Intelligence 37 (2009) 581–591

Contents lists available at ScienceDirect

Intelligence

Does a fitness factor contribute to the association between intelligence and health outcomes? Evidence from medical abnormality counts among 3654 US Veterans

Rosalind Arden [a,*], Linda S. Gottfredson [b], Geoffrey Miller [c]

[a] Social, Genetic, Developmental and Psychiatry Centre, Institute of Psychiatry, King's College London, London SE5 8AF, United Kingdom
[b] School of Education, University of Delaware, Newark, DE 19716, USA
[c] Psychology Department, Logan Hall 160, University of New Mexico, MSC03 2220 Albuquerque, NM 87131-1161, USA

ARTICLE INFO

Available online 28 April 2009

Keywords:
Fitness
Intelligence
Cognitive epidemiology
Health
Mutation load

ABSTRACT

We suggest that an over-arching 'fitness factor' (an index of general genetic quality that predicts survival and reproductive success) partially explains the observed associations between health outcomes and intelligence. As a proof of concept, we tested this idea in a sample of 3654 US Vietnam veterans aged 31–49 who completed five cognitive tests (from which we extracted a *g* factor), a detailed medical examination, and self-reports concerning lifestyle health risks (such as smoking and drinking). As indices of physical health, we aggregated 'abnormality counts' of physician-assessed neurological, morphological, and physiological abnormalities in eight categories: cranial nerves, motor nerves, peripheral sensory nerves, reflexes, head, body, skin condition, and urine tests. Since each abnormality was rare, the abnormality counts showed highly skewed, Poisson-like distributions. The correlation matrix amongst these eight abnormality counts formed only a weak positive manifold and thus yielded only a weak common factor. However, Poisson regressions showed that intelligence was a significant positive predictor of six of the eight abnormality counts, even controlling for diverse lifestyle covariates (age, obesity, combat and toxin exposure owing to service in Vietnam, and use of tobacco, alcohol, marijuana, and hard drugs). These results give preliminary support for the notion of a superordinate fitness factor above intelligence and physical health, which could be further investigated with direct genetic assessments of mutation load across individuals.

1. Introduction

A formerly steady drip of individual reports has become a regular flow of empirical studies showing positive correlations between higher intelligence and better health outcomes (Batty, Deary, & Gottfredson, 2007; Deary & Der, 2005; Gottfredson & Deary, 2004; Hart et al., 2005; Kuh, Richards, Hardy, Butterworth, & Wadsworth, 2004; Martin & Kubzansky, 2005; van Oort, van Lenthe, & Mackenbach, 2005). When a relationship—such as that between intelligence and health—becomes well substantiated, the correlation's existence often seems obvious (in retrospect), but the reasons for the correlation are often left unexamined. There are at least four pathways that could contribute to the observed intelligence–health correlations, but so far cognitive epidemiologists have only given serious attention to the fourth.

1. Intelligence and health could both be influenced by common genetic factors.
2. Intelligence and health could both be influenced by common environmental factors.
3. Health could influence intelligence.
4. Intelligence could influence health.

Let's consider each pathway through some specific examples. 1) Common genetic factors could influence both intelligence and health. These include genes or mutations that affect multiple traits. We focus on this possibility in this paper.

* Corresponding author.
E-mail address: r.arden@iop.kcl.ac.uk (R. Arden).

0160-2896/$ – see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.intell.2009.03.008

2) Environmental factors could influence both intelligence and health. These factors might include prenatal infection, malnutrition, high pathogen load, environmental toxins, or social stress (Fuhrer et al., 2002; Marmot, 2003; Singh-Manoux, Ferrie, Lynch, & Marmot, 2005). 3) General good health may boost or protect intelligence and, conversely, specific illnesses (or their treatment side-effects) such as influenza, diabetes or cancer may harm it. Illness, or treatments, may affect intelligence by affecting glucose availability, neurotransmitter balances, patterns of gene expression in neural tissue, or other aspects of neurophysiology. 4) Intelligence could influence health, especially through choice of lifestyle (Batty et al., 2007; Gottfredson & Deary, 2004; Kumari, Seeman, & Marmot, 2004; Starr et al., 2004; Whalley, Fox, Deary, & Starr, 2005). Higher-intelligence people may learn earlier and faster about the health implications of habits such as smoking, drinking, using drugs, over-eating, and being sedentary (Gottfredson, 2006). Intelligence can also influence health indirectly through improving education and career outcomes. Blue-collar jobs may entail higher accident risks, more exposure to toxins, pathogens, and mutagens, more stress, fewer holidays, worse food, and worse health insurance, all of which can undermine health.

These four pathways probably all contribute somewhat to the positive correlations between intelligence and health, but their relative importance remains unknown. Indeed, findings that intelligence and health are correlated are often taken as prima facie evidence for the second or fourth pathway. Our concern here is not to argue that these two pathways are unimportant, but that the first pathway (common genetic factors) may warrant much closer attention.

We suggest that a latent genetic 'fitness factor' influences diverse traits throughout the entire human organism including, but not limited to, general health and general intelligence. (By 'fitness' we mean a person's statistical propensity to survive and reproduce successfully in ancestrally normal environments). This hypothesized fitness factor (Houle, 2000; Miller, 2000; Prokosch, Yeo, & Miller, 2005) (see Fig. 1 below), as manifest throughout the entire phenotype, is broader in scope than physical health; it includes traits such as mental health, sexual attractiveness, social status, parental competence, and cognitive abilities. Indeed, the 'fitness factor' would be hierarchically dominant to *g* and, would partly explain the existence, stability, heritability, and predictive validity of *g*. We posit that this fitness factor reflects an individual's general genetic quality (low mutation load, low genetic inbreeding). (What evolutionary biologists call 'phenotypic condition' is affected by both genetic quality and environmental variables. Here we focus theoretically on genetic fitness but empirically on phenotypic condition.)

Fig. 1 illustrates our provisional model of how the fitness factor might influence both intelligence and health. Positive loadings from general intelligence and physical health on the hypothesized fitness factor could explain a portion of the positive correlations between intelligence and health observed in cognitive epidemiology research, without depending on pathways 2, 3, or 4.

It is well known, among readers of this journal, that any cognitive ability correlates moderately with any other cognitive ability. This tendency of abilities to go together gives rise to Spearman's *g*, the formalized factor accounting for around half the total variance in a diverse array of tests (Jensen, 1998). Is this striking matrix of positive correlations from which *g* emerges unique, or is it part of a larger phenotype-wide manifold? Evolutionary biology offers two good reasons to expect a positive manifold throughout the phenotype that can be represented by a general fitness factor. These reasons concern the nature of mate choice, and the nature of mutations.

