical assessment tool, the GOASSESS. A description of the instrument and its administration is

---

[2]From Gross & McDermott (2009, p. 162): 'The census block group is the smallest geographic area (encompassing approximately 1000 individuals) for systematic reporting and, consequently, researchers frequently equate the statistical phenomena exclusive to a given block-group unit with the phenomena that essentially define a neighborhood'.

[3]For children aged 8–10, all clinical and demographic variables (e.g. parent education) were collected from the collateral (usually parent); for ages 11–17, they were collected from the participant and the collateral; for ages 18–21, they were collected from the participant only.

[4]The Philadelphia Police Department (PPD) provides data from their Incident Transmittal System (INCT) to the University of Pennsylvania Cartographic Modeling Lab (CML), which makes it available to researchers. All incidents to which police respond are included, with some re-classified within 5 days pending investigation (cf. Gross & McDermott, 2009). These data are then linked to 2000 census block groups.

[5]The Census Bureau indicates 'A nonfamily household can be either a person living alone or a householder who shares the housing unit only with nonrelatives – for example, boarders or roommates. The non-relatives of the householder may be related to each other' (see Vespa *et al.* 2013).

[6]Using analogous crime data from 2006 (i.e. also from the CML via PPD), Gross & McDermott (2009) note that aggravated assaults include homicides, as the latter are not coded as such until an investigation has been completed.

provided by Calkins *et al*. (2014), and a description of the methods used for calculating the factor scores used here are available upon request (Calkins *et al*. unpublished data).

Finally, some individual-level variables were obtained either from the clinical interview cited above or from basic demographics collected during enrollment. These include age, race, gender, trauma exposure (a total count of traumatic experiences from a list of nine), substance use (a total count of non-pharmaceutical substances used in the last year), parent education (mean years of mother and father, unless only one is available), and whether the participant's parents were separated or divorced.

Exploratory factor analysis (EFA)

The 2010 census-based ACS variables and the neighborhood crime rate variables were analyzed separately via EFA in R (R Core Team, 2014). These were performed using many combinations of extraction method (least squares, maximum likelihood, principal axis) and oblique rotation (oblimin, promax, geomin) to check for inconsistency across method. Inconsistency was minimal, and thus results reported here are for the (default) least squares extraction method with oblimin rotation. The unidimensional, two-, and three-factor solutions of the census and crime variables were examined for interpretability, and the cleanest and most interpretable solution was selected for calculating factor scores by the Thurstone (1935) method using the factor.scores() command in the R psych package (Revelle, 2013). The scree plot for the census and crime variables was also examined, and was consistent with our judgment of the most interpretable solution. Extraction beyond three factors for either data set showed signs of over-extraction, such as factors comprising only one indicator. Note that race-related variables such as 'percent white' were not included in these analyses (or scores), because we wished to analyze specific associations of neighborhood racial composition independent of the summary variables, i.e. we wished to include a separate race-related variable ('percent white') in the structural model demonstration described below. EFAs that included neighborhood racial composition differed very little from the analyses presented here, and are available upon request.

Multilevel structural equation model demonstration

The neighborhood-level factor scores were used in combination with the other individual-level variables described above in a demonstrative structural model. Fig. 1 shows the conceptual path diagram describing the model. Due to intra-class correlation, this type of multilevel data usually requires a special kind of modeling called hierarchical linear modeling. In the structural equation modeling (SEM) framework, it is implemented as multilevel SEM (MSEM). The data used here technically involved three levels (individuals within households within neighborhoods); however, as sibling pairs (especially in the same household) enrolled in the study were relatively rare in the sample (1.3%), household-level variables were treated as individual-level variables in the structural analysis. Additionally, because the crime-related variables were measured 2 years earlier than the census-level variables and were therefore based on the 2000 census block groups, they could not be treated as neighborhood-level variables along with the 2010 census-level variables. That is, although the 2000 and 2010 block groups largely overlapped, there were some exceptions,

meaning an individual living in the same place in 2000 and 2010 might be assigned to two different block groups in 2000 and 2010. Crime-related variables were therefore treated as individual-level[7] crime exposure variables (from 2008), while the 2010 census variables were treated as neighborhood-level. The end result was a two-level model with census-based ACS variables at the neighborhood level and all other variables at the individual level.

The variables were related in a MSEM estimated using the robust maximum likelihood estimator in Mplus (Muthén & Muthén, 1998–2013). The model revolved around a single dependent variable of interest (neurocognitive test performance accuracy; see Fig. 1), to which all other variables related. To explore mediation, many of the independent variables were related to other independent variables either by direct effect or by correlation. Specific relationships among independent variables were determined by theory and by examining the model modification indices (Sörbom, 1989).

Results

Exploratory factor analysis

Table 1 shows the unidimensional, two-, and three-factor models for the 13 census-based (block-group-level) variables. The unidimensional model is dominated by socioeconomic status (SES)-related variables, including percent in poverty (−0.86), percent married (0.84), median family income (0.82), and percent with at least a high school education (0.75). Other variables seemingly unrelated to SES have negligible loadings, including average household size (0.02), percent of residents with children (0.09), and percent of households that are non-family (−0.19).

