Differential epidemiology: IQ, neuroticism, and chronic disease by the 50 U.S. states

Bryan J. Pesta [a,*], Sharon Bertsch [b], Michael A. McDaniel [c],
Christine B. Mahoney [a], Peter J. Poznanski [a]

[a] Cleveland State University, United States
[b] University of Pittsburgh at Johnstown, United States
[c] Virginia Commonwealth University, United States

ABSTRACT

Current research shows that geo-political units (e.g., the 50 U.S. states) vary meaningfully on psychological dimensions like intelligence (IQ) and neuroticism (N). A new scientific discipline has also emerged, differential epidemiology, focused on how psychological variables affect health. We integrate these areas by reporting large correlations between aggregate-level IQ and N (measured for the 50 U.S. states) and state differences in rates of chronic disease (e.g., stroke, heart disease). Controlling for health-related behaviors (e.g., smoking, exercise) reduced but did not eliminate these effects. Strong relationships also existed between IQ, N, disease, and a host of other state-level variables (e.g., income, crime, education). The nexus of inter-correlated state variables could reflect a general fitness factor hypothesized by cognitive epidemiologists, although valid inferences about causality will require more research.

© 2012 Elsevier Inc. All rights reserved.

1. Introduction

The study of individual differences — differential psychology — has recently expanded to include the study of differences across groups of people categorized by shared geography (e.g., states or nations). Aggregate-level measures now exist for both intelligence (IQ) and the *Big Five* personality traits (Lynn & Vanhanen, 2002; McDaniel, 2006; Rentfrow, Gosling, & Potter, 2008). These aggregate-level measures seem to consistently predict important geo-political outcomes, as reviewed below. The goal of the present study is to illustrate the unique capacity aggregate-level psychological variables possess in predicting disease rates across populations (here, the 50 U.S. states). These relationships persist even after controlling for state income levels, and for various health-related behaviors (smoking and exercising) that epidemiologists typically study as disease antecedents. Because we consider both dispositional and cognitive traits, we

term this area *differential epidemiology* (as opposed to either *dispositional* or *cognitive epidemiology* — for the latter, see, e.g., Deary, 2010). We begin by reviewing the predictive value of both IQ and the personality trait, neuroticism (N), measured for individuals and for geo-political units.

1.1. Individual and aggregate-level intelligence

Intelligence tests presumably measure individual differences in the brain's ability to efficiently process information (Jensen, 1998). Though controversial as a construct outside psychology, a massive literature shows that individual IQ scores predict real-world outcomes, from income levels and socioeconomic status (Strenze, 2007), to job and school performance (Kuncel, Ones, & Sackett, 2010; Schmidt & Hunter, 1998), to health and mortality (Batty, Deary, & Gottfredson, 2007; Deary, 2008; Deary, 2010; Gottfredson & Deary, 2004). For many outcome variables, IQ scores emerge as the single best (but not the only) predictor (see, e.g., Jensen, 1998).

* Corresponding author at: Cleveland State University, Department of Management, 2121 Euclid Avenue, Cleveland OH 44115, United States.
 E-mail address: b.pesta@csuohio.edu (B.J. Pesta).

In explaining links between IQ and epidemiology, Gottfredson (2004) argued that healthcare is a type of intelligence test (see also Gottfredson, 1997). Namely, health maintenance involves active participation in a series of tasks (e.g., learning health-related information), duties (e.g., dealing with health emergencies) and responsibilities (e.g., adhering to treatment). These behaviors require cognitive resources to manage effectively. Individuals (or groups of people) with high IQ would likely be in the best position to handle the complex spectrum of knowledge and behavior needed for good health.

Whether geographical units (versus individuals) differ in IQ has drawn increased attention from psychologists. In the aggregate, IQ scores have now been calculated for nations across the world (Lynn & Meisenberg, 2010; Lynn & Vanhanen, 2002), and for the 50 U.S. states (McDaniel, 2006). Both national and U.S. state IQs predict many of the things that individual IQ scores do, including socio-economic status (Pesta, McDaniel, & Bertsch, 2010), education (Lynn & Meisenberg, 2010), and crime (Pesta et al., 2010). Particularly relevant are recent studies showing links between aggregate IQ and epidemiologic outcomes (e.g., global state health: Pesta et al., 2010; life expectancy, mortality and fertility rates: Reeve, 2009; positive and negative health indicators: Reeve & Basalik, 2010).

1.2. Individual and aggregate-level neuroticism

Personality is the set of psychological traits or constructs that create consistency in how people think, act and feel (John, Robins, & Pervin, 2008). A highly regarded theoretical perspective on personality is the "Big Five" model (Costa & McCrae, 1992). The model assumes that five factors explain most of the variance in one's personality: neuroticism, extraversion, openness, agreeableness and conscientiousness. We focus here on just neuroticism, as it emerged as the only consistent *Big Five* predictor of epidemiologic outcomes (e.g., rates of heart disease or high blood pressure) and health-related behaviors (e.g., rates of smoking or exercise). Individuals scoring high on N tend to be anxious, stressed, and worry-prone, while those scoring low tend to be the opposite (Costa & McCrae, 1992).