First consider mate choice — the selective choice of sexual partners found in most animal species. Finding and choosing a mate choice has higher costs in time, energy, and risk than mating randomly, so mate choice must offset these costs by offering some compensating fitness benefits such as better genes or resources for one's offspring. Consistent with Darwin's sexual selection theory, decades of evidence (Andersson & Simmons, 2006; Kokko, Jennions, & Brooks, 2006) suggest that mate choice can bring both genetic benefits (such as fewer mutations that would reduce the survival and reproductive prospects of one's offspring) and resource benefits (such as better food, territory, or protection).

In particular, mate choice for genetic benefits ('good genes mate choice') makes more sense if there is some general dimension of genetic quality similar to our hypothesized fitness factor. Of course, for such 'good genes sexual selection' to work, animals require no conscious understanding of why they evolved mate preferences for certain traits correlated with genetic quality. Also, the fact that many species show good genes mate preferences for multiple traits suggests that no

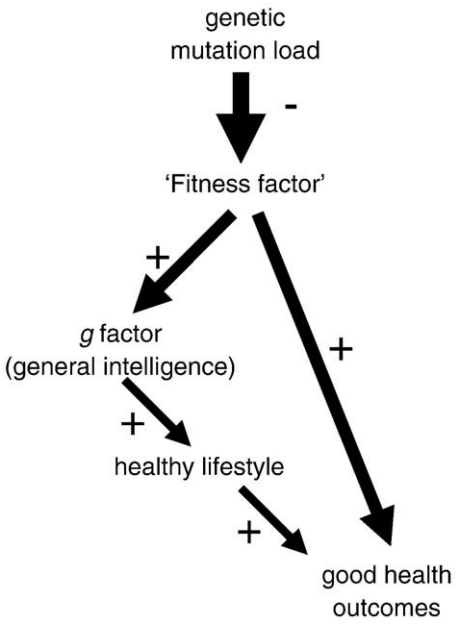

Fig. 1. A model of the fitness factor: how it might influence both general intelligence and health outcomes. Note that it may create positive correlations between intelligence and good health through two pathways: (1) intelligence influences lifestyle and health behaviors (exercise, diet, smoking, drinking, drug use, hazard exposure), which influence health outcomes; (2) intelligence and health outcomes are both positively and directly loaded onto the fitness factor, which reflects underlying genetic quality (low mutation load).

single trait has a very high correlation with overall genetic quality.

The second reason to expect a general fitness factor concerns the nature of mutations. As a species, we are perpetually subject to the opposing forces of mutation and selection. We all carry thousands of mutations but their number and harmfulness vary among us (Crow, 2000; Houle & Kondrashov, 2002; Lynch et al., 1999). If most of our genes are at least mildly pleiotropic, then such mutations will impair multiple traits to some degree. We propose that individual-level variation in mutation load likely causes a population-level, weakly visible, phenotype-wide correlation matrix, which may be seen under high enough resolution (small associations are only detectable through large sample sizes). If there is such a matrix, then traits subject to natural and sexual selection (fitness-related traits) should all be positively correlated. Since *g* is likely to have a huge genetic footprint (we know this empirically since no major studies have found genes accounting for more than 1% of the variance in *g*), our hypothesis suggests that *g* should correlate negatively with most impairments across most other traits.

In this report we use physician-diagnosed medical measures (scored dichotomously as normal or abnormal) of diverse physical traits as one might use 'items' in an IQ-type test. We formed 'tests' (or scales) of these abnormalities and examined whether these scales correlated positively with one another and negatively with *g*.

2. Methods

2.1. Participants

This study included all 3654 White non-Hispanic men among the 4462 US Vietnam-era Army veterans who were recruited, as part of the larger Vietnam Experience Study (VES), to undergo a comprehensive four-day physical and psychological examination. The men had served in Vietnam, South Korea, Germany, or the USA. The men were aged 31–49 (mean 38) at the time of these examinations in 1985–6. Their average age of induction into the army was 20 (range 16–33). We included only men whose self-reported ethnicity was White and non-Hispanic because of our evolutionary hypothesis. Selection on traits varies between populations with different geographic ancestry (think of the selection on melanin in equatorial regions, or the selection on shorter stature in Polar regions).

Careful sampling design ensured that men in this study were highly representative of the US Army at the time of their military service, 1965–1971. Army recruits during that period were probably more representative of the general population of young men than is typical of later military samples, because there was compulsory military service. Except for test-norming samples, military samples are generally the most representative samples available, and they are widely used in cognitive epidemiology. Yet, it should be noted that even the big military samples are always somewhat restricted in range. All Army recruits must pass a physical examination, so none of the VES participants had serious physical disabilities or illnesses when they were inducted. Federal law also bars the US military from enlisting individuals below the 10th percentile in cognitive ability, which means that men from the lower (left) tail of the cognitive ability distribution are under-sampled. Normally, the military services induct no one below the 16th percentile of mental ability, but Secretary of Defense Robert McNamara initiated Project 100,000 in 1966, which required all branches of the military to induct men from the 10th–15th percentiles. They were also required to meet larger quotas for men of substantially below-average intelligence; during 1967–1971, 25% of Army accessions were to come from the 10th to 30th percentiles of ability (Laurence & Ramsberger, 1991, p. 29). After induction into the Army, lower-intelligence men may have been more likely to die in combat, or, after military service, to end up in prison, hospital, or otherwise unable to participate in the VES study. Nonetheless, the sample is more representative at the low extreme than is often the case in intelligence research. At the other extreme, higher-intelligence men were reasonably well represented, and atypically so, because there was mandatory conscription during these years. Thus, despite some restriction at the low end, our sample captures a broad range of intelligence. Comprehensive details of the VES sampling design are given on the Centers for Disease Control website (also in Centers for Disease Control, 1989 p. 43).

2.2. Cognitive ability measures

The 4462 veterans took a large battery of neuropsychological tests when they were examined in 1985–6. Of these, we chose 5 cognitive ability tests for their sound psychometric properties as well as their suitability for a *g* factor. We settled on the Verbal and Arithmetic tests of the Army Classification Battery (ACB) (Montague, Williams, Gieseking, & Lubin, 1957), the Information and Block Design subtests of the Wechsler Adult Intelligence Scale-Revised (WAIS-R) (Wechsler, 1981), and the Reading subtest of the Wide Range Achievement Test (WRAT-R) (Jastak & Jastak, 1965). Both the ACB and WAIS pairs of tests, one test in each being verbal and the other nonverbal, are sometimes used to measure overall intelligence. We included a second verbal test, the WRAT-R, to compensate for the low ceiling on the ACB Verbal test. The *g* factor explained >58% of the total variance among the 5 test scores.

