**The association of Early Life State, Work Context and Neighborhood Context with Hypertension, Diabetes and Ischemic Heart Disease**

David H Rehkopf, Mark R Cullen

School of Medicine, Division of General Medical Disciplines

1265 Welch Road, MSOB 3rd Floor, MC 5411

Stanford, CA 94305 USA

650.725.0356

drehkopf@stanford.edu

Extended Abstract prepared for Population Association of American 2015 Annual Meeting

Draft date: September 26, 2014

**Abstract**

In the majority of studies on context and health, the focus, primarily due to data limitations, has been on a single type of context, most commonly neighborhood. Yet in addition to neighborhood of residence, the alternative environment of extensive exposure is the workplace. We use a unique data linkage to combine data on early life state of residence, work environment, and neighborhood environment for blue and white collar workers that were followed from 1996 to 2012, with 305,936 person-years of observation. Disease was assessed using medical claims. In our initial analyses we find small but statistically significant associations between early state of residence characteristics and later life hypertension, diabetes and ischemic heart disease. The most consistent associations were with income inequality and percent white. Our next analyses will examine trajectories from early life state to workplace and neighborhood, and the relative strength with which each predicts chronic disease outcomes.

## Introduction

Environment, whether at the level of country, state, county or neighborhood, accounts for a meaningful amount of variance in chronic disease (1-8). While earlier work was entirely ecological in nature, thus making inference to individual health outcomes problematic, more recent literature has avoided ecological bias by measuring individual level outcomes and covariates (4, 9, 10). In the United States, state is a level of geography that has varied social characteristics and policy heterogeneity – "It is one of the happy incidents of the federal system that a single courageous state may, if its citizens choose, serve as a laboratory; and try novel social and economic experiments without risk to the rest of the country" (11). The state level is also methodologically advantageous as a level of analysis because there is likely less severe self-selection bias as compared to smaller regions within metropolitan areas (12), such as neighborhood, where wealth and race are substantial drivers of within metropolitan area residential location due to socially and materially influenced housing choice (10). A further notable emphasis in the current literature is that, most typically, due to data constraints, only current context is examined, which is inconsistent with life course theories of the etiology of chronic disease (13-17). It is possible that the associations found with current measures of social and economic environment are at least in part due to impacts of earlier environments that are strongly correlated with current context as many people remain stable geographically (18).

For these reasons, the current analysis begins with state early in life. Specifically, the study examines how early life state characteristics are associated with prevalent hypertension, type 2 diabetes and ischemic heart disease. We are able to construct this environment through data linkage of individuals in a well-characterized occupational cohort using the first three digits of social security numbers to determine state of early life residence (24). The next steps of the analysis which we will present examine the extent to which these early life characteristics predict the type of neighborhood individuals live in later in life, and the type of work environments, both physical and psychological, that they are exposed to later in life. Finally, we will compare the strength with each of these types of environments are associated with chronic disease outcomes, and use machine learning methods to explore potential higher order interactions between these environments with each other and with individual level social and demographic characteristics of individuals.

**Methods**

The study population is from an occupational cohort obtained from United States manufacturing plants with a total analytic sample size of 40,804, with 305,936 person years of observation. Data were collected on all employees beginning in 1996 through their date of retirement. Individuals were censored in 2012. The data come from 62 manufacturing plants across the United States, with at least 100 individuals per plant (smaller plants were excluded from the current analysis). We used the first three digits of employee social security numbers to determine the state where their social security card was issued (25). Fewer than 1% of our sample could not be linked. We chose the following state characteristics based on previous findings in the literature as important indicators of social condition that were also available as consistently measured in the U.S. census since 1940: percent urban, percent with less than a high school level of education, median income in year 2000 dollars, unemployment rate, percent white and Gini coefficient of income inequality. Supplemental figure S1 presents a Pearson correlation plot showing how strongly each of the early life state and current state measures are correlated, with more intense colors and larger dots indicating stronger correlation, as shown in the scale to the right of the figure.