Among individuals, N correlates with many health-related variables, including depression and anxiety disorders (Jyhla & Isometsa, 2006), mortality (Deary et al., 2008; Wiebe, Drew, & Croom, 2010), coping skill (John et al., 2008), death from cardiovascular disease (Shipley, Weiss, Der, Taylor, & Deary, 2007), and whether one smokes tobacco (Munaf, Zetteler, & Clark, 2007). Recent research also shows a strong relationship between N and *metabolic syndrome;* a chronic complex of health symptoms associated with increased heart disease and mortality (Phillips et al., 2010). To explain this relationship, Phillips et al. (2010) suggest that N "may be a marker of central nervous system (CNS) excitation, with higher levels leading to biological senescence, thus, increasing susceptibility to disease"(p. 193).

As with aggregate-level IQ, psychologists have recently focused on how personality traits vary across geographical units. Estimates now exist of the *Big Five* personality traits for each of the 50 U.S. states (Rentfrow, 2010; Rentfrow et al., 2008). State personality predicts many interesting aspects of American culture, including political preference and voting patterns (Rentfrow, Jost, Gosling, & Potter, 2009). Consistent with research on individuals, N seems to be the best predictor (among the Big Five traits) of health outcomes for the 50 U.S. states (as reviewed by Rentfrow et al., 2008).

Why do geo-political units differ meaningfully on psychological dimensions? One possibility is the attraction/similarity paradigm, where people are drawn to others who closely resemble them in characteristics like cultural background, personality, or shared demographics (Lydon, Jamieson, & Zanna, 1988). Both social (e.g., religious beliefs and customs) and genetic (e.g., IQ and personality, in part) factors characterize the settlers of a particular geographic area. Settler characteristics then become the basis for local beliefs and behaviors, which either attract or repel future residents from assimilating a community's culture. These specific characteristics likely still remain represented genetically and culturally in local populations in a non-random fashion (Rentfrow et al., 2008).

1.3. Explaining links between aggregate IQ/N and health

Arden, Gottfredson, and Miller (2009) proposed four possible explanations for links between individual-level IQ and health. We generalize their discussion here to include relationships between aggregate-level IQ, N and the health of populations:

(1) IQ/N and health could be influenced by common genetic factors.
(2) IQ/N and health could be influenced by common environmental factors.
(3) Health could influence IQ/N.
(4) IQ/N could influence health (Arden et al., 2009, p. 581).

Explanations (1) and (2) contrast genetic and environmental factors. In explanation (1), genes and genetic mutations affect health, IQ and N. This explanation is preferred by Arden et al. (2009), who argued for the existence of a general fitness factor, determined by genetics. The fitness factor subsumes IQ, N and health outcomes. Links between IQ/N and health are mediated by differences in lifestyle behaviors (e.g., smoking, exercising), which then lead to differences in disease rates across individuals or populations. In explanation (2), the relationship between IQ/N and health is caused by environmental variables. Examples include prenatal care, social stress, and pathogen loads.

The last two explanations differ on the direction of presumed causality. In explanation (3), health influences IQ/N, while the reverse holds in explanation (4). For the former, perhaps good health increases brain efficiency (as measured by IQ) and reduces stress (as measured by N); whereas disease decreases brain efficiency and increases stress. For explanation (4), high IQ/low N individuals might be more likely to engage in behaviors (e.g., exercise, eating healthy) conducive to good health. Though similar to explanation (1) in terms of what it predicts, explanation (4) does not necessarily implicate genetics. For example, high IQ might indirectly affect health by improving educational and career opportunities (Arden et al., 2009).

All four explanations probably contribute to the relationship between IQ/N and health (Arden et al., 2009). Consistent with this conclusion, most important socio-political variables (including health outcomes) are strongly inter-correlated at the aggregate level. For example, Pesta et al. (2010) identified

a robust general factor of state "well-being," comprised of the following sub-domains: intelligence, crime, education, income, health, and religious fundamentalism. Inter-dependence among state-level outcome variables seems to be the rule, rather than the exception (Pesta et al., 2010). Interestingly, the well-being nexus identified by Pesta et al. (2010) might largely reflect the general fitness factor hypothesized by Arden et al. (2009). The nexus may also result from the joint effects of all four explanations reviewed above. More research is needed in order to make inferences about the relative importance of each explanation for links between IQ/N and health.

Toward that end, we provide an examination of how IQ and N link to the behavioral antecedents of disease (e.g., smoking, exercising) and to disease itself. Then, we examine how these variables correlate with other important sub-domains (e.g., income, crime) of state well-being. Finally, we describe challenges for researchers interested in the causal mechanisms (i.e., the four explanations reviewed above) that best explain these relationships.