We also extracted two additional *g* scores to replicate key analyses: an alternative *g* from tests taken at time of physical examination ($N = 3627$), and an induction-age *g* ($N = 3572$). The second 'middle-age *g*' was derived from ten tests, the five tests in our 5-test *g* plus five other, more neurodiagnostically-oriented tests. The 5- and 10-test *g*s correlated $r = .97$. Lastly, we extracted an earlier-life *g* factor from four mental tests that had been administered when all 18,313 veterans in the larger VES study had first been inducted into the US army. These were the verbal comprehension test (VE) and arithmetic reasoning (AR) tests from the general technical (GT) examination (which is often used by cognitive epidemiologists alone as an index of general intelligence), plus a Pattern Analysis (PA) test and General Information Test (GIT). This 'induction-age *g*' correlated $r = .87$ with both of the 'middle-age' general factor scores.

2.3. Medical abnormality counts

As well as neuropsychological test scores, the VES researchers collected other health information including

self-reported, physician-assessed, and laboratory-test data. We used count data to construct abnormality 'test' scores that would sample different aspects of health so that we could examine how they related to each other and to *g*. We selected 'items' that were either physician-assessed or laboratory data (rather than self-reported items — with the exception of covariates discussed below). We used mostly dichotomous items where the physician response was unambiguous ('normal' versus 'abnormal'). For the few items that used continuous measures (such as the urinalysis results), we followed the reference values used in the analyst's codebook, provided by the Centers for Disease Control, to code 'normal' versus 'abnormal'. The health examinations followed a protocol where, in the case of ocular health, for example, a series of observations were made about disorders such as 'Retinal abnormality: arterio-venous nicking, arteriolar spasm, exudates' (and so on). Each 'item' was rated 'normal' or 'abnormal' by the physician. We scored each 'normal' as zero and each 'abnormal' as one. In the case of bilateral abnormalities, we scored a double impairment as two. Continuing the ocular example, a man with no impairments would score zero and a man who had arteriolar spasm in both eyes and exudates in one eye would score three.

These items were aggregated into eight abnormality counts corresponding to different categories of health indicators in the VES, without any weighting of constituent items. The first four abnormality counts included physician-diagnosed signs of neurological abnormalities: Cranial nerves, Motor nerves, Reflexes, and Peripheral Sensory nerves. The other four abnormality counts concerned Head, Urinalysis, General Physical, and Skin items. Abnormalities of the eyes, ears, nose, throat, sinus and salivary glands were summed to make the 'Head scale'; abnormalities from urinalysis laboratory data (given the VES reference norms) were summed to make a 'Urinalysis scale'; dermatology items were summed to make a 'Skin scale' and, a heterogeneous set of conditions (such as club toes, joint swelling, or hernia) were summed to create a 'General Physical Abnormalities scale'. Bearing in mind our 'fitness factor hypothesis', we excluded measures where the relationship to genetic fitness seemed especially vexed. For example, a 'missing limb' in this sample is likely to reflect a combat injury in Vietnam rather than a congenital abnormality arising from genetic mutations. Appendix A gives the list of the variables used to create each scale.

2.4. Covariates

The covariates included age (at time of physical examination in 1985–6), body mass index (BMI at time of examination), alcohol (self-reported alcoholic drinks per month in 1985), and smoking (self-reported cigarettes smoked per day in 1985). We did not expect to see much of an age effect in our age-restricted sample, but since age affects both health and intelligence, we included it as a covariate. Although unconnected with our central hypothesis, we also examined service in Vietnam (because of possible exposure to toxins that might harm both intelligence and health) compared with service elsewhere, and self-reported use of marijuana or hard drugs.

We focused on covariates that might influence health, rather than covariates that might influence intelligence. By age 38 (our sample average), the heritability of intelligence is both high and stable. Twin and adoption studies show that shared family environment has zero effect on intelligence in adulthood (see for review Plomin & Petrill, 1997). We did not include socio-economic status as a covariate because behavioural genetic studies suggest that socio-economic status in mobile societies is mostly a consequence, rather than a cause, of intelligence differences (Scarr & Weinberg, 1978). (This fact also reduces the plausibility of the cognitive epidemiology model in which intelligence and health are both influenced by common environmental factors.) Socio-economic differences certainly influence *what* people know (a bright person from a blue collar environment may not recognize an escutcheon, whereas a dull member of the British aristocracy will readily identify a bend sinister), but the evidence is compelling that such socio-economic differences do not contribute to the latent intelligence factor under analysis here.

3. Analyses and results

3.1. Descriptives

Table 1 shows the basic descriptive statistics for the key measures. The *g* factor is distributed fairly normally. Compared to the full sample of 4,462, our subsample of non-Hispanic White veterans showed a slightly higher mean (.16 versus 0.0) and a slightly lower SD (.88 versus 1.0). Most of the abnormality counts are, as expected, not normally distributed and form Poisson-like distributions — because that is the nature of count data. For any given item (such as 'eye exudates') most people score normal (0), and only a small percentage score abnormal (1 or perhaps 2 for bilateral items). The number of items that constitute each scale varies, but no man scored the maximum possible on any scale.

3.2. Pearson correlations

If the *g* factor reflects the general efficiency of neuropsychological functioning, then reverse-coded *g* should reflect

Table 1
Descriptive statistics for *g* factor, eight abnormality counts, and six covariates.

	Mean	SD	Range	Skewness	Kurtosis	(*N*)
g factor	.16	.88	−2.76–1.78	−.73	.02	(3653)
Abnormality-count scales (# items)						
Cranial (31)	.42	.86	0–13	3.22	21.55	(3654)
Motor (44)	.32	1.43	0–29	10.65	157.88	(3654)
Reflexes (11)	2.40	3.03	0–11	1.08	−.14	(3654)
Peripheral sensory (24)	.29	.96	0–19	7.12	91.68	(3654)
Head (29)	.86	1.16	0–8	1.50	2.61	(3654)
Urine tests (12)	.29	.66	0–7	3.09	13.00	(3654)
General physical (22)	.42	.70	0–5	1.92	4.47	(3654)
Skin (62)	4.06	2.44	0–14	.56	.03	(3654)
Smoking (current cig/day)	19.85	16.31	0–80	.50	−.17	(3652)
Alcohol (current drinks/mo)	33.98	55.17	0–720	3.97	26.59	(3637)
Drug use, past or present (no/yes)	.27	.45	0–1	1.01	−.98	(3654)
Body mass index (BMI)	26.80	4.41	16–67	1.58	7.16	(3653)
Age at exam	38.35	2.50	31–49	.13	.06	(3654)
Vietnam service (no/yes)	.56	.50	0–1	−.25	−1.94	(3654)

Table 2
Correlations among *g* factor (reverse-scored), eight abnormality-count scales, and four covariates ($N = 3633$ White Army veterans).