The two-factor model in Table 1 retains the SES factor (F1), while the second factor is determined by aspects of household sizes and knowledge of English (regardless of whether it is their first language). As in the unidimensional model, the SES-related factor (F1) is dominated by the percent of residents in poverty, the percent of residents who are married, and the median family income. The household-related factor (F2) is dominated by the percent of residents with children (0.90) and the percent of residents who are English speakers (−0.54). Overall, the two-factor model has a simple structure, with the exception of the cross-loading (−0.31) of percent employed on F2; specifically, those who live in areas with large households and few English speakers are slightly less likely than average to be employed.

The three-factor model in Table 1 is mostly identical to the two-factor model, except that median age has 'broken away' from F1 to form its own factor (F3). F3 is completely dominated by median age (loading = 0.92), with only two small negative loadings for population density (−0.27) and percent of households that are non-family (−0.30). That is, older people tend to live in neighborhoods that are less dense and with more family households.

---

[7]We believe this is justified by the moderate-to-small intra-class correlation (0.057) for the dependent variable of interest, combined with the very small average number of persons per block group (~2.4). Either of these alone would suggest multilevel modeling might be unnecessary even for the census data (Hox, 1998), but we decided to do so for the census data out of caution and for the sake of demonstration.

Due to the simple structure and interpretability of the two-factor model (and the lack thereof for the three-factor model), we decided to use the two-factor model for calculating scores. Inspection of the scree plot (Cattell, 1966; see Fig. 2) lends moderate support for this choice of two factors, because that is arguably where the 'elbow' of the scree function occurs (see Bentler & Yuan, 1998).

Table 2 shows the unidimensional, two-, and three-factor models of crimes (per 100 persons) committed in Philadelphia neighborhoods in 2008. The unidimensional model is dominated by domestic crimes (disturbance = 0.86, abuse = 0.81) and common assaults (non-aggravated = 0.86; aggravated without guns = 0.83). The weakest indicators are minor crimes (dangerous dog, false police report, gambling, and liquor law violation), all with loadings <0.15. The mean unidimensional loading is 0.52, and the scree plot (Fig. 3) shows a dramatic drop in explained variance when a second factor is extracted (1st:2nd eigenvalue ratio = 5.13). In the two-factor model, F2 largely comprises common, non-violent crimes (theft from auto, auto accidents, embezzlement), whereas F1 retains the violent crimes, as well as other miscellaneous crimes (drug possession, curfew violation, traffic violation). Additionally, there are some notable cross-loadings in the two-factor model; namely, vandalism, grand theft auto, auto-tag theft, lost property, check fraud, and robbery without guns all load on both factors at least 0.35. The three-factor model retains much of the same structure as the two-factor model. One important exception is that the six variables with cross-loadings on F2 in the two-factor model (robbery with guns, vandalism, residential burglary, grand theft auto, harassment, and auto-tag theft) all shift to F2 in the three-factor model. F3 appears to be a contrast factor positively indicated by animal incidents, aggravated assault with guns, and missing persons, and negatively indicated by pickpocketing, embezzlement, and retail theft. Due to the (1) questionable interpretability of the two- and three-factor models, (2) large number of cross-loadings in both models, (3) moderate correlation between F1 and F2 in the two-factor model, and (4) high ratio of 1st:2nd eigenvalues in the unidimensional model, we decided to use the unidimensional model for calculating scores. That is, each individual received a single score for the amount of crime (*per capita*) in his/her area.

Multilevel structural equation model demonstration

A highly complex structural model is difficult to display in graphical form; thus, Table 3 provides an example of the results of such a model predicting Computerized Neurocognitive Battery (CNB) accuracy. The first nine are direct associations of various individual-level variables with the dependent variable of interest (CNB accuracy). Further, the 27th, 28th and 29th effects listed in Table 3 are direct associations of neighborhood-level variables with CNB accuracy. All other reported associations are among the independent variables themselves. The fit of the model is acceptable (comparative fit index = 0.98; root mean square error of approximation = 0.036; standardized root mean square residual = 0.032).

The model in Table 3 is an example of a full model including age, sex, and race as covariates. Two examples of notable phenomena detailed in Table 3 are:

1. The most powerful direct associations are those of parent education (individual level) and neighborhood SES (area level). The percentage of one's neighbors

who are white is also a strong predictor, though it is difficult to distinguish its individual associations from those of neighborhood SES.

2. Parent education mediates the association of white race with CNB accuracy. That is, the direct effect of white race on parent education is 0.253, and the direct effect of parent education on CNB accuracy is 0.229, for a combined (mediated) association of $0.253 \times 0.229 = 0.058$. Note that the direct association of white race with CNB accuracy is 0.083, which is only slightly larger than the mediated association. Thus, someone modeling only the direct association of white race with CNB accuracy (ignoring parent education) without considering mediating effects would acquire an incomplete picture of the overall phenomena. Indeed, the associations of other important variables (e.g. mother age at birth) in the present model make it clear that mediating effects need to be modeled. Detection of such mediations is a key strength of structural modeling of the type presented here.

Performing multiple hierarchical linear regressions for specific associations (including interactions) would indicate whether the sample should be stratified (e.g. modeling males and females separately) to further investigate the associations of demographic variables with the phenomena being modeled. Most significant interactions will suggest that stratified models are necessary.