We used administrative records to obtain individual level covariates and outcome data. These data include age, race, gender, whether the individual was an hourly or salaried worker and a continuous measure of employment grade. Prior work has shown that level of employment grade, which ranges from 1 to 76 and whether the individual was an hourly worker are strong predictors of chronic disease health outcomes (26). Prevalence of health outcomes were defined as having two or more diagnoses for each of the health outcomes of analysis over the study period: hypertension, type 2 diabetes mellitus and ischemic heart disease. Prior work has determined the accuracy of medical claims data for identifying health outcomes from administrative data (28).

All statistical models were fit within a pooled logistic regression framework and random effects were used to account for the clustering within both individuals (n=40,804) and manufacturing plant (n=62) (a three level model). We fit four primary models for each of the three health outcomes focusing on interpreting the coefficients for state characteristics in early life (coefficients for other covariates in the models are available from authors). In addition to the six early life state characteristics previously described, the models contain the following covariates: Model one included age, age-squared, race/Ethnicity (black, Latino or other) and gender. Model two additionally included the smelter workplace indicator, and a four-category early life census region. Model three additionally includes current state characteristics. Model four additionally includes whether the individual was an hourly worker and the individual level employment grade. In order to account for differential censoring within the cohort, we fit all models using inverse probability of censoring weights.

**Results**

Table 1 shows the demographic, geographic and health characteristics of the sample, which is more white, more male, and more working class than the general U.S. population. Table 2 presents the coefficients from random effect regression models with hypertension, diabetes and ischemic heart disease as the dependent variables. Each set of six rows of coefficients per outcome are from a single regression model, thus this table shows results from the four different models for each of 3 outcomes (12 different regressions). For hypertension, all six measures of the early life state environment are statistically significantly associated with the outcome in each of the four models. For model one, the model of our primary focus, a higher % white, more urban, more high school graduates, a lower percent of unemployed, a higher median income and a lower Gini are associated with lower levels of hypertension. Results are similar for model 2, but the direction of association reverses for high school and median income after controlling for census region. Both models 3 and 4 are similar in direction and magnitude to model 2, although the direction of sign changes for unemployment. The regression coefficients indicate the log odds of hypertension associated with a 1 standard deviation difference in the contextual factor. For example, the coefficient for Gini in model 1 is 0.018, which translates to a 1.02 odds ratio for hypertension associated with one standard deviation difference in the Gini. For diabetes, only percent unemployed and the Gini measure of income inequality are statistically significantly associated in the first model, and the associations remain of similar magnitude after controlling for additional covariates in models two, three and four. After controlling for region in Model 2, a higher percent of white in early life state of residence is associated with increased prevalence of diabetes. For ischemic heart disease, all of the early life state characteristics except percent urban are associated in model 1. After controlling for region, median income is no longer associated, and after controlling for current state characteristics percent high school is no longer associated. In the final model 4, only percent white and Gini remain associated with ischemic heart disease prevalence later in life. The figure presents the data from table 2, model 4, but with using Gini and unemployment on a scale running from high to low to be consistent with the other measures.

**Discussion**

We show for the first time how measures of early life state social context remain associated with chronic disease many years later, even after statistical control for individual sociodemographic characteristics and characteristics of current state context. The most consistent associations across models and health outcomes were for percent white in early life state and the Gini index of income inequality, although the direction of association with percent white varied by outcome examined. These findings are unlikely to be driven by migration, as associations were generally consistent or stronger when examined within the strata of the population who did not move states, thus associations conditional on current state were based on changes in state characteristics over time. Finally, most measures of association did not change when including current health behavior and biomarkers in the statistical models. The exception to this was the Gini, for which associations with hypertension, diabetes and ischemic heart disease were all substantially attenuated, suggesting that a potential pathway through which factors correlated with early life state income inequality impacts current chronic disease may be through current health behaviors, as also shown in other work (29). Finally, even without identifying mechanisms or specific causal factors, our findings, and the method of using the first three digits of social security numbers, can be used to inexpensively obtain information on factors that may act as confounders in studies of contemporaneous environments and chronic disease.