	Abnormality counts ($\log_{10}$ transformed)									Covariates		
	g (rev)	Cranial	Motor	Reflex	Periph	Head	Urine	Gen phys	Skin	Smok	Alcoh ($\log_{10}$)	BMI age ($\log_{10}$)
Cranial	.038*											
Motor	.108**	.143**										
Reflexes	.053**	.081**	.087**									
Peripheral sen	.071**	.141**	.167**	.079**								
Head	.189**	.054*	049**	.036*	.039*							
Urine tests	.006	−.005	.053**	−.006	.014	052**						
General phys	.068**	.027	.079**	.015	.020	.089**	.026					
Skin	−.014	−.020	−.019	.032	.061**	.005	.000	.005				
Smoking	.042*	.006	.041*	.045**	.012	.099**	.072**	.038*	.026			
Alcohol (log)	−.037*	−.048**	.001	.012	−.020	.045**	.049**	.000	−.058**	.142**		
BMI (log)	.032	.003	−.032	.059**	−.010	−.008	−.096**	.016	.164**	−.020	−.062**	
Age	−.071**	.004	−.003	.022	.054**	−.011	−.016	−.012	.105**	−.055**	−.073**	.065**

*Indicates $p < .05$ (two-tailed); **indicates $p < .001$ (two-tailed).

general neuropsychological impairment, and should be positively correlated with physical impairments as indexed by our abnormality counts. However, the abnormality counts are not normally distributed, so we log-normalized them before correlational analysis. Table 2 shows the simple bivariate Pearson correlations among reverse-scored *g*, the logged abnormality counts, and the four continuous covariates, two of which (alcohol use and BMI) were also logged owing to non-normality. Five abnormality-count scales (the four neurological scales plus the Head scale) correlate with each other ($rs = .036$–$.167$) as well as with lower *g* ($rs = .038$–$.189$). Correlations among the three other abnormality counts are fewer and smaller, and only one (General Physical) correlates significantly with lower *g*. Higher *g* men in this sample tended to be slightly older and drink more, but smoke less. Men who smoked more had higher abnormality counts on five of the eight scales. There was no pattern, however, to the correlations for alcohol consumption and BMI; they each correlated with 3–4 of the 8 abnormality counts, half the time negatively and half positively, but never both in the same direction.

We re-ran the correlations using the two alternative *g*s, the alternative 10-test 'middle-age *g*' and the 4-test 'induction g'. Both yielded the same pattern of significant correlations as seen in Table 2, the only difference being that the former were sometimes a bit higher and the latter sometimes a bit lower.

3.3. Effects of prevalence of abnormalities on their correlations with intelligence

This variability in correlations among the abnormality scales may be influenced somewhat by statistical artifacts, such as differences in the reliability of physician ratings across abnormalities, or in prevalences of those abnormalities. There are only spotty data on reliability, but we could examine the effect of prevalence. The issue is this. Almost all of the individual abnormalities were very rare (low prevalence), typically just a few percent in the sample. Not only are lower-prevalence items subject to more sampling error, but the rarer they are, the smaller their variance, and hence the lower their correlations are constrained to be for statistical (not substantive) reasons. Thus, we expected that lower-prevalence items might show weaker correlations with general intelligence. This was indeed the case: across all of the individual abnormality items (see Appendix A), there was a correlation of $r = .30$ ($p < .000$) between an abnormality's prevalence and its correlation with *g*. This suggests that there would be a stronger positive manifold among items and scales if the medical abnormalities were more common, or measured more precisely as continua rather than normal/abnormal dichotomies.

3.4. Poisson regressions

We examined the relations between the abnormality counts and *g* in a second way. Instead of log transforming the abnormality counts, as in the correlational analyses, we regressed each on *g* and the covariates using Poisson regression. Although the method does not allow us to analyze all the counts simultaneously, it is more statistically suited to individual count variables such as ours. Each abnormality count (Cranial, Motor, Reflexes, Peripheral Sensory, Head, Urinalysis, General Physical, and Skin anomalies) was regressed on *g* and the six covariates, to identify any significant predictors.

These analyses, shown in Table 3, tell basically the same story as the correlations in Table 2. The regression coefficients for *g* are significant for the same six of eight abnormality counts: Cranial, Motor, Reflexes, Peripheral Sensory, Head, and General Physical. None of the six covariates had such consistent influence. Only for Skin abnormalities did more than two of them have significant coefficients. The *g* factor appears to be the only consistent predictor of abnormality counts.

Once again, we replicated the analyses using the alternative, 10-test *g* at middle-age and then the 'induction g'. Both sets of regressions yielded the same pattern of results as shown in Table 3, with one exception. The coefficients for 'induction *g*' were no longer significant for cranial nerve and reflex abnormality counts, but four others remained significant.

Taken together these results show that the relationship between the abnormality count scales and various measures of intelligence do not vary much even when the measure of *g* changes and when influences from the environment sustained over two decades of life, including army service, are

Table 3
Poisson regressions of eight physiological abnormality counts on the *g* factor and six covariates (3633 White Army veterans).

	Neurological counts								Other counts							
	Cranial (31 items)		Motor (44)		Reflexes (11)		Peripheral sensory (24)		Head (29)		Urine tests (12)		General physical (22)		Skin (62)	
	B	*p*	*B*	*p*	*B*	*p*	*B*	*p*	*B*	*p*	*B*	*p*	*B*	*p*	*B*	*p*
Intercept	−.72	.452	1.09	.588	−1.29	.470	**−3.99**	.005	−.46	.432	**3.07**	.003	**−1.63**	.032	**−1.73**	.000
g factor	**−.09**	.037	**−.35**	.000	**−.06**	.014	**−.26**	.000	**−.27**	.000	−.02	.701	**−.13**	.000	.01	.477
Smoking	.00	.408	.01	.072	.00	.132	.00	.682	**.01**	.000	**.01**	.000	.00	.075	**.00**	.002
Alcohol	**−.13**	.004	−.07	.504	.02	.185	−.07	.372	**.07**	.027	**−.12**	.023	−.00	.941	**−.03**	.014
Drugs ever	−.01	.892	.00	.908	−.01	.312	.20	.155	.04	.506	.08	.407	.11	.085	.00	.882
BMI	−.01	.983	−1.46	.231	**1.10**	.011	−.09	.926	−.09	.802	**−3.25**	.000	−.45	.315	**1.55**	.000
Age	.00	.962	−.00	.910	.01	.074	**.07**	.001	.01	.486	−.00	.843	.00	.867	**.02**	.000
Vietnam	.02	.785	−.16	.271	.05	.524	.03	.760	−.02	.687	.20	.121	−.02	.728	−.02	.310

Notes: statistically significant results are in boldface type. All abnormality counts, alcohol (current drinks per month) and BMI were $\log_{10}$ transformed.

interposed between the first and second time of measuring intelligence. This supports our suggestion that the relationships among the abnormality counts and *g* are mediated by more than lifestyle factors.