Studying IQ and N at the aggregate level allows researchers to capitalize on large, reliable data bases maintained by the U.S. federal government. One example is the Behavioral Risk Factor Surveillance System (BRFSS). The BRFSS tracks health and welfare across the 50 U.S. states. The system is updated annually, via population-representative surveys of residents in each state. Data exist on both the incidence of health behaviors (e.g., exercise; smoking) and chronic conditions (e.g., heart disease; high cholesterol) by U.S. geography.

We coded data comprising ten variables reported in a current BRFSS *Surveillance Summary* (BRFSS, 2010). We selected this specific summary because it is timely and includes many common, chronic health problems impacting the well-being of millions of people. Via regression, we first tested whether IQ/N predict chronic disease, and then whether these effects are attenuated by including health behaviors and state income levels in the model. Thereafter, we add other state-level variables illustrating a nexus of inter-correlated psychological, epidemiological, and environmental outcomes.

2. Method

2.1. State IQ and state N

State IQ estimates come from McDaniel (2006), and have a mean of 100.3 with a standard deviation of 2.70. State N estimates come from Rentfrow et al. (2008) and are reported as Z scores (mean = 0; $\sigma = 1$). All remaining variables were coded from an on-line, Behavioral Risk Factor Surveillance System Summary (2010). Unless noted, all BRFSS variables represent percentages for residents age 18 years or older in each state.

2.2. BRFSS variables

2.2.1. Health behaviors

We coded four BRFSS variables representing behaviors that epidemiologists traditionally study when predicting disease rates across populations. These included: (1) Activity and Exercise (created via factor analysis on BRFSS variables measuring light, moderate, and vigorous activity — the Appendix displays factor loadings and alpha reliabilities for

all scaled variables used in this study), (2) Smoking (every day or occasionally), (3) Alcohol Consumption (consumption of more than one [women] or two [men] alcoholic beverages per day), and (4) Healthy Eating (consuming at least five servings of fruits and/or vegetables per day). These four variables were highly correlated, so we also combined them into a single factor, *Health Behaviors*, for use in regression analyses.

2.2.2. Chronic disease

We coded the following measures of chronic disease or impairment (as a percentage of state residents) from the BRFSS summary: (1) Obesity (BMI > 30), (2) Diabetes, (3) High Blood Pressure, (4) High Cholesterol, (5) Coronary Heart Disease, (6) Stroke. We report data on these variables separately, and then together as scaled into a single factor (*Chronic Disease*) via factor analysis. Finally, we also created factor scores for state *Metabolic Syndrome* (using the first four of six variables representing the Chronic Disease factor) to see if recent results by Phillips et al. (2010) replicate at the U.S. state level.

2.2.3. Data analyses

We first report descriptives and simple correlations for all variables. Next, we report two regressions — one features Chronic Disease as the dependent variable, the other features Metabolic Syndrome. For each regression, IQ and N were entered in Step 1, and Health Behaviors was entered in Step 2. Lastly, we incorporate state income levels and additional variables (i.e., the sub-domains of well-being reported by Pesta et al., 2010) to see where Health Behaviors and Chronic Disease fit within the U.S. state well-being nexus.

3. Results

3.1. Descriptives and Pearson correlations

Table 1 shows rankings by U.S. state for IQ, N and the three BRFSS factor scores (Health Behaviors, Chronic Disease, and Metabolic Syndrome). Note for example that West Virginia has the highest incidence of Chronic Disease. It is also the most neurotic state in the U.S. Conversely, Utah is the least neurotic state; whereas, Massachusetts has the highest IQ, and Vermont ranks first in Health Behaviors.

Table 2 shows means, standard deviations, and zero-order correlations for IQ, N, all BRFSS variables, and the three BRFSS factor scores (with $N = 50$, a correlation of .28 is significant at $p < .05$, one tailed). In the table, state IQ and N are essentially uncorrelated ($r = -.08$). However, IQ significantly predicts two of the four behavioral variables (activity/exercise, and smoking), and correlates .45 with the Health Behaviors factor score. Similarly, N significantly predicts the same two behavioral variables that IQ does, and N correlates −.40 with Health Behaviors. Both IQ and N, however, failed to correlate with either alcohol consumption or healthy eating.

State IQ correlated moderately with four of the six disease variables (IQ predicted neither high cholesterol nor heart disease). IQ also correlated −.51 and −.53 with Chronic Disease and Metabolic Syndrome, respectively. Relative to state IQ, the correlations between N and disease were stronger and more consistent. For example, state N correlated significantly with all six disease variables. It also correlated .59 and .62 with the Chronic Disease and Metabolic Syndrome, respectively.

Table 1
State ranks and mean values for IQ, N, and the BRFSS factor score composites.