**Tables**

Table 1. Demographic characteristics, Alcoa cohort, 1996-2012

| | Mean/percent (n=40,804) |
|---|---|
| age (mean) | 49 |
| male | 78% |
| race | |
| black | 9.8% |
| white | 75% |
| latino | 11% |
| other | 4.2% |
| hourly worker | 71% |
| smelter worker | 31% |
| current census region | |
| west | 17% |
| central | 36% |
| east | 14% |
| south | 33% |
| early life census region | |
| west | 8.1% |
| central | 37% |
| east | 18% |
| south | 37% |
| hypertension prevalence | 16% |
| diabetes prevalence | 5.7% |
| ischemic heart disease prevalence | 3.5% |

table note: person years of observation are 305,936

Table 2. Association of early life state characteristics with later life health, Alcoa cohort, 1997-2012

| | Model 1 | | | Model 2 | | | Model 3 | | | Model 4 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Hypertension | estimate | SE | p-value | estimate | SE | p-value | estimate | SE | p-value | estimate | SE | p-value |
| % white | -0.0080* | 0.0007 | <.0001 | -0.0094* | 0.0008 | <.0001 | -0.0075* | 0.0009 | <.0001 | -0.0079* | 0.0010 | <.0001 |
| % urban | -0.0028* | 0.0007 | 0.0001 | -0.0040* | 0.0008 | <.0001 | -0.0112* | 0.0010 | <.0001 | -0.0111* | 0.0010 | <.0001 |
| % high school | -0.0058 | 0.0018 | 0.0011 | 0.0202* | 0.0022 | <.0001 | 0.0326* | 0.0023 | <.0001 | 0.0316* | 0.0023 | <.0001 |
| % unemployed | 0.0020 | 0.0007 | 0.0027 | 0.0019 | 0.0007 | 0.0045 | -0.0035* | 0.0007 | <.0001 | -0.0033* | 0.0008 | <.0001 |
| median income | 0.0068* | 0.0020 | 0.0006 | -0.0063 | 0.0021 | 0.0034 | -0.0078 | 0.0024 | 0.0010 | -0.0081 | 0.0024 | 0.0007 |