3.5. Factor analysis of abnormality count scales

We also ran some exploratory factor analyses to test the notion of a fitness factor influencing phenotype-wide traits. It is important to bear in mind that the Poisson-distributed medical sign scales have much weaker measurement properties for this purpose than do well-designed IQ-type items and tests, so would be expected to underestimate a general factor, whatever its true magnitude.

We examined an unrotated principal axis factor that included the eight logged abnormality counts. The first factor, fairly normally distributed, explains around 7% of the variance among these eight measures. This general abnormalities factor correlates .16 with reverse-scored *g* ($p<.001$), and the abnormality scales loading most heavily on it are the same as those with significant *g*-correlations in Table 2. The Kaiser–Meyer–Olkin measure of sampling adequacy is .588. Bartlett's test of sphericity indicates that the correlation matrix is not an identity matrix, but has sufficient non-random structure to allow factor analysis (Chi-squared = 399.17, df 28, $p<.000$).

3.6. Factor analysis of (dichotomous) abnormality items

We further investigated the factor structure of medical abnormalities with an item-level analysis, by factor-analyzing a subset of the individual abnormality items (as opposed to aggregate abnormality counts) — specifically, those 24 items that had the desirable measurement properties of being unambiguous with regard to their good/bad status, and of showing decent prevalence (at least 40 out of 3654, or about 1%). From these 24 items, we performed factor analyses using Promax rotation (because the fitness factor hypothesis suggests that the factors should correlate), using weighted least-squares extraction with mean and variance adjustment. We found that a four-factor solution fit the data reasonably well (Chi-squared = 100.53, df 85, $p=.12$). The root mean square error of approximation (RMSEA) was estimated at .01; the root mean square residual (RMSR) was estimated at .09. This is above the RMSR cut-off of .07 suggested by experts (Hu & Bentler, 1999), but given other fit indices and our low-frequency count items we think the solution worth reporting. These four correlated factors suggest that the data are somewhat structured although we do not draw strong conclusions about the meanings of the factors.

4. Discussion

Among these 3654 Vietnam-era veterans (mean age 38), the bivariate correlations were positive, although small, amongst general intelligence (*g*) and 6 of the 8 medically-assessed abnormality counts (Cranial nerves, Motor nerves, Reflexes, Peripheral Sensory nerves, Head, General Physical). The Poisson regressions also showed that intelligence significantly predicts 6 out of the 8 medical abnormality counts, even after controlling for 6 key lifestyle covariates and risk factors (age, smoking, drinking, drug use, BMI, age, place of military service). The factor analysis of individual abnormality count scales revealed a weak general abnormalities factor that also correlated with intelligence. Further, the abnormality-item prevalences correlated moderately with their *g*-correlations. The exploratory factor analysis of 24 individual abnormality items revealed some correlated factors as well. Finally, all of these results were highly similar whether analysis used the five-test *g* factor, the ten-test *g* factor or the induction-age *g* factor assessed two decades earlier.

Together, these results give some preliminary support for the hypothesis that a general genetic fitness factor may partly explain the many correlations between intelligence and health that have been observed in cognitive epidemiology. Other recent findings provide convergent support (Miller & Penke, 2007; Posthuma et al., 2003; Prokosch et al., 2005; Silventoinen, Posthuma, van Beijsterveldt, Bartels, & Boomsma, 2006; Sundet, Tambs, Harris, Magnus, & Torjussen, 2005). For example, in a previous analysis of VES data, we reported positive correlations (.14–.19) between intelligence and three measures of semen quality—sperm count, sperm concentration, and sperm motility—which is likely to be a highly fitness-related trait in males (Arden, Gottfredson, Miller, & Pierce, 2009).

The lifestyle covariates also predicted abnormality counts to some degree. Smoking tobacco predicted higher counts in five out of the eight scales, and alcohol use predicted in four — half positively and half negatively. However, intelligence predicted abnormality counts more consistently than any of

these lifestyle factors did, and the number of significant relations between lifestyle and abnormalities was halved when including *g* in the Poisson regressions. These results suggest that intelligence's correlations with health are not just due to lifestyle.

4.1. How the findings relate to the fitness factor hypothesis

Since most genes influence several traits in parallel, many harmful mutations are likely to disrupt several traits in parallel. These correlated disruptions should give rise to positive (but small) genetic correlations across traits in their adaptive efficiency (Houle, 2000; Miller, 2000). Moreover, there are likely to be substantial differences across individuals in their overall 'mutation load' — the number and severity of harmful mutations (Crow, 2000). In a species such as ours, mutual mate choice amplifies variation in mutation load through assortative mating for genetic quality (Hooper & Miller, 2008). Together, these three effects–pleiotropic mutations, variation in mutation load, and assortative mating for genetic quality–are likely to create generally modest positive correlations across almost all phenotypic traits.

4.2. Limitations of the study

Our data have some serious limitations for purposes of testing the fitness factor hypothesis. These limitations could be overcome by further research with other datasets and health measures. In particular, while count data are commonly used in medical research, they are statistically problematic.

First, although many statistical procedures are available for analyzing normally distributed data, there are few techniques for analyzing count data, where most individuals score at the minimum (here, 0) and the number of people scoring at successively higher levels drops dramatically. Thus, while we could use Poisson regression to analyze one abnormality count at a time, there is no appropriate factor analytic procedure, to our knowledge, for extracting a general factor from a matrix of such scales. This is why we tried to analyze the data in several different ways.

Second, abnormality counts depend on aggregating somewhat arbitrary medical judgments about where to draw the line between what is normal versus abnormal, which lowers the psychometric reliability of such measures. For many medical signs, there are few objective bases for scoring a trait as statistically 'abnormal', much less as evolutionarily 'maladaptive'. For example, a cholesterol value that is lower than the reference norms for Westernized populations might be adaptively normal among hunter–gatherer societies with lower-fat diets. We have tried to avoid these ambiguities by focusing on items that seem likely to impair normal functioning in any population. Physicians must also rely on fallible personal judgment when assessing most abnormalities. This may partly account for the notorious unreliability of clinical judgments. Not surprisingly, the levels of inter-rater agreement reported for a subset of our dichotomous measures ranged widely but tended to be low overall, despite careful training of physician raters. Example kappa values are illustrative: varicocele, .28; acne (grade 1), .24; mouth dental status, .14 (Centers for Disease Control, 2007) (Medical and Psychological Data Quality Supplement B page 9). Low kappas work against our finding a fitness factor because measurement unreliability reduces observed correlations.