State	IQ rank/score	N rank/score	Health behaviors rank/score	Chronic disease rank/score	Metabolic syndrome rank/score
Alabama	45.5/95.7	30/−0.26	44/−1.24	4/1.74	4/1.84
Alaska	36/99.0	46/−1.2	11/0.77	47/−1.18	45/−0.88
Arizona	43/97.4	44/−1.09	16/0.65	24/−0.17	44/−0.82
Arkansas	42/97.5	10/1.01	41/−0.84	8/1.06	7/1.17
California	48/95.5	36/−0.53	3/1.21	40/−0.82	42/−0.80
Colorado	20/101.6	49/−1.97	10/0.82	49/−2.08	49/−2.16
Connecticut	9/103.1	15/0.54	4/1.18	46/−0.96	39/−0.63
Delaware	28/100.4	20/0.21	24/0.18	17/0.47	11/0.65
Florida	38.5/98.4	35/−0.5	22/0.26	19/0.31	22/0.09
Georgia	40/98.0	32/−0.39	30/−0.03	9/0.96	9/0.91
Hawaii	47/95.6	39/−0.74	2/1.36	30/−0.30	21/0.13
Idaho	22/101.4	31/−0.36	18/0.36	32/−0.38	35/−0.55
Illinois	31/99.9	19/0.21	26/0.11	22/0.07	24/0.04
Indiana	19/101.7	13/0.88	38/−0.64	18/0.44	19/0.24
Iowa	8/103.2	22/0.15	35/−0.32	27/−0.24	39/−0.22
Kansas	12/102.8	33/−0.44	37/−0.51	29/−0.27	30/−0.26
Kentucky	34/99.4	7/1.17	50/−2.09	7/1.24	10/0.82
Louisiana	49/95.3	8/1.14	46/−1.37	6/1.25	5/1.39
Maine	6.5/103.4	12/0.9	7/0.95	20/0.15	20/0.21
Maryland	32/99.7	17/0.45	19/0.33	23/−0.03	16/0.34
Massachusetts	1/104.3	11/0.98	8/0.94	41/−0.82	38/−0.61
Michigan	27/100.5	26/−0.09	28/−0.01	15/0.57	12/0.53
Minnesota	5/103.7	40/−0.8	34/−0.22	48/−1.68	48/−1.89
Mississippi	50/94.2	4/1.5	49/−1.82	3/1.85	1/2.17
Missouri	25/101.0	25/−0.09	42/−0.93	12/0.70	14/0.48
Montana	6.5/103.4	38/−0.71	15/0.69	42/−0.85	46/−0.92
Nebraska	15/102.3	43/−1.0	25/0.12	34/−0.46	33/−0.49
Nevada	44/96.5	41/−0.83	29/−0.01	26/−0.24	27/−0.19
New Hampshire	2/104.2	14/0.7	9/0.86	35/−0.49	34/−0.50
New Jersey	12/102.8	5/1.47	20/0.30	21/0.07	23/0.06
New Mexico	45.5/95.7	29/−0.2	33/−0.19	33/−0.46	36/−0.58
New York	26/100.7	3/1.55	21/0.27	31/−0.32	26/−0.12
North Carolina	29/100.2	24/−0.06	43/−0.94	14/0.60	13/0.51
North Dakota	3.5/103.8	42/−0.84	32/−0.12	38/−0.68	37/−0.59
Ohio	18/101.8	9/1.1	36/−0.47	11/0.75	15/0.47
Oklahoma	35/99.3	27/−0.15	48/−1.77	5/1.48	6/1.38
Oregon	23/101.2	47/−1.27	5/1.11	37/−0.58	31/−0.35
Pennsylvania	21/101.5	6/1.22	27/0.10	13/0.61	18/0.29
Rhode Island	33/99.5	2/1.61	13/0.73	28/−0.26	25/−0.02
South Carolina	38.5/98.4	16/0.53	39/−0.71	10/0.81	8/0.91
South Dakota	12/102.8	48/−1.68	40/−0.71	36/−0.57	41/−0.74
Tennessee	41/97.7	23/0.11	45/−1.35	2/1.94	2/1.95
Texas	30/100.0	28/−0.17	31/−0.10	16/0.56	17/0.30
Utah	24/101.1	50/−2.52	12/0.75	50/−2.11	50/−2.32
Vermont	3.5/103.8	18/0.43	1/1.57	45/−0.94	47/−1.00
Virginia	16.5/101.9	21/0.18	17/0.56	25/−0.23	28/−0.20
Washington	16.5/101.9	45/−1.1	6/0.99	39/−0.76	40/−0.70
West Virginia	37/98.7	1/2.36	47/−1.71	1/1.99	3/1.91
Wisconsin	10/102.9	34/−0.45	14/0.72	43/−0.86	32/−0.48
Wyoming	14/102.4	37/−0.59	23/0.21	44/−0.86	43/−0.81

Note. IQ (intelligence) has a mean of 100.3 and a standard deviation of 2.7. N (neuroticism) is a Z score.
The remaining variables are factor scores.

An unexpected finding in Table 2 is the relationship between alcohol consumption, Health Behaviors and Chronic Disease. At the state level, drinking alcohol correlates positively with exercising and eating fruits and vegetables; whereas it correlates negatively with rates of smoking and many of the tabled chronic diseases. These data are consistent with a growing but mixed literature showing that alcohol consumption correlates inversely with chronic disease rates (see e.g., Holahan et al., 2010; see also, Arden et al., 2009, for mixed results of alcohol consumption on a battery of health-related variables).

