Heritability and Validity of Healthy Physical Aging (Wellness) in Elderly Male Twins

Terry Reed and Danielle M. Dick

Department of Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, USA

The authors define a trait "wellness" for good health in 6109 men in the National Academy of Sciences-National Research (NAS-NRC) twin panel aged 70 years and up surveyed by mail in the fall of 1998. Men who responded that they had not had a heart attack, coronary surgery, stroke, diabetes or prostate cancer in the survey questionnaire (Q8) met the broad definition of wellness. A more narrow definition included the absence of hypertension. Genetic analysis indicated that over 50% of the population variance for liability to either wellness definition was genetic. A subset of the NAS-NRC twins also participates in the National Heart, Lung and Blood Institute (NHLBI) twin study. NHLBI examinations and medical record review was done in 1986–1987 and 1995–1997 for 389 individuals who completed Q8. Excellent agreement (kappa > 0.8) was found between Q8 and outcome review for each condition comprising the wellness definition, ranging from 0.81 for coronary surgery to 0.88 for diabetes. Substantial agreement (kappa = 0.67) was found for hypertension. Kappa values for wellness were 0.82 for the broader definition and 0.74 if high blood pressure was included. Fraternal twin-pairs concordant for the wellness definitions are currently being recruited for linkage studies.

Most genetic studies involved with aging have been disease specific. More recently, it has been recognized that it is valuable to also examine genetic factors related to disease free or healthy aging (Hadley et al., 2000). For example, the correlation of longevity between parent and child has long been recognized (Pearl, 1931). This rather broad phenotype has moderate (0.20–0.30) heritability (Herskind et al., 1996; Ljungquist et al., 1998; McGue et al., 1993). A recent focus in the genetics of aging in humans has been to study families of centenarians. These kindreds show strong familial effects (Perls et al., 2002), and a locus on chromosome 4 has been linked with extreme longevity in such families (Puca et al., 2001).

Siblings who are concordant for survival to a set age free of any specified diseases can be used to search for genes involved in healthy aging. Although genes related to good health may also include some of those that promote extreme familial longevity, there is no reason to expect genetic factors related to healthy aging are only associated with extreme longevity. We utilized twins from the National Academy of Sciences — National Research Council (NAS-NRC) twin panel, and by using twin-pairs we have also matched for sex and perhaps less environmental variability than between other pairs of relatives. To search for genes for healthy aging, we defined a trait "wellness" from a health history questionnaire mailed to the NAS-NRC twins aged 70 and over in 1998. In this paper we report on two aspects of this study: (1) An analysis of the wellness trait to ensure there are significant genetic effects and (2) a comparison of questionnaire responses about health to physical examinations and medical records on a subset of the NAS-NRC twins to judge the validity of the measure for wellness.

Materials and Methods

Participants

In 1955, the NAS-NRC initiated the development of a veteran twin registry by matching birth certificate information, collected from 39 of the then 48 states plus all of Louisiana, except for the city of New Orleans, to Veteran Administration records. The resulting registry of 15,924 white male twin-pairs represented 93% of male twin births from 1917 through 1927. A more detailed description of the creation of the registry is published (Jablon et al., 1967; Hrubec & Neel, 1978). It has been estimated that zygosity is correctly determined for at least 95% of the 13,487 twin-pairs assigned a zygosity in the panel (Jablon et al., 1967) using a combination of blood typing, fingerprints, and the twins' own assessment of their zygosity. Approximately 80% of pairs assigned a zygosity were classified solely on their own assessment from a zygosity questionnaire mailed in 1965 (Hrubec & Omenn, 1981).

For the phenotype of wellness in this study, responses were used from a health history questionnaire (Q8) mailed in the fall of 1998. Q8 was pilot tested on a small sample of male spouses of patients attending the Indiana University Alzheimer's disease clinic, and refined from feedback and analysis of the responses given by these elderly male caregivers who were the approximate same ages as the NAS-NRC twins to be surveyed. Q8 was mailed to all complete NAS-NRC twin-pairs (4086) known to be alive with a current address and 906 singles who completed two earlier

Address for correspondence: Terry E. Reed, PhD, Department of Medical & Molecular Genetics, Indiana University School of Medicine, 975 W. Walnut Street, IB — 130, Indianapolis, IN 46202 USA E-mail: treed@iupui.edu

health history questionnaires in the 1970s and 1980s. Excluding those with invalid addresses or the death of the subject, 6109 of 8848 (69%) of Q8 were completed. Among this group were 2055 complete pairs with a mean age of 74.3 years.