Third, by definition, abnormality counts aggregate scores on a series of dichotomous variables. A dichotomous variable differs from a continuous one in that its mean (proportion affected, p) strongly influences its variance ($p(1-p)$). The closer p is to .5, the greater the measure's variance and, in turn, its capacity to correlate with other variables. This fact creates statistical artifacts that, unless appreciated, can mislead when interpreting results. For example, the .3 correlation we found between each individual abnormality's prevalence and its correlation with *g* (Section 3.3) suggests that the underlying relationships between physical health and intelligence are being obscured by differential prevalence across abnormality items. Such artifacts would be greatest when individual abnormalities ('items') rather than counts of them ('scales') are being analyzed. Hence our reluctance to place much weight on our factor analysis of 24 specific abnormalities even though at first glance it might seem the most conceptually direct way of testing the fitness factor hypothesis.

Fourth, while the sample size and comprehensiveness of the medical examinations are impressive, we theorize that any particular medical abnormality is, at best, a very weak indicator of the proposed fitness factor. Just as individual IQ test items have relatively low *g* loadings, specific medical abnormalities will have very low 'fitness loadings'. Our hope was that by aggregating a number of abnormalities, the tiny valid variance in each item would accumulate, much as it does across items in a lengthy IQ test. The challenge for testing the fitness hypothesis is that particular medical abnormalities may be far weaker markers of general fitness than rigorously designed intelligence test items are of *g*.

Fifth, our abnormality scales treat all rated abnormalities as equally important, but we expect that they actually vary substantially in their medical severity, stability over time, implications for survival and reproduction, fitness loadings, mutational target size, and the number and diversity of constituent biological processes that they reflect. For example, the *g* factor is likely to summarize the efficiencies of thousands of constituent neurophysiological processes and the number and severity of mutations across thousands of genes, so could easily show correlations with other higher-order summary characteristics such as longevity, social success, and sexual attractiveness. By contrast, urine analysis abnormalities may reflect more localized problems in liver, kidney, or metabolic function, which could reflect fewer or very rare single-gene polymorphisms. To take another example, the General Physical abnormality scale includes some items that seem manifestly more important to reproductive success and more fitness-loaded (such as whether retinas are normal) than others (whether pubic hair pattern is normal). While it is not ideal to treat all abnormalities as equivalent, it is biologically parsimonious given our limited understanding of how medical disorders relate to biological fitness. It is also statistically parsimonious, given that unit-weighted composite variables usually perform about as well statistically as do more diversely-weighted composites (Bobko, Roth, & Buster, 2007; Dawes, 1979; Raju, Bilgic, Edwards, & Fleer, 1997).

Given these limitations, the surprise is not that the correlations we observed were small, but that they kept appearing. Although larger effect sizes (following Cohen's

suggestion of how to characterize sizes (Cohen, 1988) are the bread and butter of social scientists, miniscule correlations are the bread and butter of evolution. For evolution, what matters is robustness over time (which we cannot test, here), not size of effect at any one time. The critical question is 'are the effects we saw tiny but robust or, tiny and fugitive?' Only further research can answer this question, but the pervasiveness of small correlations with *g* across various organ systems and types of analysis suggests that it is worth pursuing.

4.3. Genetic tests of the fitness factor idea

The fitness factor hypothesis is amenable to more direct empirical tests based on new genetic data. We would predict that a range of phenotypes, including general intelligence, will covary with a reasonable index of mutation load (see for example (Boyko et al., 2008; Gorlov, Gorlova, Sunyaev, Spitz, & Amos, 2008; Mitchell-Olds, Willis, & Goldstein, 2007). These mutation load scores would include summing the total number of very low-frequency SNPs and copy-number variants across an individual's genome. If reliable indices of mutation load do not correlate with a wide range of important phenotypes, then the fitness factor idea may be safely discarded.

5. Conclusion

The field of cognitive epidemiology should be concerned with all of the possible causal relationships between intelligence on health, not just the relations among phenotypic intelligence, lifestyle, social environments, and health. The elimination of health inequalities is a frequently stated goal of cognitive epidemiology. If our fitness factor hypothesis is correct, we might view health inequalities in a somewhat different light. Some health disparities across socio-economic groups may not be prima facie evidence of a dysfunctional society, but may reflect genetic variance in mutation load that affects both physical health and general intelligence (which, in turn, influences socio-economic success). Evolution itself, through pleiotropic mutations and assortative mating, may maximize the range of genetic quality across individuals and the strength of genetic correlations across traits, with the side-effect that it maximizes the apparent unfairness of medical, educational, and socio-economic outcomes. This is not a cause for gloom, however. Evolution has also supplied us with insight, empathy and a sense of fairness. Equipped with a clear understanding of the world as it exists, good men and women have always found opportunities to reduce avoidable suffering by exercising these gifts.

Appendix A. List of variables constituting each of the eight abnormality-count scales

Cranial scale (neurology) variables, 31 (as ordered in Centers for Disease Control, Table G.1, Vol. III)

Smell-rt
Smell-lt
Visual field-rt
Visual field-lt
Optic disc-rt
Optic disc-lt
Pupil size-rt-(mm)
Pupil size-lt-(mm)
Light react-rt
Light react-lt
Ptosis-rt
Ptosis-lt
Occular mobil-rt
Occular mobil-lt
Nystagmus-rt
Nystagmus-lt
Jaw strength
Jaw jerk
Facial pain
Corneal reflex-rt
Corneal reflex-lt
Facial muscles-rt
Facial muscles-lt
Palate motion-rt
Palate motion-lt
Gag reflex
Acc nerves-rt
Acc nerves-lt
Tongue motion-rt
Tongue motion-lt
Other cranial cond

Motor scale (neurology) variables, 44 (Centers for Disease Control Table G.1 in Vol. III)

Amput loss
Gait
Arm swing-rt
Arm swing-lt
Tandem gait
Eyes open
Eyes closed
Abnorm muscle
Muscle "tone"
Atrophy
Strength-rt deltoids
Strength-lt deltoids
Strength-rt biceps
Strength-lt biceps
Strength-rt-triceps
Strength-lt triceps
Strength-rt wrist ext
Strength-lt wrist ext
Strength-rt grip
Strength-lt grip
Strength-rt fing abdu
Strength-lt fing abdu
Strength-rt hip flex
Strength-lt hip flex
Strength-rt knee ext
Strength-lt knee ext
Strength-rt knee flex
Strength-lt knee flex
Strength-rt dorsiflex
Strength-lt dorsiflex
Strength-rt plan flex
Strength-lt plan flex