| Gini | 0.0188* | 0.0012 | <.0001 | 0.0156* | 0.0014 | <.0001 | 0.0132* | 0.0016 | <.0001 | 0.0132* | 0.0016 | <.0001 |
| Diabetes | | | | | | | | | | | | |
| % white | 0.0008 | 0.0004 | 0.0590 | 0.0012 | 0.0005 | 0.0080 | 0.0023* | 0.0006 | <.0001 | 0.0022* | 0.0006 | <.0001 |
| % urban | 0.0001 | 0.0004 | 0.7342 | 0.0008 | 0.0005 | 0.1034 | 0.0007 | 0.0006 | 0.2503 | 0.0009 | 0.0006 | 0.1331 |
| % high school | 0.0001 | 0.0010 | 0.9053 | 0.0024 | 0.0013 | 0.0528 | 0.0027 | 0.0013 | 0.0420 | 0.0024 | 0.0013 | 0.0765 |
| % unemployed | 0.0011 | 0.0004 | 0.0034 | 0.0014* | 0.0004 | 0.0004 | 0.0015 | 0.0004 | 0.0007 | 0.0014 | 0.0004 | 0.0011 |
| median income | -0.0003 | 0.0012 | 0.8164 | -0.0008 | 0.0012 | 0.5327 | -0.0023 | 0.0014 | 0.0929 | -0.0028 | 0.0014 | 0.0426 |
| Gini | 0.0032* | 0.0007 | <.0001 | 0.0021 | 0.0008 | 0.0068 | 0.0037* | 0.0009 | <.0001 | 0.0039* | 0.0009 | <.0001 |
| Ischemic Heart Disease | | | | | | | | | | | | |
| % white | 0.0010* | 0.0003 | 0.0001 | 0.0011* | 0.0003 | 0.0001 | 0.0018* | 0.0003 | <.0001 | 0.0018* | 0.0003 | <.0001 |
| % urban | 0.0005 | 0.0003 | 0.062 | 0.0009 | 0.0003 | 0.0013 | 0.0008 | 0.0004 | 0.0238 | 0.0008 | 0.0004 | 0.0225 |
| % high school | -0.0041* | 0.0006 | <.0001 | -0.0025 | 0.0008 | 0.0012 | -0.0016 | 0.0008 | 0.0575 | -0.0018 | 0.0008 | 0.0286 |
| % unemployed | 0.0006 | 0.0002 | 0.0053 | 0.0008* | 0.0002 | 0.0005 | 0.0002 | 0.0003 | 0.3799 | 0.0003 | 0.0003 | 0.3445 |
| median income | 0.0021 | 0.0007 | 0.003 | 0.0015 | 0.0008 | 0.0516 | 0.0015 | 0.0008 | 0.0698 | 0.0017 | 0.0008 | 0.0466 |
| Gini | 0.0019* | 0.0004 | <.0001 | 0.0015 | 0.0005 | 0.0026 | 0.0015 | 0.0006 | 0.0068 | 0.0015 | 0.0006 | 0.0100 |