Wellness

For this study, a definition of good health, termed "wellness", from the Q8 responses was determined for 1019 identical (monozygous, MZ) and 986 fraternal (dizygous, DZ) complete twin-pairs. An individual met the wellness definition if he answered "NO" to all of the following questions: (1) Has a doctor ever told you that you had a heart attack? (2) Have you ever had coronary bypass surgery or angioplasty? (3) Have you ever been told by a doctor that you had a stroke? (4) Has a doctor ever told you that you have diabetes? (5) Have you ever been diagnosed with prostate cancer? These traits were selected to define an individual who has successfully aged into his 70s by eliminating most of the major health problems. Prostate cancer was the only type of cancer with specific questions in Q8. Prostate cancer was included in the wellness definition because it was felt to be an important cancer compatible with lengthy survival. It was judged that if an individual were diagnosed with most other cancers, he would be less likely to participate by completing the Q8 questionnaire and would probably be deceased or too ill to participate at the time of the follow-up linkage study. A more restrictive definition of wellness also included answering "NO" to the following additional question: Has a doctor or other medical person every told you that you have high blood pressure? Both the broader and narrow definition of wellness were analyzed. If an individual answered "YES" to any item, they were counted as not meeting the wellness definition. Otherwise if an individual answered any individual item "don't know" or left any item blank, they were excluded from analyses.

It is, of course, possible that other conditions not included in the wellness definition might also impact on functional abilities and quality of life. Many of these chronic conditions were not available in Q8 and/or for record verification in the validity study. Individuals with severe limitations because of such conditions also were judged less likely to participate in a future linkage study. However, an item in Q8 from the SF-36 health survey (Ware, 1997) asked the individual to rate his own health on a scale from 1 to 5. In twins meeting the broader wellness definition, 91.5% rated their health as good to excellent (codes 1–3) compared with 74.2% in those not meeting the wellness criteria. The percentage rating health as excellent was 21.3% in the wellness group versus 6.2% in those not meeting criteria for wellness, and the figures for those reporting very good health were 39.9% versus 26.3%, respectively. Only 1.6% of those in the wellness category rated their health as poor (code 5). Difference in ratings between the wellness group and all other subjects was tested using Proc MIXED from the SAS statistical package. This procedure allows testing of the mean difference in rating with an F test while also accounting for the correlation between twins in complete pairs and it allows the correlation to differ between zygosity groups. The better self-health ratings of subjects meeting both the broad and narrow definitions of wellness were highly significant: $F_{1,1801} = 613.7$, $p < .0001$: $F_{1,1808} = 441.5$, $p < .0001$. The only other part of the SF-36 included in Q8 was the physical function scale. This scale is moderately correlated with the self-health rating item, and also showed significantly ($p < .0001$) better physical function scores in those meeting the wellness definition. Thus, we feel that our definition of wellness accurately captures twins who have enjoyed healthy aging.

Validity of Q8 Responses

A subset of the NAS-NRC twin panel, the National Heart, Lung, and Blood Institute (NHLBI) twin study comprising 514 twin-pairs, was created in 1969–73 at five examination centers to study the genetics of cardiovascular disease risk factors. Details of recruitment into the NHLBI twin study from the NAS-NRC registry, participation rate, determination of zygosity, and initial examination protocol are published (Feinleib et al., 1977). There have been a total of five examinations over a roughly 30-year period. Mean age of the cohort at entry was 48 years. At the third examination that took place in 1986–1987, 622 (268 pairs and 86 singletons) participated at the age of 63 and validated events from medical record review were first undertaken. Events were classified as definite when all diagnostic criteria were met, probable when some of the criteria were met and possible when a physician stated the participant had an event but no confirmatory medical records were able to be obtained (Reed et al., 1991). At Exams 4 (July 1995–April 1997) and 5 (1999–2000) events were classified as definite if confirmed by records and possible if there were no confirmatory record of an event ever obtained.