Strength-rt toe ext
Strength-lt toe ext
Tremors-rt arm
Tremors-lt arm
Finger-nose ataxia
Hand pronation
Heel-shin ataxia
Finger tapping
Arm drift
Excess rebound
Speech
Other motor cond

Reflexes scale (neurology) variables, 11 (Centers for Disease Control, Table G.1 in Vol. III)

Reflex-rt biceps
Reflex-lt biceps
Reflex-rt triceps
Reflex-lt triceps
Reflex-rt knee
Reflex-lt knee
Reflex-rt ankle
Reflex-lt ankle
Reflex-rt plantar
Reflex-lt plantar
Other reflex cond

Peripheral sensory scale (neurology) variables, 24 (Centers for Disease Control Table G.1 in Vol. III)

Pinprick-rt arm-pd (proximal dorsal)
Pinprick-lt arm-pd
Pinprick-rt-arm-pv (proximal ventral)
Pinprick-lt-arm-pv
Pinprick-rt arm-dd (distal dorsal)
Pinprick-lt arm-dd
Pinprick-rt arm-dv (distal ventral)
Pinprick-lt arm-dv
Pinprick-rt leg-pd
Pinprick-lt leg-pd
Pinprick-rt leg-pv
Pinprick-lt leg-pv
Pinprick-rt leg-dd
Pinprick-lt leg-dd
Pinprick-rt leg-dv
Pinprick-lt leg-dv
Vib-rt-lat mall
Vib-lt-lat mall
Vib-rt-patella
Vib-lt patella
Sens ext-face
Sens ext-arms
Sens ext-legs
Sens ext-visual

Head scale variables, 29 (Centers for Disease Control, Table C.1, Vol. III)

Eye-scarring
Eye-cataract
Eye-scleral icterus
Eye-arterio-venous nicking
Eye-arteriolar spasm
Eye-exudates
Eye-papilledema
Eye-cupping
Eye-disc pallor
Eye-hemorrhages
Ear-cerumen impact
Ear-inflammation
Ear-drum perforated
Ear-drum retracted
Ear-drum scarred
Ear-drum bulging
Ear-drum inflamed
Nose-abnormality (inc. septum, polyps, ulceration, bleeding)
Throat-pharyngitis
Throat-tonsils
Mouth-dental status
Mouth-ulcers
Mouth-plaques
Mouth-mass
Mouth-glossitis
Mouth-gums
Sinuses-frontal
Sinuses-maxillary
Salivary glands

Urinalysis scale variables, 12 (Centers for Disease Control, Table 12.7, Vol. III)

Glucose
Ketones
Protein
Bilirubin
Urobilinogen
Haemoglobin
Red blood cells
White blood cells
Hyaline casts
Granular casts
Red cell casts
White cell casts

General physical scale variables, 22 (Centers for Disease Control, Table C.1, Vol. III)

Breast-gynecomastia
Abdominal -visible abnorm
Abdominal-palpable mass
Abdominal -tenderness
Abdominal -palpable liver
Spleen palpable
Costovertebral angle -tender
Hernia
Pubic hair abnormal male pattern
Penis abnormal
Epididymis
Varicocele
Scrotal mass
Prostate

Ext-club fingers
Ext-club toes
Oedema
Ext-acrocyanosis
Ext-soft tissue mass
Ext-joint swell
Scoliosis-spine
ln-lymph nodes

Skin scale variables, 62 (Centers for Disease Control, Table B.1, Vol. III)

Hyperpigmentation
Hypopigmentation
Birthmarks
Other condition
Alopecia
Alopecia, scarring
Alopecia, nonscarring
Hirsuitism
Other hair condition
Acne, grade i
Acne, grade ii
Acne, grade iii
Acne, grade iv
Acne, atypical
Comedones only
Folliculitis
Hidraden suppur
Tinea of nails
Candida
Tinea versicolor
Tinea other
Infect cond other
Neoplastic
Cancer of skin
Dermato-fibromas
Epidermal cysts
Kera actinic
Kera seborrheic
Lipomas
Milia
Nevi atypical
Sebaceous hyperplasia
Warts, nongenital
Neoplastic cond other
Capillarities
Hemangioma
Palmar erythema
Poikiloderma civatte
Spider angiomas
Telangiectasias
Vasculitis
Varicosities
Vascular cond other
Aphthosis
Bullae
Vesicles
Eczematous derm
Dyshidrosis
Lichen simplex
Lichen planus
Psoriasis
Excoriations
Rosacea
Seborr dermatitis
Angular stomatitis
Urticaria
Other inflam cond
Asteatosis
Keratosis pilaris
Pityriasis alba
Striae
Other condition

References

Andersson, M., & Simmons, L. W. (2006). Sexual selection and mate choice. *Trends in Ecology and Evolution*, *21*(6), 296−302.

Arden, R., Gottfredson, L. S., Miller, G., & Pierce, A. (2009). Intelligence and semen quality are positively correlated. *Intelligence*, *37*, 277−282.

Batty, G. D., Deary, I. J., & Gottfredson, L. S. (2007). Premorbid (early life) IQ and later mortality risk: Systematic review. *Annals of Epidemiology*, *17*(4), 278−288.

Bobko, P., Roth, P. L., & Buster, M. A. (2007). The usefulness of unit weights in creating composite scores — A literature review, application to content validity, and meta-analysis.*Organizational Research Methods*, *10*(4), 689−709 (ß).

Boyko, A. R., Williamson, S. H., Indap, A. R., Degenhardt, J. D., Hernandez, R. D., Lohmueller, K. E., et al. (2008). Assessing the evolutionary impact of amino acid mutations in the human genome. *PloS Genetics*, *4*(5).

Centers for Disease Control (1989). *The health status of Vietnam veterans.* Atlanta, Georgia: US Dept. of Health and Human Services.

Centers for Disease Control (2007). *The health status of Vietnam veterans.* from http://www.cdc.gov/nceh/veterans/default1c.htm

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*, (2nd Ed.) Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Crow, J. F. (2000). The origins, patterns, and implications of human spontaneous mutation. *Nature Reviews. Genetics*, *1*(1), 40−47.

Dawes, R. M. (1979). Robust beauty of improper linear models in decision-making. *American Psychologist*, *34*(7), 571−582.

Deary, I. J., & Der, G. (2005). Reaction time explains IQ's association with death. *Psychological Science*, *16*(1), 64−69.

Fuhrer, R., Shipley, M. J., Chastang, J. F., Schmaus, A., Niedhammer, I., Stansfeld, S. A., et al. (2002). Socioeconomic position, health, and possible explanations: A tale of two cohorts. *American Journal of Public Health*, *92*(8), 1290−1294.