table notes: Each set of six rows of coefficients per outcome (and SE, standard error) are from a single regression model, thus this table shows results from 12 different regression models (four models for each of three outcomes). All models include random effects for individual and plant location. In addition to the regression coefficients shown here, the models contain the following covariates: Model one includes age, age-squared, and race (black, Latino or other) and gender. Model two additionally includes smelter workplace indicator, and the four category early life census region. Model three additionally includes current state characteristics (the same six as the early life characteristics). Model four additionally includes whether the individual was an hourly work and their employment grade. Estimates marked with an * indicate statistically significant at $p<0.05$ after Bonferroni adjustment (equivalent p-value of $p<0.0006944$) for multiple hypothesis testing.

**Figure**

Radar plot of Odds ratios of association of early life state characteristics with later life health, Alcoa cohort, 1997-2012

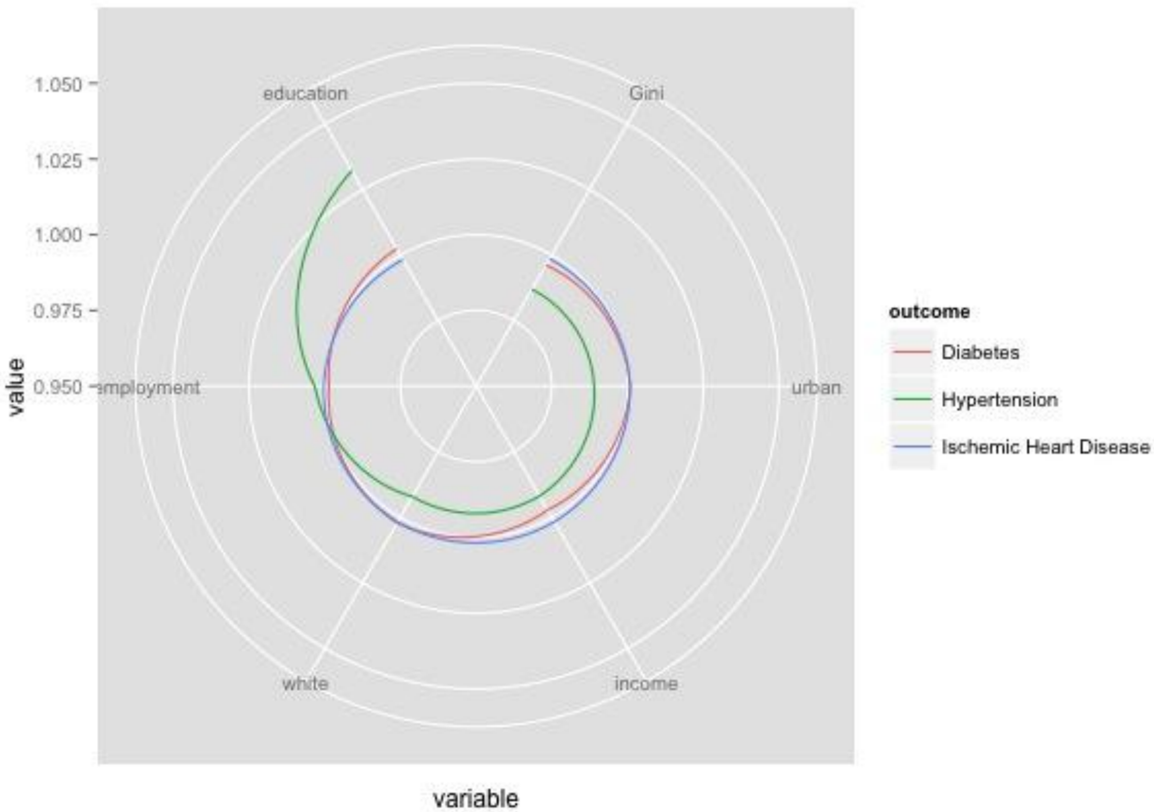


Figure notes: The figure presents the data from table 2, model 4, but converts coefficients to odds ratios, and using Gini and unemployment on a scale running from high to low to be consistent with the other measures. The location of the colored line along radii corresponds to odds ratios of association between the factor indicated on each radii and the health outcome as indicated in the legend to the right. The order of radii from 0 degrees going clockwise is of factors associated with decreased to increased risk of hypertension. The values on the left hand side of the plot indicate magnitude of odds ratio associated with each outcome. The range is from an odds ratio of 0.95 (center of circle) to 1.05 (outer ring of circle).

## References


1. Cullen MR, Cummins C, Fuchs VR. Geographic and racial variation in premature mortality in the U.S.: analyzing the disparities. PLoS One. 2012;7(4):e32930. doi: 10.1371/journal.pone.0032930. PubMed PMID: 22529892; PubMed Central PMCID: PMC3328498.
2. Murray CJ, Kulkarni SC, Michaud C, Tomijima N, Bulzacchelli MT, Iandiorio TJ, et al. Eight Americas: investigating mortality disparities across races, counties, and race-counties in the United States. PLoS Med. 2006;3(9):e260. Epub 2006/09/14. doi: 05-PLME-RA-0571R2 [pii] 10.1371/journal.pmed.0030260. PubMed PMID: 16968116.
3. Chen JT, Rehkopf DH, Waterman PD, Subramanian SV, Coull BA, Cohen B, et al. Mapping and measuring social disparities in premature mortality: the impact of census tract poverty within and across Boston neighborhoods, 1999-2001. J Urban Health. 2006;83(6):1063-84. Epub 2006/09/27. doi: 10.1007/s11524-006-9089-7. PubMed PMID: 17001522.