In sum, the zero-order correlations in Table 2 show IQ and N to be relatively consistent predictors of both behaviors associated with chronic diseases, and the diseases themselves. Next we test whether these relationships are attenuated by considering behaviors like smoking and exercising.

3.2. Multiple regressions

To avoid multi-colinearity, we used factor scores for the BRFSS behavioral and disease variables. High scores on the Health Behaviors factor correspond to higher rates of exercising,

Table 2
Descriptive statistics and correlation matrix for state IQ, N, and all BRFSS variables.

	M	SD	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. I 1 IQ	100.3	2.7	−.08	.51	−.29	.12	.19	−.37	−.57	−.52	−.10	−.20	−.49	.45	−.51	−.53
2 N	0.0	1.0	–	−.50	.39	−.07	−.04	.28	.57	.62	.53	.61	.40	−.40	.59	.62
3 Activity and Exercise[1]	51.6	4.4		–	−.59	.54	.52	−.68	−.83	−.80	−.28	−.73	−.76	.92	−.85	−.82
4 Smoking	20.1	3.2			–	−.37	−.57	.68	.60	.65	.53	.73	.73	−.78	.74	.68
5 Alcohol Consumption	5.1	1.3				–	.52	−.49	−.37	−.23	.01	−.36	−.37	.69	−.36	−.28
6 Healthy Eating	23.7	3.5					–	−.65	−.28	−.35	−.23	−.43	−.43	.74	−.42	−.37
7 Obesity	26.9	2.9						–	.67	.71	.46	.59	.69	−.78	.77	.75
8 Diabetes	8.2	1.5							–	.87	.48	.75	.81	−.77	.94	.91
9 High Blood Pressure	28.5	3.0								–	.57	.73	.80	−.74	.95	.99
10 High Cholesterol	37.8	2.2									–	.60	.49	−.36	.61	.60
11 Heart Disease	6.5	1.3										–	.77	−.76	.84	.76
12 Stroke	2.6	0.5											–	−.78	.91	.83
13 Health Behaviors[2]	0.0	1.0												–	−.83	−.77
14 Chronic Disease[2]	0.0	1.0													–	.98
15 Metabolic Syndrome[2]	0.0	1.0														–

Note: N (neuroticism) is a Z score. The remaining variables (except as noted) represent % means for the 50 U.S. States.
[1] The reported mean is the average for three variables comprising this factor — see the Appendix A.
[2] Created via Maximum likelihood factor analysis — see the Appendix A.

eating fruits/vegetables, and drinking alcohol, but lower rates of smoking. For both the Chronic Disease and Metabolic Syndrome factors, high scores indicate higher disease rates (as a percentage of residents) across states.

Table 3 shows results of hierarchical regressions predicting Chronic Disease and Metabolic Syndrome from IQ, N (entered at Step 1) and Health Behaviors (entered at Step 2). At Step 1, the linear combination of IQ and N alone explained 57% and 61% of the variance in Chronic Disease and Metabolic Syndrome, respectively. Both IQ and N remained significant (but attenuated) predictors of disease, after entering Health Behaviors at Step 2. Not surprisingly, Health Behaviors itself explained large amounts of variance (over IQ and N) in both Chronic Disease and Metabolic Syndrome. Note that the variance explained at Step 2 is unusually large for social science research. Fully 80% of the variance in Chronic Disease (77% in Metabolic Syndrome) was explained by the combination of IQ, N and Health Behaviors. The size of the effects here, though, could exemplify the "high resolution" that aggregate-level data offer, relative to studies that use individuals (see Arden et al., 2009, p. 582).

Any number of third variables could be included in the regression models above. Perhaps most obvious, socioeconomic differences across states may largely explain why IQ and N emerge as strong predictors of disease. For various reasons, poorer states might score lower on IQ and higher on N. To test this hypothesis, we included a composite measure of state income from Pesta et al. (2010). The state income measure was a factor derived from U.S. census data. Variables included: income per capita, disposable income per capita, the percentage of families living in poverty, and the percentage of individuals living in poverty (Pesta et al., 2010). We re-conducted the regression analysis reported above (predicting Chronic Disease in Table 3) and included state income levels at Step 3. However, IQ (Beta = −.18), N (Beta = .35) and Health Behaviors (Beta = −.53) all remained significant as predictors of chronic disease, even after controlling for state income (Beta = −.12, ns).

3.3. nexus of state outcome variables

Adding additional third variables to the regression model seems arbitrary, given the strong correlations between most state-level measures. Also, showing that a variable predicts over and above another variable does not necessarily mean that the chosen variable is the best explanation for the data. Differences in construct validity, base rates, or variance across variables would affect conclusions reached via regression, independent of a variable's true effect on health. Instead, we believe the most compelling aspect of these results is that any given state-level outcome is consistently and non-trivially correlated with nearly every other outcome.