Discussion

Previous literature corroborates our findings that individual-level socio-demographic characteristics (e.g. race and gender) of our youth participants, and aspects of their familial social capital (e.g. parental education) have statistical relationships with their neurocognitive performance (Hackman & Farah, 2009; Hackman *et al.* 2010). The importance of neighborhood-level demography and crime, to further characterize the environment around these youth at the time of entry into the cohort, has also been noted (Noble *et al.* 2007; McEwen & Gianaros, 2010). This work presents a novel conceptual approach to context ualizing neurodevelopmental assessment, in that it attempts to incorporate proximal (direct cognitive performance measurements), intermediate (individual socio-demographic and familial attributes) and distal (neighborhood-level demography) characteristics along the continuum of social determinants of health (see Warnecke *et al.* 2008 for elaboration of terms).

The two factors identified utilizing exploratory factor analysis of the ACS data for our cohort study participants align with previous social epidemiology research on complex diseases. The factors highlight the importance of neighborhood-level SES, household composition, and language. Language spoken, which reflects the density of immigrants, is likely a proxy for more complex constructs that we are unable to measure without ethnographic methods (e.g. heritage based norms, social support network dynamics that impact rearing, and acculturation).

Healthy behaviors associated with mental well-being, such as participation in the arts and physical activity, are negatively associated with crime levels (Ferreira *et al.* 2007; McGinn *et*

al. 2008; Lovasi *et al.* 2009). People in crime-ridden areas are less likely to participate in healthy lifestyles and are more likely to feel stressed and depressed (Branas *et al.* 2011). Of particular interest in our data are the strong loadings of domestic crimes (e.g. abuse and disturbance), because these crimes would likely be the most disruptive to a young person's sense of security and would likely be associated with poor mental health prognoses. Our finding that higher crime scores were associated, directly and indirectly, with lower CNB performance, is consistent with the aforementioned literature and is an important addition to our understanding of the role of social context in cognitive development.

Our multilevel models demonstrate the importance of accounting for the often complex mediating or confounding relationships between individual and neighborhood-level factors and age- and gender-related neurocognitive developmental milestones. Parental educational attainment emerges as a key example of a complex mediator of CNB performance. Our findings suggest that the developmental (household) environment created for the developing youth is a manifestation of the parents' education. The results also suggest that parental education mediates race effects, e.g. white race is associated with higher parental education and better CNB accuracy.

Further research on the resilience of those non-white youth who had high accuracy despite lower parental education may be key to developing interventions that address the need to improve parental achievement for the sake of youth cognitive development. Such interventions might also target directly involved youths whose parents have low educational attainment to supplement their environments. Parental marital status, overall household composition and maternal age at birth are linked to parental educational achievement. The temporality and directionality of those relationships require further research in this cohort. Nonetheless, these variables are associated with household SES and neighborhood composition. Neighborhood SES and composition are significant predictors of CNB performance and thus worthy targets for intervention to reduce disparities in assessment performance and improve the overall mental well-being of youth.

Acknowledgments

References

Barros AJ, Victora CG (2005). A nationwide wealth score based on the 2000 Brazilian demographic census. Revista de Saúde Pública 39, 523–529. [PubMed: 16113899]

Bentler PM, Yuan K-H (1998). Tests for linear trend in the smallest eigenvalues of the correlation matrix. Psychometrika 63, 131–144.

Berkman LF, Kawachi I, Glymour M (eds) (2014). Social Epidemiology. Oxford University Press: New York, NY.

Branas CC, Cheney RA, MacDonald JM, Tam VW, Jackson TD, Ten Have TR (2011). A difference-in-differences analysis of health, safety, and greening vacant urban space. American Journal of Epidemiology 171, 1296–1306.

Calkins ME, Merikangas KR, Moore TM, Burstein M, Behr MA, Satterthwaite TD, Ruparel K, Wolf DH, Roalf DR, Menth FD, Qiu H, Chiavacci R, Connolly JJ, Sleiman PMA, Gur RC, Hakonarson H, Gur RE (2015). The Philadelphia Neurodevelopmental Cohort: constructing a deep phenotyping

collaborative. Journal of Child Psychology and Psychiatry. Published online: 10 May 2015, doi:10.1111/jcpp.12416.

Calkins ME, Moore TM, Merikangas KR, Burstein M, Satterthwaite TD, Bilker WB, Ruparel K, Chiavacci R, Wolf DH, Mentch F, Qiu H, Connolly JJ, Sleiman PA, Hakonarson H, Gur RC, Gur RE (2014). The psychosis spectrum in a young US community sample: findings from the Philadelphia Neurodevelopmental Cohort. World Psychiatry 13, 296–305. [PubMed: 25273303]

Carey GW (1966). The regional interpretation of Manhattan population and housing patterns through factor analysis. Geographical Review 56, 551–569.

Cattell RB (1966). The scree test for the number of factors. Multivariate Behavioral Research 1, 245–276. [PubMed: 26828106]

Ernst JS (2001). Community-level factors and child maltreatment in a suburban county. Social Work Research 25, 133–142.