4. Diez Roux AV. Estimating neighborhood health effects: the challenges of causal inference in a complex world. Soc Sci Med. 2004;58(10):1953-60. PubMed PMID: 15020010.
5. Leventhal T, Brooks-Gunn J. Moving to opportunity: an experimental study of neighborhood effects on mental health. Am J Public Health. 2003;93(9):1576-82. Epub 2003/09/02. PubMed PMID: 12948983.
6. Diez Roux AV, Merkin SS, Arnett D, Chambless L, Massing M, Nieto FJ, et al. Neighborhood of residence and incidence of coronary heart disease. N Engl J Med. 2001;345(2):99-106. PubMed PMID: 11450679.
7. Yen IH, Kaplan GA. Neighborhood social environment and risk of death: multilevel evidence from the Alameda County Study. Am J Epidemiol. 1999;149(10):898-907. Epub 1999/05/26. PubMed PMID: 10342798.
8. Diez-Roux AV, Nieto FJ, Muntaner C, Tyroler HA, Comstock GW, Shahar E, et al. Neighborhood environments and coronary heart disease: a multilevel analysis. Am J Epidemiol. 1997;146(1):48-63. PubMed PMID: 9215223.
9. Diez Roux AV. Investigating neighborhood and area effects on health. Am J Public Health. 2001;91(11):1783-9. PubMed PMID: 11684601.
10. Houseman EA, Christensen BC, Yeh RF, Marsit CJ, Karagas MR, Wrensch M, et al. Model-based clustering of DNA methylation array data: a recursive-partitioning algorithm for high-dimensional data arising as a mixture of beta distributions. BMC Bioinformatics. 2008;9:365. Epub 2008/09/11. doi: 10.1186/1471-2105-9-365
1471-2105-9-365 [pii]. PubMed PMID: 18782434; PubMed Central PMCID: PMC2553421.
11. NEW STATE ICE CO. V. LIEBMANN, 285 U.S. 262, (1932).
12. Smith GD. Equal, but different? Ecological, individual and instrumental approaches to understanding determinants of health. Int J Epidemiol. 2005;34(6):1179-80. PubMed PMID: 16416513.
13. Power C, Kuh D, Morton S. From developmental origins of adult disease to life course research on adult disease and aging: insights from birth cohort studies. Annu Rev Public Health. 2013;34:7-28. doi: 10.1146/annurev-publhealth-031912-114423. PubMed PMID: 23514315.
14. Lawlor DA, Patel R, Fraser A, Smith GD, Ebrahim S. The association of life course socio-economic position with diagnosis, treatment, control and survival of women with diabetes: findings from the British Women's Heart and Health Study. Diabet Med. 2007;24(8):892-900. doi: 10.1111/j.1464-5491.2007.02187.x. PubMed PMID: 17509070.
15. Lynch J, Smith GD. A life course approach to chronic disease epidemiology. Annu Rev Public Health. 2005;26:1-35.
16. Kuh D, Ben-Shlomo Y, Lynch J, Hallqvist J, Power C. Life course epidemiology. Journal of epidemiology and community health. 2003;57(10):778.
17. Harper S, Lynch J, Hsu WL, Everson SA, Hillemeier MM, Raghunathan TE, et al. Life course socioeconomic conditions and adult psychosocial functioning. Int J Epidemiol. 2002;31(2):395-403. PubMed PMID: 11980802.