Gorlov, I. P., Gorlova, O. Y., Sunyaev, S. R., Spitz, M. R., & Amos, C. I. (2008). Shifting paradigm of association studies: Value of rare single-nucleotide polymorphisms. *American Journal of Human Genetics*, *82*(1), 100−112.

Gottfredson, L. S. (2006). Innovation, fatal accidents and the evolution of general intelligence. In R. M. J. (Ed.), *Integrating the Mind* (pp. 387−425). Hove, UK: Psychology press.

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

Hart, C. L., Taylor, M. D., Smith, G. D., Whalley, L. J., Starr, J. M., Hole, D. J., et al. (2005). Childhood IQ and all-cause mortality before and after age 65: Prospective observational study linking the Scottish Mental Survey 1932 and the Midspan studies. *British Journal of Health Psychology*, *10*, 153−165.

Hooper, P. L., & Miller, G. F. (2008). Mutual mate choice can drive costly signaling even under perfect monogamy. *Adaptive Behavior*, *16*, 53−70.

Houle, D. (2000). Is there a g factor for fitness? *The nature of intelligence*, *vol. 233*. (pp. 149−170)Chichester: John Wiley and Sons, Ltd.

Houle, D., & Kondrashov, A. S. (2002). Coevolution of costly mate choice and condition-dependent display of good genes. *Proceedings of the Royal Society Of London. Series B, Biological Sciences*, *269*(1486), 97−104.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria in fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*(1), 1−55.

Jastak, J. F., & Jastak, S. R. (1965). *Wide range achievement test: WRAT.* Wilmington, Delaware: Guidance Associates.

Jensen, A. R. (1998). *The g factor: The science of mental ability.* Westport: Conn. Praeger.

Kokko, H., Jennions, M. D., & Brooks, R. (2006). Unifying and testing models of sexual selection. *Annual Review of Ecology Evolution and Systematics*, *37*, 43−66.

Kuh, D., Richards, M., Hardy, R., Butterworth, S., & Wadsworth, M. E. J. (2004). Childhood cognitive ability and deaths up until middle age: A post-war birth cohort study. *International Journal of Epidemiology*, *33*(2), 408−413.

Kumari, M., Seeman, T., & Marmot, M. (2004). Biological predictors of change in functioning in the Whitehall II study. *Annals of Epidemiology*, *14*(4), 250−257.

Laurence, J., & Ramsberger, P. F. (1991). *Low-aptitude men in the military. Who profits? Who pays?* New York: Praeger.

Lynch, M., Blanchard, J., Houle, D., Kibota, T., Schultz, S., Vassilieva, L., et al. (1999). Perspective: Spontaneous deleterious mutation. *Evolution*, *53*(3), 645−663.

Marmot, M. G. (2003). Understanding social inequalities in health. *Perspectives in Biology and Medicine*, *46*(3), S9−S23.

Martin, L. T., & Kubzansky, L. D. (2005). Childhood cognitive performance and risk of mortality: A prospective cohort study of gifted individuals. *American Journal of Epidemiology*, *162*(9), 887−890.

Miller, G. F. (2000). Sexual selection for indicators of intelligence. *The nature of intelligence*, *vol. 233*. (pp. 260−275) Chichester: John Wiley and Sons Ltd.

Miller, G. F., & Penke, L. (2007). The evolution of human intelligence and the coefficient of additive genetic variance in human brain size. *Intelligence*, *35*(2), 97−114.

Mitchell-Olds, T., Willis, J. H., & Goldstein, D. B. (2007). Which evolutionary processes influence natural genetic variation for phenotypic traits? *Nature Reviews. Genetics*, *8*, 845−856.

Montague, E. K., Williams, H. L., Gieseking, C. F., & Lubin, A. (1957). Army tests for assessment of intellectual deficit. *US Armed Forces Medical Journal*, *8*, 883−892.

Plomin, R., & Petrill, S. A. (1997). Genetics and intelligence: What's new? *Intelligence*, *24*(1), 53−77.

Posthuma, D., Baare, W. F. C., Pol, H. E. H., Kahn, R. S., Boomsma, D. I., & De Geus, E. J. C. (2003). Genetic correlations between brain volumes and the WAIS-III dimensions of verbal comprehension, working memory, perceptual organization, and processing speed. *Twin Research*, *6*(2), 131−139.

Prokosch, M. D., Yeo, R. A., & Miller, G. F. (2005). Intelligence tests with higher g-loadings show higher correlations with body symmetry: Evidence for a general fitness factor mediated by developmental stability. *Intelligence*, *33*(2), 203.

Raju, N. S., Bilgic, R., Edwards, J. E., & Fleer, P. F. (1997). Methodology review: Estimation of population validity and cross-validity, and the use of equal weights in prediction. *Applied Psychological Measurement*, *21*(4), 291−305.

Scarr, S., & Weinberg, R. A. (1978). The influence of "family background" on intellectual attainment. *American Sociological Review*, *43*, 674−692.

Silventoinen, K., Posthuma, D., van Beijsterveldt, T., Bartels, M., & Boomsma, D. I. (2006). Genetic contributions to the association between height and intelligence: Evidence from Dutch twin data from childhood to middle age. *Genes, Brain, and Behavior*, *5*(8), 585−595.

Singh-Manoux, A., Ferrie, J. E., Lynch, J. W., & Marmot, M. (2005). The role of cognitive ability (intelligence) in explaining the association between socioeconomic position and health: Evidence from the Whitehall II prospective cohort study. *American Journal of Epidemiology*, *161*(9), 831−839.

Starr, J. M., Taylor, M. D., Hart, C. L., Davey Smith, G., Whalley, L. J., Hole, D. J., et al. (2004). Childhood mental ability and blood pressure at midlife: Linking the Scottish Mental Survey 1932 and the Midspan studies. *Journal of Hypertension*, *22*(5), 893−897.

Sundet, J. M., Tambs, K., Harris, J. R., Magnus, P., & Torjussen, T. M. (2005). Resolving the genetic and environmental sources of the correlation between height and intelligence: A study of nearly 2600 Norwegian male twin pairs. *Twin Research and Human Genetics*, *8*(4), 307−311.

van Oort, F. V. A., van Lenthe, F. J., & Mackenbach, J. P. (2005). Education and mortality: A role for intelligence? Reply. *Journal of Epidemiology and Community Health*, *59*(9), 810−810.

Wechsler, D. (1981). *WAIS-R manual: Wechsler adult intelligence scale-revised.* New York, NY: Harcourt Brace Jovanovich for Psychological Corp.

Whalley, L. J., Fox, H. C., Deary, I. J., & Starr, J. M. (2005). Childhood IQ, smoking, and cognitive change from age 11 to 64 years. *Addictive Behavior*, *30*(1), 77−88.