Ferreira I, Van Der Horst K, Wendel-Vos W, Kremers S, Van Lenthe FJ, Brug J (2007). Environmental correlates of physical activity in youth – a review and update. Obesity Reviews 8, 129–154. [PubMed: 17300279]

Fuentes M, Hart-Johnson T, Green CR (2007). The association among neighborhood socioeconomic status, race and chronic pain in black and white older adults. Journal of the National Medical Association 99, 1160. [PubMed: 17987920]

Greenwood TA, Swerdlow NR, Gur RE, Cadenhead KS, Calkins ME, Dobie DJ, Freedman R, Green MF, Gur RC, Lazzeroni LC, Nuechterlein KH, Olincy A, Radant AD, Ray A, Schork NJ, Seidman LJ, Siever LJ, Silverman JM, Stone WS, Sugar CA, Tsuang DW, Tsuang MT, Turetsky BI, Light GA, Braff DL (2013). Genome-wide linkage analyses of 12 endophenotypes for schizophrenia from the Consortium on the Genetics of Schizophrenia. American Journal of Psychiatry 170, 521–532. [PubMed: 23511790]

Gross KS, McDermott PA (2009). Use of city-archival data to inform dimensional structure of neighborhoods. Journal of Urban Health 86, 161–182. [PubMed: 19104943]

Gur RC, Richard J, Hughett P, Calkins ME, Macy L, Bilker WB, Brensinger C, Gur RE (2010). A cognitive neuroscience-based computerized battery for efficient measurement of individual differences: standardization and initial construct validation. Journal of Neuroscience Methods 187, 254–262. [PubMed: 19945485]

Gur RE, Nimgaonkar VL, Almasy L, Calkins ME, Ragland JD, Pogue-Geile MF, Kanes S, Blanjero J, Gur RC (2007). Neurocognitive endophenotypes in a multiplex multigenerational family study of schizophrenia. American Journal of Psychiatry 164, 813–819. [PubMed: 17475741]

Hackman DA, Farah MJ (2009). Socioeconomic status and the developing brain. Trends in Cognitive Sciences 13, 65–73. [PubMed: 19135405]

Hackman DA, Farah MJ, Meaney MJ (2010). Socioeconomic status and the brain: mechanistic insights from human and animal research. Nature Reviews Neuroscience 11, 651–659. [PubMed: 20725096]

Havard S, Deguen S, Bodin J, Louis K, Laurent O, Bard D (2008). A small-area index of socioeconomic deprivation to capture health inequalities in France. Social Science & Medicine 67, 2007–2016. [PubMed: 18950926]

Herbert DT (1968). Principal components analysis and British studies of urban-social structure. The Professional Geographer 20, 280–283.

Hox JJ (1998). Multilevel modeling: when and why In Classification, Data Analysis, and Data Highways (ed. Balderjahn I, Mathar R and Schader M), pp. 147–154. New York: Springer Verlag.

James SA, Kleinbaum DG (1976). Socioecologic stress and hypertension related mortality rates in North Carolina. American Journal of Public Health 66, 354–358. [PubMed: 1267079]

Jones FL (1965). A social profile of Canberra, 1961. Journal of Sociology 1, 107–120.

Krabbendam L, Hooker CI, Aleman A (2014). Neural effects of the social environment. Schizophrenia Bulletin 40, 248–251. [PubMed: 24442853]

Langlois A, Kitchen P (2001). Identifying and measuring dimensions of urban deprivation in Montreal: an analysis of the 1996 census data. Urban Studies 38, 119–139.

Lauer K (1994). The risk of multiple sclerosis in the USA in relation to sociogeographic features: a factor-analytic study. Journal of Clinical Epidemiology 47, 43–48. [PubMed: 8283194]

Li G, Weng Q (2007). Measuring the quality of life in city of Indianapolis by integration of remote sensing and census data. International Journal of Remote Sensing 28, 249–267.

Lo CP, Faber BJ (1997). Integration of Landsat Thematic Mapper and census data for quality of life assessment. Remote Sensing of Environment 62, 143–157.

Lovasi GS, Hutson MA, Guerra M, Neckerman KM (2009). Built environments and obesity in disadvantaged populations. Epidemiologic Reviews 31, 7–20. [PubMed: 19589839]

Manolio TA, Bailey-Wilson JE, Collins FS (2006). Genes, environment and the value of prospective cohort studies. Nature Reviews Genetics 7, 812–820.

McEwen BS (2012). Brain on stress: how the social environment gets under the skin. Proceedings of the National Academy of Sciences of the United States of America 109(Suppl. 2), 17180–17185. [PubMed: 23045648]

McEwen BS, Gianaros PJ (2010). Central role of the brain in stress and adaptation: links to socioeconomic status, health, and disease. Annals of the New York Academy of Sciences 1186, 190–222. [PubMed: 20201874]

McGinn AP, Evenson KR, Herring AH, Huston SL, Rodriguez DA (2008). The association of perceived and objectively measured crime with physical activity: a cross-sectional analysis. Journal of Physical Activity & Health 5, 117. [PubMed: 18209258]

Mezuk B, Li X, Cederin K, Concha J, Kendler KS, Sundquist J, Sundquist K (2015). Ethnic enclaves and risk of psychiatric disorders among first- and second-generation immigrants in Sweden. Social Psychiatry and Psychiatric Epidemiology. Published online: 27 August 2015. doi:10.1007/s00127-015-1107-1.

Moore TM, Reise SP, Gur RE, Hakonarson H, Gur RC (2015). Psychometric properties of the Penn Computerized Neurocognitive Battery. Neuropsychology 29, 235–246. [PubMed: 25180981]

Muthén LK, Muthén BO (1998–2013). Mplus User's Guide, 7th edn Muthén & Muthén: Los Angeles, CA.

Noble KG, McCandliss BD, Farah MJ (2007). Socioeconomic gradients predict individual differences in neurocognitive abilities. Developmental Science 10, 464–480. [PubMed: 17552936]

R Core Team (2014). R: a Language and Environment for Statistical Computing. R Foundation for Statistical Computing: Vienna, Austria (http://www.R-project.org/).