The Q8 survey was undertaken between Exams 4 and 5 of the NHLBI twin study. Four hundred and twenty-three of the NHLBI subset returned a completed Q8. Eliminating individuals who only participated at the NHLBI entry examination or did not fully complete all the Q8 health history questions defining wellness, 389 individuals were utilized for validation of events in Q8. NHLBI Exam 5 occurred at least one year after the Q8 survey. A "NO" response for a Q8 item but a "YES" for the same item determined from medical records showed greater disparity at Exam 5 than Exam 4. Given the ages of the twins it is likely that "YES" to events at Exam 5 reflect events mostly occurring after Q8 was returned. The reverse ("NO" for records and "YES" at Q8) were much more consistent between Exams 4 and 5. As a result responses on Q8 questions about heart attack, coronary surgery, stroke, diabetes, prostate cancer and hypertension and the broad and narrow definitions of wellness were compared with record review through Exam 4 of the NHLBI twins. In addition to medical records, outcome assessment of the NHLBI twins included a detailed physical examination by a physician at Exams 1–3. At all exams, blood pressures were measured, information concerning medications including anti-hypertensives, insulin, diabetic diet or oral hypoglycemic agents were collected, and questions at each exam were asked about heart attack, coronary surgery, stroke, diabetes or cancer (specify the type) since the previous examination.

The measure of agreement between the Q8 questionnaire data and the outcome assessment of events of the NHLBI twins was determined by using the kappa statistic (Fleiss, 1981). Kappa is the proportion of agreement above chance and has been shown to be a satisfactory measure for dichotomous variables, as in the present study with a 2 by 2 table (Maclure & Willett, 1987). The proportion of agreement for an individual event (e.g., heart attack, diabetes, wellness) was also calculated (Grant, 1991). In addition, sensitivity, specificity, positive and negative predictive powers were calculated from the 2 by 2 table, assuming the NHLBI outcome review as a gold standard. Most studies using kappa to assess agreement employ the criteria of Landis & Koch (1977); values between 0.4 and 0.6 are considered moderate .6 to .8 substantial and over .8 excellent agreement.

Genetic Analyses

Contingent upon certain assumptions, comparisons of MZ and DZ twin pairs yield estimates of the amount of variance in a trait that is because of genetic and environmental effects (Kendler, 1993). This is because MZ twins share all of their genes in common, whereas DZ twins share, on average, just half of their genetic variation, as do ordinary siblings. Two types of genetic effects can be assessed: additive effects (A) and dominance (D). Dominant genetic effects are implied when the DZ correlation is less than half of the MZ correlation; for a trait solely influenced by additive genetic effects, we would expect the DZ correlation (0.50) to be exactly half that of the MZ correlation (1.0). Two types of environmental influences can also be assessed with comparisons of twins: common environmental effects (C), which make siblings more alike than would be expected solely based on shared genes, and unique environmental effects (E), which make siblings different from one another. When twins are the only family members available for analysis, C and D parameters cannot be estimated simultaneously in a genetic model.

To estimate genetic and environmental influences on wellness, we used the program Mx (Neale et al., 1999), designed specifically to deal with genetically informative data. Because age influences health risk, the average age of the twins at the time of Q8 response was included in the model. This allowed the amount of influence attributed to age effects to be estimated in addition to genetic and environmental effects. The average age in twin-pairs with complete questionnaire data was 74.6 years (SD = 2.74). Only 4 twins out of 1829 pairs had an age difference > 1 year at the time of questionnaire completion; these pairs were omitted from modeling analyses. When data is ordinal rather than continuous, polychoric correlation matrices and asymptotic covariance matrices are used as input into Mx rather than variance-covariance matrices. These were computed using Prelis 2.3 (Joreskog & Sorbom, 1993). The fits of the various models were determined by evaluating the log-likelihood statistic. The p-values > .05 indicate that the model provides a good fit to the data. Additionally, Akaike's information criterion (AIC) provides another index of the adequacy of a model fit; more negative AIC values indicate a better fit. Finally, the Root Mean Squared Error Approximation (RMSEA) statistic should be .05 or less for a very good fit, or between .05 and .10 for a good fit (Browne & Cudek, 1993). After fitting the full model, a series of sub-models was fit to test the significance of each parameter. This was done by systematically dropping parameters from the model and determining whether this caused a subsequent decrease in fit. Differences between the log-likelihoods of nested models were compared with a χ^2 distribution. When there was a significant χ^2 ($p < .05$) for the difference in degrees of freedom between the models, the model with fewer degrees of freedom was adopted.