Table 3
Predicting chronic disease and metabolic syndrome from IQ, N, and health behaviors.

Variable	Chronic disease				Metabolic syndrome			
	β	B	SE B		β	B	SE B	
Step 1								
Intelligence (IQ)	−.47	−.168	.035		−.48	−.174	.033	
Neuroticism (N)	.55	.534	.093		.58	.559	.089	
R/R^2				.75/.57				.78/.61
Step 2								
Intelligence (IQ)	−.22	−.079	.027		−.28	−.101	.029	
Neuroticism (N)	.33	.322	.070		.40	.385	.076	
Health behaviors	−.60	−.637	.086		−.49	−.523	.093	
R/R^2				.90/.80				.88/.77

Note: IQ (intelligence) is a scaled score with a mean of 100 and a standard deviation of 2.7. N (neuroticism) is a Z score, and all other variables are factor scores.

Table 4
A nexus of inter-correlated variables potentially representing a general fitness factor.

	1	2	3	4	5	6	7	8	9
1 IQ	–	−.08	.45	−.51	−.53	−.76	.41	−.55	.57
2 N		–	−.40	.59	.62	.01	−.30	.05	−.12
3 Health behaviors			–	−.83	−.77	−.47	.74	−.78	.66
4 Chronic disease				–	.98	.49	−.72	.65	−.62
5 Metabolic syndrome					–	.48	−.67	.61	−.55
6 Crime[1]						–	−.26	.51	−.42
7 Education[1]							–	−.62	.66
8 Religiosity[1]								–	−.72
9 Income[1]									–

[1] These variables are factor scores borrowed from Pesta et al. (2010).

To illustrate, Table 4 shows correlations between the variables presented here and various sub-domains of well-being reported by Pesta et al. (2010). Of the 36 correlations presented in the table, eight (22%) have values of $r = .70$ or higher; fifteen (42%) have values between $r = .50$ and $r = .69$; and eight (22%) have values between $r = .30$ to $r = .49$. Only 5 (14%) correlations are non-significant (four of these occur for N predicting other variables in the table).

The correlations in Table 4 could reflect the existence of the general fitness factor, as proposed by Arden et al. (2009). However, they are also consistent with the interplay of all four explanations for the IQ/N and health link, reviewed above. Isolating causality for highly-correlated, aggregate-level variables (that are not experimentally manipulated) is a daunting task. What is clear, though, is that a nexus of inter-correlated variables exists, and it reliably measures psychological, environmental and epidemiologic differences across the 50 U.S. states.

4. Discussion

The present data show that psychological variables uniquely predict differences in chronic disease rates across the 50 U.S. states. Moderate to strong relationships exist between IQ/N and a variety of chronic health problems, together with the behavioral antecedents of these problems. These relationships persist, even after controlling for income, and for many behaviors (e.g., exercising, smoking) epidemiologists typically study as the causes of disease.

4.1. Causality

We reviewed four possible explanations for links between IQ/N and health. Arden et al. (2009) argued that each explanation partly contributes to these relationships. However, measuring the relative importance of each explanation is complicated by the correlational nature of the data. One strategy is to seek replication across different units of analysis. Consistent patterns across individuals, states, and nations would strengthen inferences about a variable's role in affecting health.

A second strategy would be to isolate key third variables derived from testable theories. For example, Arden et al. argued for the existence of a genetic fitness factor that influences IQ/N and health. A reliable measure of "mutation load" for the 50 U.S. states would offer an informative (but non-conclusive) test of this hypothesis (see Arden et al., 2009). Unfortunately, we know of no such measure. However, aggregate-level genetic data do exist, using US schools as

the unit of analysis. Beaver and Wright (2011) examined data for 132 middle and high schools across the USA. They showed that genetic variation (i.e., allelic distributions of dopaminergic polymorphisms) predicted verbal IQ scores for different schools, even after controlling for race.

A second example of a key third variable is parasite prevalence (a measure of biological stress caused by infectious disease; see, e.g., Eppig, Fincher, & Thornhill, 2011). Eppig et al. (2011) showed that this variable uniquely predicted U.S. state IQ, even after controlling for income and educational differences across the states. As a final example, income inequality (the wealth difference between the richest and poorest members of a population) seems to be the third-variable of choice for economists (see, e.g., Diener & Oishi, 2000).

Nonetheless, we caution against making strong causal inferences just because a specific variable "won" by explaining the most unique variance in a regression model (for a discussion, see Gottfredson, 2009). For example, education and IQ are strongly correlated (Kuncel et al., 2010). Often compelling theoretical reasons exist for controlling education when testing the effects of IQ (or vice versa) on some outcome. But, if education and IQ are co-causal, the resulting attenuation of IQ (by partialing out education) would give other independent variables an unfair advantage in potential to explain unique variance (thereby leading researchers to faulty inferences about presumed causality). Methods beyond regression are needed that triangulate possible causal mechanisms for the relationships observed here.