Ray DM (1971). From factorial to canonical ecology: the spatial interrelationships of economic and cultural differences in Canada. Economic Geography, 47, 344–367.

Revelle W (2013). psych: Procedures for personality and psychological research. Northwestern University: Evanston, Illinois, USA (http://CRAN.R-project.org/package=psych).

Roberts RE, McBee GW (1968). Modernization and economic development in Mexico: a factor analytic approach. Economic Development and Cultural Change, 16, 603–612.

Smoller JW (2015). The genetics of stress-related disorders: PTSD, depression and anxiety disorders. Neuropsychopharmacology. Published online: 31 August 2015. doi:10.1038/npp.2015.266.

Sörbom D (1989). Model modification. Psychometrika 54, 371–384.

Tello JE, Jones J, Bonizzato P, Mazzi M, Amaddeo F, Tansella M (2005). A census-based socio-economic status (SES) index as a tool to examine the relationship between mental health services use and deprivation. Social Science & Medicine 61, 2096–2105. [PubMed: 15922500]

Temkin K, Rohe WM (1998). Social capital and neighborhood stability: an empirical investigation. Housing Policy Debate 9, 61–88.

Thurstone LL (1935). The Vectors of Mind. University of Chicago Press: Chicago.

Vespa J, Lewis JM, Kreider RM (2013). America's families and living arrangements: 2012 In Current Population Reports, pp. 20–570. U.S. Census Bureau: Washington, DC.

Wang MC, Kim S, Gonzalez AA, MacLeod KE, Winkleby MA (2007). Socioeconomic and food-related physical characteristics of the neighbourhood environment are associated with body mass index. Journal of Epidemiology and Community Health 61, 491–498. [PubMed: 17496257]

Warnecke RB, Oh A, Breen N, Gehlert S, Paskett E, Tucker KL, Lurie N, Rebbeck T, Goodwin J, Flack J, Srinivasan S, Kerner J, Heurtin-Roberts S, Abeles R, Tyson FL, Patmios G, Hiatt RA (2008). Approaching health disparities from a population perspective: the National Institutes of

Health Centers for Population Health and Health Disparities. American Journal of Public Health 98, 1608–1615. [PubMed: 18633099]

Yen IH, Syme SL (1999). The social environment and health: a discussion of the epidemiologic literature. Annual Review of Public Health 20, 287–308.