Results

Table 1 shows kappa values, agreement for the medical condition, sensitivity, specificity, and positive and negative predictive power for each individual condition used to define wellness. All kappa values are in the excellent range except for hypertension which would be considered as substantial agreement. Of the eight false positive individuals who indicated that they had had a heart attack in the Q8 response, two from medical records were diagnosed with coronary

Table 1

Kappa, Agreement for Event, Sensitivity, Specificity, Positive and Negative Predictive Powers for the Individual Conditions used in the Wellness Definitions

	NHLBI Exam Outcome Assessment					
	Heart Attack	Coronary Surgery	Stroke	Diabetes	Prostate Cancer	Hypertension
Q8 Response	Yes No	Yes No	Yes No	Yes No	Yes No	Yes No
Yes	44 8	59 5	38 5	46 3	34 12	149 9
No	5 330	16 303	8 335	8 329	1 340	56 175
κ (95% CI)	0.85 (0.79–0.91)	0.81 (0.76–0.87)	0.83 (0.78–0.88)	0.88 (0.82–0.93)	0.82 (0.77–0.87)	0.67 (0.61–0.72)
Agreement for event	0.77 (0.66–0.88)	0.74 (0.64–0.83)	0.74 (0.62–0.86)	0.81 (0.71–0.91)	0.72 (0.60–0.85)	0.70 (0.64–0.76)
Sensitivity	80.0%	78.7%	82.6%	85.2%	97.1%	72.7%
Specificity	97.6%	98.4%	98.5%	99.1%	96.6%	95.1%
+ pred. power	84.6%	92.2%	88.4%	93.9%	73.9%	94.3%
– pred. power	98.5%	97.7%	97.7%	97.6%	99.7%	75.8%

insufficiency, and another had congestive heart failure. Hypertension had the lowest negative predictive power and sensitivities and 14.4% of the Q8 respondents indicated they did not have hypertension, but were either on hypertension medication or had elevated blood pressure at the various NHLBI examinations.

Table 2 displays the estimates of agreement for the broad and narrow definitions of wellness. The broad definition of wellness had excellent agreement; the narrow definition substantial agreement. Since hypertension had the least agreement among the individual medical conditions, it is not surprising that the narrow definition including hypertension had decreased concordance for wellness.

For the broad definition of wellness, excluding consideration of hypertension, there were 950 MZ pairs and 864 DZ pairs with complete data. The tetrachoric correlation for the definition of wellness excluding hypertension was 0.55 for MZ pairs and 0.23 for DZ pairs. The results of model fitting for the broad definition of wellness are presented in top half of Table 3. Since the DZ correlation was close to half the MZ correlation both the ACE (1) and ADE (2) models were fit to the data; however, as would be expected from the correlations, the ADE model provided a better fit to the data, as indicated by a higher p-value and lower AIC. Thus all sub-models were subsequently fitted from the full ADE model with age effects. Dropping the effect of age (3) caused a significant decrease in fit of the model $\Delta\chi^2 = 8.90$ (1 df), $p = .003$. Dominance effects (4) were not significant $\Delta\chi^2 = 0.59$ (1 df), $p = 0.44$. However, additive genetic effects were highly significant, and dropping them from the model (5) caused a significant decrease in fit $\Delta\chi^2 = 156.21$ (2 df), $p < .001$.