18. Gilman SE, Loucks EB. Invited commentary: does the childhood environment influence the association between every x and every y in adulthood? Am J Epidemiol. 2012;176(8):684-8. doi: 10.1093/aje/kws228. PubMed PMID: 23024136; PubMed Central PMCID: PMC3571248.
19. Goodin DS, Jones J, Li D, Traboulsee A, Reder AT, Beckmann K, et al. Establishing long-term efficacy in chronic disease: use of recursive partitioning and propensity score adjustment to estimate outcome in MS. PLoS One. 2011;6(11):e22444. Epub 2011/12/06. doi: 10.1371/journal.pone.0022444 PONE-D-11-05335 [pii]. PubMed PMID: 22140424; PubMed Central PMCID: PMC3227563.
20. Cohen AK, Rehkopf DH, Deardorff J, Abrams B. Education and obesity at age 40 among American adults. Soc Sci Med. 2013;78:34-41. doi: 10.1016/j.socscimed.2012.11.025. PubMed PMID: 23246398; PubMed Central PMCID: PMC3545063.
21. Musick K, Seltzer JA, Schwartz CR. Neighborhood Norms and Substance Use among Teens. Soc Sci Res. 2008;37(1):138-55. Epub 2008/05/23. doi: 10.1016/j.ssresearch.2007.02.003. PubMed PMID: 18496598.
22. Molnar BE, Cerda M, Roberts AL, Buka SL. Effects of neighborhood resources on aggressive and delinquent behaviors among urban youths. Am J Public Health. 2008;98(6):1086-93. Epub 2007/09/29. doi: AJPH.2006.098913 [pii]
10.2105/AJPH.2006.098913. PubMed PMID: 17901441.
23. Dubowitz T, Subramanian SV, Acevedo-Garcia D, Osypuk TL, Peterson KE. Individual and neighborhood differences in diet among low-income foreign and U.S.-born women. Womens Health Issues. 2008;18(3):181-90. Epub 2008/01/29. doi: S1049-3867(07)00182-X [pii]
10.1016/j.whi.2007.11.001. PubMed PMID: 18222706.
24. Block G, Matanoski GM, Seltser RS. A method for estimating year of birth using social security number. Am J Epidemiol. 1983;118(3):377-95. PubMed PMID: 6613981.
25. SSA. Social Security Numbers. Available from: http://www.ssa.gov/history/ssn/geocard.html.
26. Clougherty JE, Eisen EA, Slade MD, Kawachi I, Cullen MR. Gender and sex differences in job status and hypertension. Occup Environ Med. 2011;68(1):16-23. doi: 10.1136/oem.2009.049908. PubMed PMID: 20864467; PubMed Central PMCID: PMC3718674.
27. Neophytou A, Brown D, Costello S, Noth E, Hammond K, Cullen M, et al. 0223 Marginal Structural Models in Occupational Epidemiology: An Application in the US Aluminium industry. Occup Environ Med. 2014;71 Suppl 1:A30. doi: 10.1136/oemed-2014-102362.94. PubMed PMID: 25018317.
28. Solberg LI, Engebretson KI, Sperl-Hillen JM, Hroscikoski MC, O'Connor PJ. Are claims data accurate enough to identify patients for performance measures or quality improvement? The case of diabetes, heart disease, and depression. American journal of medical quality : the official journal of the American College of Medical Quality. 2006;21(4):238-45. doi: 10.1177/1062860606288243. PubMed PMID: 16849780.
29. Modrek S, Ahern J. Longitudinal relation of community-level income inequality and mortality in Costa Rica. Health & place. 2011;17(6):1249-57. doi: 10.1016/j.healthplace.2011.07.006. PubMed PMID: 21873102; PubMed Central PMCID: PMC3190651.
30. Eriksson J, Forsen T, Tuomilehto J, Osmond C, Barker D. Fetal and childhood growth and hypertension in adult life. Hypertension. 2000;36(5):790-4. PubMed PMID: 11082144.
31. Huxley R, Owen CG, Whincup PH, Cook DG, Colman S, Collins R. Birth weight and subsequent cholesterol levels: exploration of the "fetal origins" hypothesis. JAMA. 2004;292(22):2755-64. doi: 10.1001/jama.292.22.2755. PubMed PMID: 15585736.
32. Huxley R, Neil A, Collins R. Unravelling the fetal origins hypothesis: is there really an inverse association between birthweight and subsequent blood pressure? Lancet. 2002;360(9334):659-65. doi: 10.1016/S0140-6736(02)09834-3. PubMed PMID: 12241871.
33. Imai K, Keele L, Tingley D. A general approach to causal mediation analysis. Psychol Methods. 2010;15(4):309-34. doi: 10.1037/a0020761. PubMed PMID: 20954780.
34. Neophytou AM, Picciotto S, Hart JE, Garshick E, Eisen EA, Laden F. A structural approach to address the healthy-worker survivor effect in occupational cohorts: an application in the trucking industry