Fig. 1.
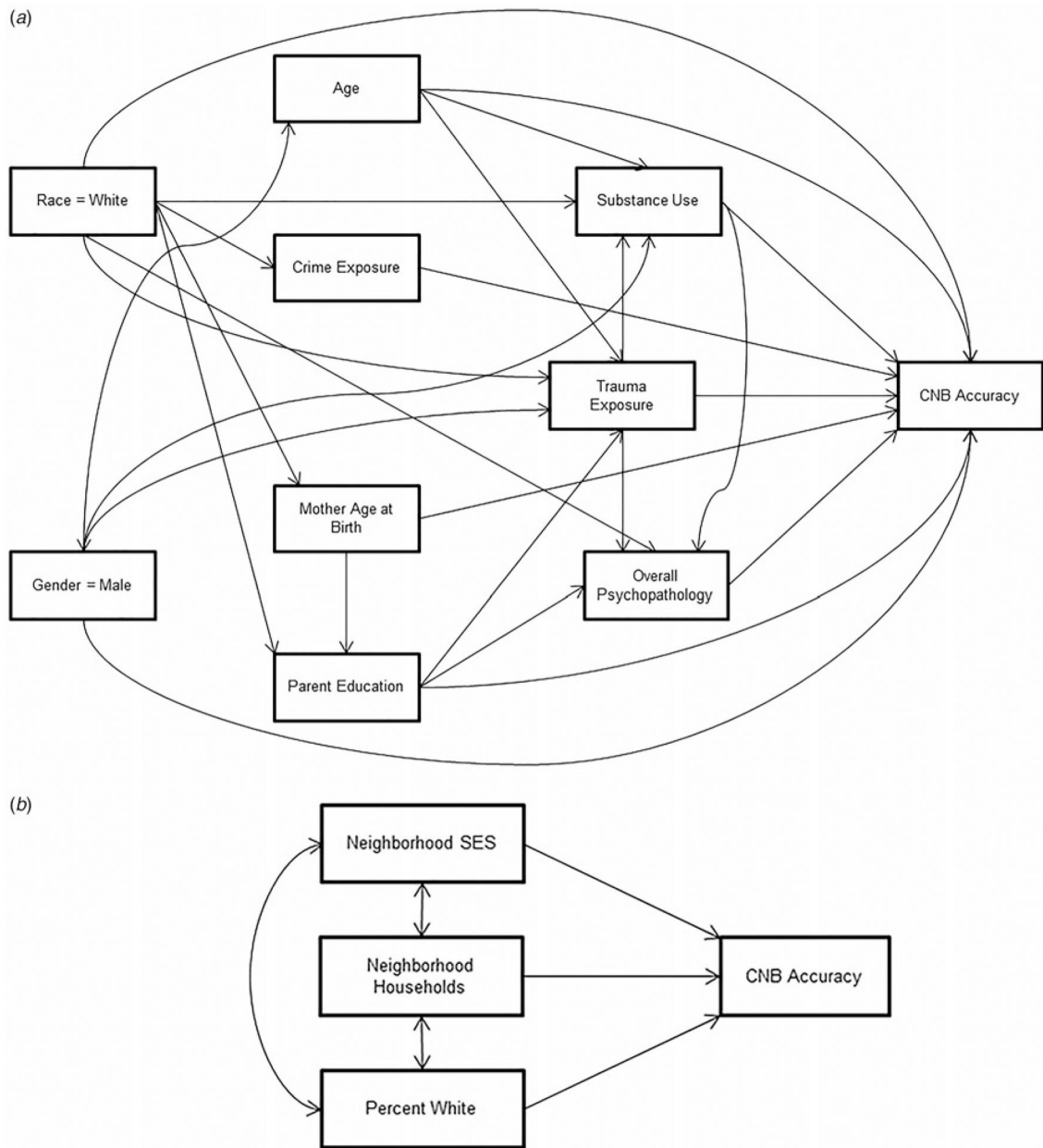(*a*) Structural model (within-level) relating individual-level variables to neurocognitive performance accuracy. (*b*) Structural model (between-level) relating neighborhood-level variables to neurocognitive performance accuracy.