For the narrow definition of wellness including hypertension, there were 954 MZ and 867 DZ pairs with complete data. The tetrachoric correlation was 0.59 for MZ pairs and 0.20 for DZ twins. Because the DZ correlation was substantially less than half the MZ correlation for the narrow definition of wellness, an ADE model was fit to the data rather than the ACE model. The results of model fitting for the narrow phenotypic definition are presented in the bottom half of Table 3. These results were similar to those for the broad definition of wellness. The full

Table 2

Kappa, Agreement for Wellness, Sensitivity, Specificity, Positive and Negative Predictive Powers for both Definitions

	NHLBI Exam Outcome Assessment	
	Wellness (narrow)	Wellness (broad)
Q8 Response*	Yes No	Yes No
Yes	113 37	202 20
No	8 227	14 146
κ (95% CI)	0.74 (0.69–0.79)	0.82 (0.76–0.87)
Agreement for wellness	0.71 (0.64–0.79)	0.86 (0.81–0.90)
Sensitivity	93.4%	93.5%
Specificity	86.0%	88.0%
+ predictive power	76.0%	91.0%
– predictive power	96.6%	91.2%

Note: * Narrow definition: answered no to ever having or been told you have had a heart attack, bypass surgery or angioplasty, stroke, diabetes, prostate cancer, and high blood pressure. Broader definition excludes high blood pressure.

Table 3

Model-Fitting Results for Definitions of Wellness: Narrow Excluding Those with Hypertension, Broader Ignoring Hypertension Status

		Fit Statistics			
Model	*df*	χ^2	*p*	AIC	RMSEA
Wellness-Broader definition					
1 ACE model with age effects	2	1.30	0.52	–2.70	0.00
2 ADE model with age effects	2	0.71	0.70	–3.29	0.00
3 ADE model	3	9.61	0.02	3.61	0.01
4 AE model with age effects	3	1.30	0.73	–4.70	0.00
5 E model with age effects	4	156.92	0.00	148.92	0.13
Wellness-Narrow definition					
1 ADE model with age effects	2	0.24	0.89	–3.76	0.00
2 ADE model	3	8.45	0.04	2.45	0.01
3 AE model with age effects	3	3.62	0.31	–2.38	0.00
4 E model with age effects	4	310.68	0.00	302.68	0.19

Table 4

Variance Partitioning for Wellness Definitions into Genetic (A), Environmental (E), and Age-Related Influences. Estimate under the Best-Fitting Models (95% Confidence Intervals)

Variance Component	Narrow (including hypertension)	Broad (excluding hypertension)
A	0.57 (0.52–0.63)	0.53 (0.46–0.61)
E	0.42 (0.35–0.49)	0.45 (0.37–0.54)
Age	0.01 (0.001–0.022)	0.01 (0.001–0.033)

model (1) fit the data quite well by all fit indices. Dropping the effect of age (2) on wellness caused a significant decrease in fit of the model: $\Delta\chi^2 = 8.21$ (1 df), $p = .004$. Although the model allowing for dominance genetic effects had a slightly better fit as indexed by a lower AIC value, dropping the dominance effects (3) did not cause a significant decrease in fit of the model: $\Delta\chi^2 = 3.38$ (1 df), $p = .066$. Additive genetic effects were significant, and could not be dropped from the model (4) without causing a highly significant decrease in fit: $\Delta\chi^2 = 310.44$ (2 df), $p < .001$.

Thus, for both definitions of wellness, the AE model with age effects was the best fitting model. The amount of variance that can be attributed to A, E, and age for each phenotype, as well as the 95% confidence intervals on these estimates, can be found in Table 4. Both heritability estimates (A) for wellness are over 50%.

Discussion

The men who are part of the NAS-NRC panel, and particularly the NHLBI subset have been involved in examinations related to research studies for over the past 30 years. Thus, they provide a valuable resource for the study of aging. The responses of the NHLBI subset on the Q8 health history items in this study were compared to physical examination measurements and collection of medical records over this same period to test the agreement of the questionnaire assessment with the medical exams. It has been suggested that veterans may be a subset of the population that provide highly accurate reports of chronic disease (Horner et al., 2001). Our data seem to support this assertion as we find high kappas for nearly all outcomes studied. We would caution that the kappa values from the NHLBI subset who have regularly attended examinations may represent the upper level of agreement that could be expected from responses in the full cohort.