4.2. Other limitations

Beyond issues with correlation and causation, the present study has other limitations. First, the substantial positive manifold of the variables makes it hard to identify any single measure's contribution to health differences across the 50 U.S. states. Second, the potential for committing the ecological fallacy (Robinson, 1950) exists when interpreting these data. The ecological fallacy sometimes occurs when making inferences about people (or groups) from data derived from groups (or people).

For example, it does not follow that all residents of Massachusetts (a state with a high IQ) are smarter, less neurotic and healthier, relative to all residents of Mississippi (a state with a low IQ). Nor does it follow that group-level effects necessarily apply to individuals comprising the groups. Finally, the explanations that link IQ/N and health among individuals may be different from those that link these variables in aggregate-level

data. Crossing levels of analysis — from states to individuals within a state — could be an invalid extrapolation.

5. Conclusion

Epidemiologists should regularly employee psychological variables when studying disease patterns. The present data show that psychological variables are non-trivially correlated with health differences across the 50 U.S states, and with behaviors epidemiologists study as disease antecedents. Including psychological variables could offer epidemiologists increased leverage when predicting, interpreting and explaining differences in disease rates across populations.

Coding variables by geo-political units (versus individuals) allows researchers to capitalize on the power of aggregation (see, e.g., Lubinski & Humphreys, 1996). Specifically, IQ, N and Health Behaviors jointly explained 80% of the variance in Chronic Disease (77% for Metabolic Syndrome). The amount of variance explained by these variables is unusually large for social science research. These effects could reflect the potential (i.e., "high resolution," Arden et al., 2009) that comes by aggregating data. Continued use of aggregate-level data could help researchers design powerful tests of competing theories, thereby identifying the relative importance of proposed explanations for these relationships.

Appendix A

Maximum likelihood factor analyses, percentage of variance explained, and alpha reliabilities for selected BRFSS variables.

Factor	BRFSS item factor loading	% Variance explained	Alpha reliability
Activity and exercise	Light activity[1] .92 moderate activity[2] .98 vigorous activity[3] .95	90%	.94
Health behaviors	Activity and exercise .77 smoking −.73 alcohol consumption .64 healthy eating .74	52%	.88
Chronic disease	Obesity .75 diabetes .92 high blood pressure .93 high cholesterol .59 heart disease .82 stroke .89	68%	.86
Metabolic syndrome	Obesity .73 diabetes .89 high blood pressure .97 high cholesterol .58	65%	.85

Note. Factors appear in capital letters; individual BRFSS variables appear in lower case.

[1] State residents (%) participating in light activity or exercise (e.g., calisthenics) during the preceding month.

[2] State residents (%) participating in moderate activity or exercise (e.g., brisk walking, bicycling, or anything else that causes a small increase in breathing and heart rate) on at least 5 days per week for at least 30 min each day.

[3] State residents (%) participating in vigorous activity or exercise (e.g., running, aerobics, or anything else that causes a large increase in breathing and heart rate) on 3 or more days per week for at least 20 min each day.

References

Arden, R., Gottfredson, L., & Miller, G. (2009). Does a fitness factor contribute to the association between intelligence and health outcomes? Evidence from medical abnormality counts from 3654 US veterans. *Intelligence*, *37*, 581–591.

Batty, G., Deary, I., & Gottfredson, L. (2007). Premorbid (early life) IQ and later mortality risk: A systematic review. *Annals of Epidemiology*, *17*, 278–288.

Beaver, K., & Wright, J. (2011). School-level genetic variation predicts school-level verbal IQ scores: Results from a sample of American middle and high schools. *Intelligence*, *39*, 193–197.

Behavioral Risk Factor Surveillance System (2010). Retrieved September 1, 2011 from. www.cdc.gov/brfss/

Behavioral Risk Factor Surveillance System Summary (2010). Retrieved September 1, 2011, from. http://www.cdc.gov/mmwr/preview/mmwrhtml/ss5901a1.htm

Costa, P., & McCrae, R. (1992). Normal personality assessment in clinical practice: The NEO Personality Inventory. *Psychological Assessment*, *4*, 5–13.

Deary, I. (2008). Why do intelligent people live longer? *Nature*, *456*, 175–176.

Deary, I. (2010). Cognitive epidemiology: Its rise, its current issues, and its challenges. *Personality and Individual Differences*, *49*, 337–343.

Deary, I. J., Batty, G. D., Pattie, A., & Gale, C. R. (2008). More intelligent, more dependable children live longer: a 55-year longitudinal sudy of a representaitive sample of the Scottish nation. *Psychological Science*, *19*, 874–880.

Diener, E., & Oishi, S. (2000). Money and happiness: Income and subjective well-being across nations. In E. M. Suh (Ed.), *Culture and subjective well-being* (pp. 185–205). Boston: MIT Press.

Eppig, C., Fincher, C., & Thornhill, R. (2011). Parasite prevalence and the distribution of intelligence among the states of the USA. *Intelligence*, *39*, 155–160.