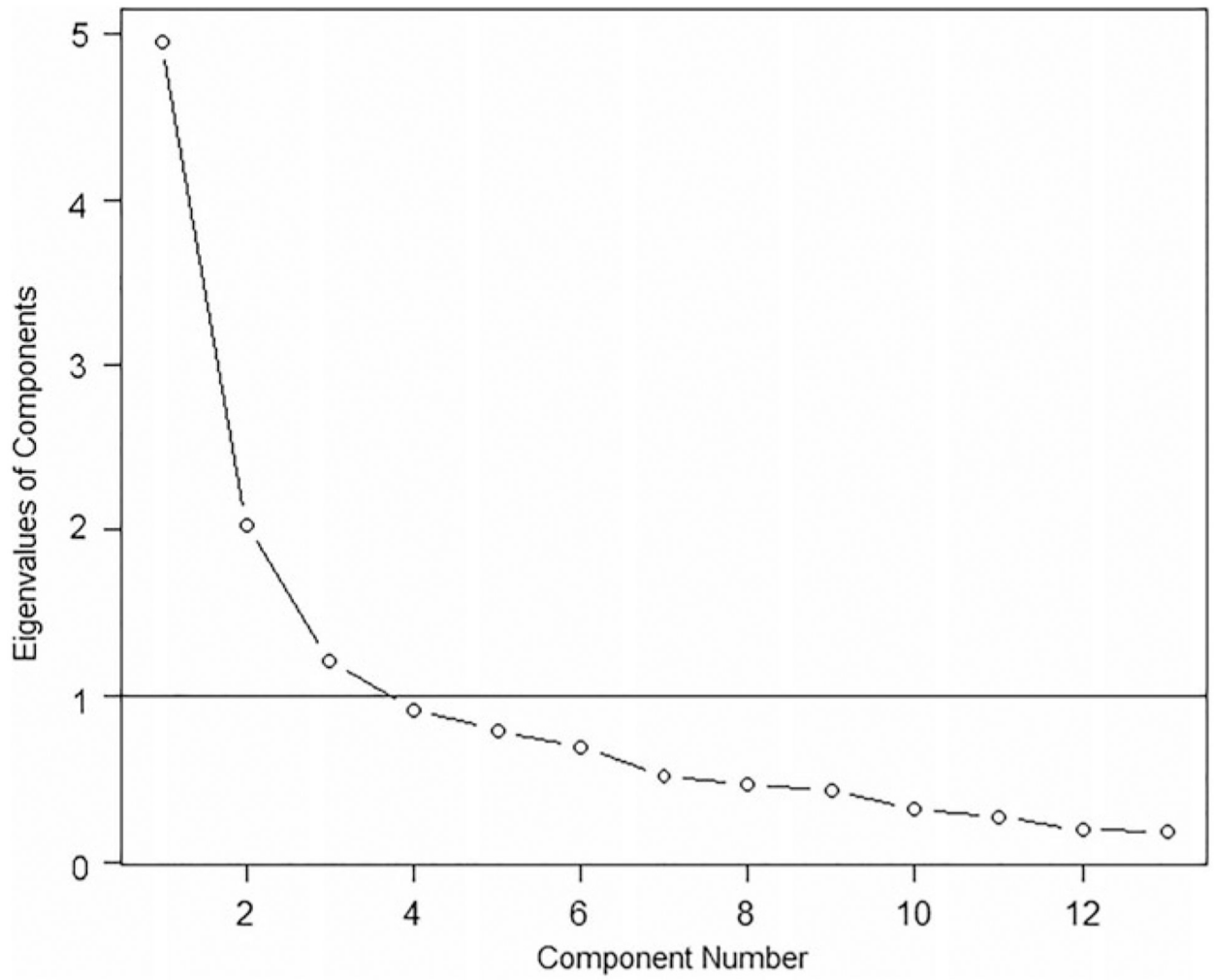
Fig. 2.
Scree plot of eigenvalues for 13 census-based American Community Survey variables.
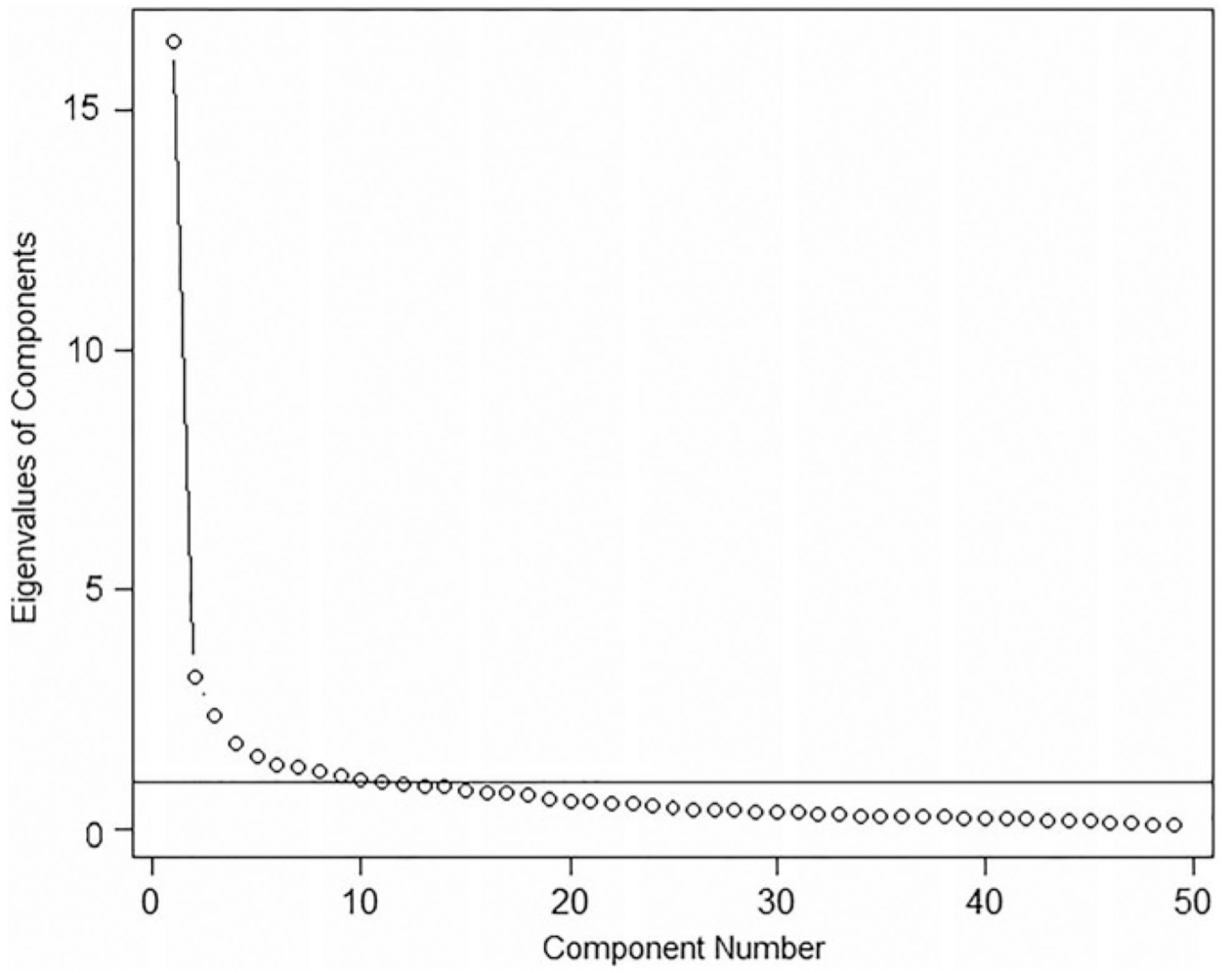
Fig. 3.
Scree plot of eigenvalues for 49 crimes committed in the Philadelphia area.