cohort. Occup Environ Med. 2014;71(6):442-7. doi: 10.1136/oemed-2013-102017. PubMed PMID: 24727736; PubMed Central PMCID: PMC4051133.
35. Costello S, Brown DM, Noth EM, Cantley L, Slade MD, Tessier-Sherman B, et al. Incident ischemic heart disease and recent occupational exposure to particulate matter in an aluminum cohort. Journal of exposure science & environmental epidemiology. 2014;24(1):82-8. doi: 10.1038/jes.2013.47. PubMed PMID: 23982120; PubMed Central PMCID: PMC4045503.
36. Newton-Cheh C, Johnson T, Gateva V, Tobin MD, Bochud M, Coin L, et al. Genome-wide association study identifies eight loci associated with blood pressure. Nat Genet. 2009;41(6):666-76. doi: 10.1038/ng.361. PubMed PMID: 19430483; PubMed Central PMCID: PMC2891673.

**Supplemental Figure**

Figure S1. Correlation of current and early life state characteristics

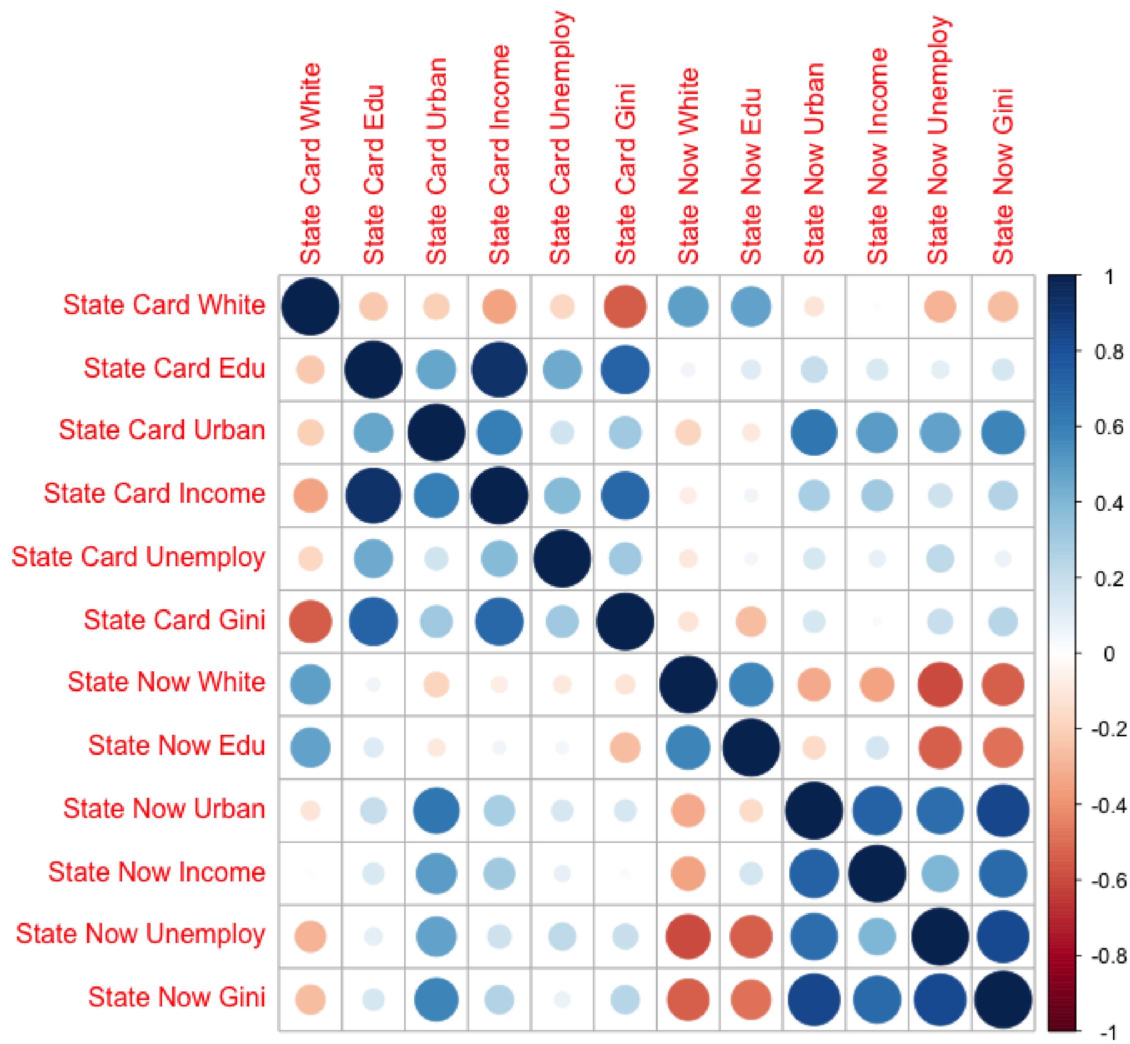


table notes: State Card factors indicate characteristic at the time the social security card was issued, State Now indicates current state characteristics. Due to this being a correlation matrix where factors along the side and top are repeated, the diagonal shows correlations of 1.