Gottfredson, L. S. (1997). Why g matters: The complexity of everyday life. *Intelligence*, *24*, 79–132.

Gottfredson, L. S. (2004). Intelligence: Is it the epidemiologists' elusive "fundamental cause" of social class inequalities in health? *Journal of Personality and Social Psychology*, *86*, 174–199.

Gottfredson, L.S. (2009). Pursuing patterns, puzzles and paradoxes. Unpublished manuscript.

Gottfredson, L. S., & Deary, I. J. (2004). Intelligence predicts health and longevity, but why? *Current Directions in Psychological Science*, *13*, 1–4.

Holahan, C., Schutte, K., Brennan, P., Holahan, C., Moos, B., & Moos, R. (2010). Late-life alcohol consumption and 20-year mortality. *Alcoholism, Clinical and Experimental Research*, doi:10.1111/j.1530-0277.2010.01286.x.

Jensen, A. (1998). *The g factor: The science of mental ability.* Santa Barbara, CA: Praeger Publishers.

John, O. P., Robins, R. W., & Pervin, L. A. (2008). *Handbook of Personality, 3rd edition: Theory and research.* New York, NY: Guilford Press.

Jyhla, P., & Isometsa, E. (2006). The relationship of neuroticism and extraversion to symptoms of anxiety and depression in the general population. *Depression and Anxiety*, *23*, 281–290.

Kuncel, N. R., Ones, D. S., & Sackett, P. R. (2010). Individual differences as predictors of work, educational, and broad life outcomes. *Personality and Individual Differences*, *49*, 331–336.

Lubinski, D., & Humphreys, L. G. (1996). Seeing the forest from the trees: When predicting the behavior or status of groups, correlate means. *Psychology, Public Policy, and Law*, *2*, 363–376.

Lydon, J. E., Jamieson, D. W., & Zanna, M. P. (1988). Interpersonal similarity and the social and intellectual dimensions of first impressions. *Social Cognition*, *6*, 269–286.

Lynn, R., & Meisenberg, G. (2010). National IQs validated for 108 nations. *Intelligence*, *38*, 353–360.

Lynn, R., & Vanhanen, T. (2002). *IQ and the Wealth of Nations.* Santa Barbara, CA: Praeger\Publishers.

McDaniel, M. A. (2006). Estimating state IQ: Measurement challenges and preliminary correlates. *Intelligence*, *34*, 607–619.

Munaf, M. R., Zetteler, J. I., & Clark, T. G. (2007). Personality and smoking status: A meta-analysis. *Nicotine & Tobacco Research*, *9*, 405–413.

Pesta, B. J., McDaniel, M. A., & Bertsch, S. (2010). Toward an index of well-being for the fifty U.S. states. *Intelligence*, *38*, 160–168.

Phillips, A. C., Batty, G. D., Weiss, A., Deary, I., Gale, C. R., Thomas, G. N., & Carroll, D. (2010). Neuroticism, cognitive ability, and the metabolic syndrome: The Vietnam Experience Study. *Journal of Psychosomatic Research*, *69*(2), 193–201.

Reeve, C. L. (2009). Expanding the g nexus: Further evidence regarding the relationship among national IQ, religiosity, and national health outcomes. *Intelligence*, *37*, 495–505.

Reeve, C. L., & Basalik, D. (2010). Average state IQ, state wealth and racial composition as predictors of state health statistics: Partial support for 'g' as a fundamental cause of health disparities. *Intelligence*, *38*, 282–289.

Rentfrow, P. (2010). Statewide differences in personality: Toward a psychological geography of the United States. *American Psychologist*, *65*, 548–558.

Rentfrow, P., Gosling, S., & Potter, J. (2008). A theory of the emergence, persistence, and expression of geographic variation in psychological characteristics. *Perspectives on Psychological Science*, *3*, 339–369.

Rentfrow, P., Jost, J., Gosling, S., & Potter, J. (2009). Statewide differences in personality predict voting patterns in 1996–2004 U.S. presidential elections. In J. T. Jost, A. C. Kay, & H. Thorisdottir (Eds.), *Social and psychological bases of ideology and system justification* (pp. 314–347). Oxford University Press.

Robinson, W. S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review*, *3*, 351–357.

Schmidt, F., & Hunter, J. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, *124*, 262–275.

Shipley, B. A., Weiss, A., Der, G., Taylor, M. D., & Deary, I. J. (2007). Neuroticism, extraversion, and mortality in the UK health and lifestyle survey: A 21-year prospective cohort study. *Psychosomatic Medicine*, *69*, 923–931.

Strenze, T. (2007). Intelligence and socioeconomic success: A meta-analytic review of longitudinal research. *Intelligence*, *35*, 401–426.

Wiebe, D. J., Drew, L. M., & Croom, A. (2010). Personality, health and illness. In D. French, K. Vedhara, A. Kaptein, & J. Weinman (Eds.), *Psychology* (pp. 294–302). New York: Wiley & Sons.