Table 1.

Unidimensional, two-, and three-factor solutions of the social environment census variables

Variable	Uni	Two-factor		Three-factor		
		F1	F2	F1	F2	F3
Percent married	0.84	0.85		0.68	0.35	
Percent in poverty	−0.86	−0.86		−0.81		
Median family income	0.82	0.82		0.68		
Percent high school plus	0.75	0.74		0.69		
Population density	−0.71	−0.71		−0.53		−0.27
Percent employed	0.66	0.68	−0.31	0.89	−0.31	
Percent vacant lots	−0.60	−0.60		−0.61		
Median age	0.60	0.61				0.92
Percent female	−0.26	−0.26		−0.37		
Percent with children	0.09		0.90		0.83	
Percent English speakers	0.24		−0.54		−0.48	
Avg household size	0.02		0.37		0.41	
Percent non-family households	−0.19		−0.34	−0.42	−0.42	−0.30
Inter-factor correlations (Phi)						
F1		-		-		
F2		0.02	-	0.05	-	
F3				0.47	0.05	-

Uni, unidimensional; Avg, average; extraction method, least squares; rotation, oblimin.

Loadings with absolute value <0.25 removed in the two- and three-factor models.

Table 2.

Unidimensional, two-, and three-factor solutions of 49 crimes (per 100 persons) reported by the Philadelphia Police Department

Crime	Uni F1	Two-factor		Three-factor		
		F1	F2	F1	F2	F3
Domestic abuse	0.81	0.92		0.62		0.39
Agg assault no guns	0.83	0.85		0.61		0.29
Drug possession	0.66	0.85		0.86		
Non-agg assault	0.86	0.77		0.53	0.33	
Agg assault with guns	0.64	0.76		0.44		0.42
Domestic disturbance	0.86	0.76		0.60	0.31	
Curfew violation	0.74	0.74		0.73		
Traffic violation	0.75	0.72		0.79		
Drug sales	0.53	0.72		0.70		
UFA weapons violation	0.61	0.65		0.51		
Missing person	0.60	0.64		0.30		0.42
Animal incidents	0.61	0.62				0.46
Truancy	0.56	0.60		0.62		
DUI	0.72	0.58		0.61		
Tenant landlord violation	0.45	0.57		0.34		0.31
Loitering and prowling	0.59	0.53		0.52		
Robbery with guns	0.72	0.51	0.30	0.36	0.41	
City ordinance violation	0.58	0.51		0.76		−0.29
Vandalism	0.81	0.51	0.44		0.67	0.33
Burglary residential	0.73	0.49	0.34		0.54	0.31
Grand theft auto	0.75	0.49	0.37		0.57	0.30
Harassment	0.63	0.42	0.30		0.46	
Counterfeit and forgery	0.51	0.40		0.50		
Auto-tag theft	0.61	0.36	0.35		0.44	
Parole violation	0.35	0.36		0.60		−0.28
Purse theft	0.43	0.29		0.39		
Prostitution	0.18	0.24		0.25		

Crime	Uni F1	Two-factor		Three-factor		
		F1	F2	F1	F2	F3
Gambling	0.12	0.23		0.22		
Dumping	0.19	0.16		0.12		
False police report	0.07	0.10		0.16		
Dangerous dog	0.04	0.08		0.19		
Theft from auto	0.53		0.88	−0.26	0.95	
Auto accidents	0.71		0.69		0.73	
Embezzlement	0.26		0.66		0.56	−0.40
Bicycle-theft	0.63		0.62		0.70	
Graffiti	0.35		0.59		0.56	
Lost property	0.77	0.37	0.55	0.32	0.61	
Report to city department	0.66	0.28	0.53		0.58	
Check fraud	0.75	0.39	0.50		0.66	
Robbery no guns	0.76	0.42	0.47	0.46	0.48	
Burglary commercial	0.43		0.46		0.44	
Criminal trespass	0.47		0.41		0.44	
Pickpocketing	0.23		0.38			−0.46
Retail theft	0.32		0.37	0.28	0.26	−0.38
Public drunkenness	0.24		0.31		0.28	
Abduction	0.19		0.27		0.32	
Arson	0.22		0.21		0.21	
Receiving stolen goods	0.17		0.21	0.28		−0.35
Liquor law violations	0.14		0.08		0.09	

Inter-factor correlations

	F1	F2	F1	F2	F3
F1	1		1		
F2	0.53	1	0.56	1	
F3			0.29	0.17	1

Agg, Aggravated; UFA, Uniform Firearms Act, DUI, driving under the influence of a substance. Loadings <0.25 removed unless all loadings in that row were <0.25.

Table 3.

Standardized path coefficients for structural model predicting CNB accuracy across all ages, races, and genders

Variable	Effect	Variable	Std. coef.	p value
Parent education	↑	CNB accuracy	0.229	<0.0005
Trauma exposure	↑	CNB accuracy	−0.011	0.418
Mother age at birth	↑	CNB accuracy	0.055	<0.0005
Substance use	↑	CNB accuracy	0.018	0.354
White race	↑	CNB accuracy	0.083	<0.0005
Age	↑	CNB accuracy	0.013	0.412
Male sex	↑	CNB accuracy	−0.008	0.532
Crime exposure	↑	CNB accuracy	0.024	0.361
Psychopathology	↑	CNB accuracy	0.025	0.054
White race	↑	Substance use	0.105	<0.0005
Male sex	↑	Substance use	0.026	0.052
Trauma exposure	↑	Substance use	0.218	<0.0005
Age	↑	Substance use	0.403	<0.0005
White race	↑	Trauma exposure	−0.142	<0.0005
Male sex	↑	Trauma exposure	0.068	<0.0005
Age	↑	Trauma exposure	0.221	<0.0005
Parent education	↑	Trauma exposure	−0.101	<0.0005
White race	↑	Mother age at birth	0.367	<0.0005
Mother age at birth	↑	Parent education	0.301	<0.0005
White race	↑	Parent education	0.253	<0.0005
White race	↑	Crime exposure	−0.270	<0.0005
White race	↑	Psychopathology	−0.064	<0.0005
Parent education	↑	Psychopathology	−0.054	<0.0005
Trauma exposure	↑	Psychopathology	0.336	<0.0005
Substance use	↑	Psychopathology	0.116	<0.0005
Age	↔	Male sex	−0.067	<0.0005
Neighborhood SES	↑	CNB accuracy	0.406	<0.0005
Neighborhood liam	↑	CNB accuracy	−0.049	0.346

Variable	Effect	Variable	Std. coef.	p value
Percent white	→	CNB accuracy	0.347	<0.0005
Percent white	↔	*Neighborhood SES*	0.754	<0.0005
Neighborhood fam	↔	*Neighborhood SES*	0.044	0.040
Neighborhood fam	↔	*Percent white*	−0.046	0.020

CNB, Computerized Neurocognitive Battery; Std. coef., standardized coefficient; Fam, family; SES, socioeconomic status. *Italics* indicate variables measured at the neighborhood level; '→' indicates a direct effect; '↔' indicates a correlation.