Despite the potential for upward bias in estimates of validity from the Veteran twin sample, other studies also support the use of questionnaire data to assess health history. Most validation studies of self-assessed disease have been in diabetes and hypertension. Survey data for these two conditions have even been used as a measurement standard (Hebert et al., 1999; Quam et al., 1993). For diabetes, most studies show moderate to high sensitivity and consistently high specificity paralleling our data (Bowlin et al., 1993; Edwards et al., 1994; Heliovaara et al., 1993; Kehoe et al., 1994; Kriegsman et al., 1996; Haapanen et al., 1997; Martin et al., 2000; Midthjell et al., 1992; Paganini-Hill & Ross, 1982; Robinson et al., 1997). Kappa values for hypertension are slightly lower than those reported for diabetes (Edwards et al., 1994; Kehoe et al., 1994; Haapanen et al., 1997; Halabri et al. 1992; Heliovaara et al., 1993; Paganini-Hill & Ross, 1982) and are in line with our results. The decrease in reliability of high blood pressure reporting may relate to differences in opinion of what level really is high blood pressure and how many elevated readings are necessary to make the diagnosis.

For both stroke and heart attack, we find high specificity and negative predictive power. This has been reported in several previous studies of stroke (Engstad et al., 2000; Heliovaara et al., 1993; Horner et al., 2001; Olsson et al., 1989) and heart attack or myocardial infarction (Haapanen et al., 1997; Harlow & Linet, 1989; Olsson et al., 1989; Walker et al., 1998). Transient ischemic attacks have been suggested to result in a substantial fraction of false positive responses and hence decreased sensitivity for stroke (Engstad et al. 2000; Walker et al., 1998). Some false positive responses for heart attack in Q8 were due to coronary insufficiency or congestive heart failure diagnoses upon review of the medical records. These "false positives" are not especially troubling for our study because regardless such individuals would not be considered to have a successful aging phenotype.

The results of genetic analyses indicated that both our broad and narrow wellness definitions have reasonable heritability to make the search for genes related to this definition of good aging worthwhile. These heritability estimates are in line with previous studies that have investigated several of the individual elements that are included in our measure of good health. The Danish twin registry estimated that > 50% of the variance in heart diseases was attributable to genetic effects in both males and females (Weinke et al., 2001). Data from the HERITAGE family study found that genetic influences accounted for 34% and 53% of the variance in coronary heart disease risk respectively in whites and blacks (Katzmarzyk et al., 2000). The San Antonio family heart study found that genes accounted for 15–45% of the variation in a variety of measures of cardiovascular risk factors (Mitchell et al., 1996). Previous studies have also provided evidence of significant heritability for diabetes and diabetes-related quantitative traits, such as glucose and insulin levels and urinary albumin excretion (Fogarty et al., 2000; Hsueh et al., 2000). Prostate cancer has also been demonstrated to be under genetic influence: a large study of twins from Swedish, Danish, and Finnish twin registries estimated that the heritability of prostate cancer was 42% (Lichtenstein et al., 2000). This estimate is slightly lower than the 57% heritability estimate reported in this sample of World War II veteran twins (Page et al., 1997). Finally, the Danish twin register estimated that 32% of the variation in liability to stroke death was due to genetic factors (Bak et al., 2002). A number of factors that contribute to stroke, such as platelet aggregation, have also been shown to be under significant genetic influence (O'Donnell et al., 2001). Thus, the heritability of our wellness phenotype is in good agreement with previous studies of aging outcomes, paralleling and often exceeding the heritabilities of related aging phenotypes.

The best fitting models suggest that genetic influences are additive. This suggests that healthy aging fits a polygenic model of inheritance, where one's aging phenotype likely results from the effects of many genes, which act additively. Dominance effects were not significant, although the test comparing AE versus ADE (or ACE) models is not very powerful, so it is possible that dominance may play a small role.

In all models, the contribution of age was very small. This probably reflects the narrow age range of our subjects (69–82). Over half of participants were between the ages of 71 and 74 and 80% between 71 and 77. Had our subjects been of a wider age range, then age may have been

a more important factor. Analyses ignoring age led to very similar heritability results in this sample.

It is important to address potential selection in our sample, based both on our phenotype definition and the sample itself. Our definition of healthy aging doesn't include other potential chronic conditions and is focused primarily on physical wellness. We cannot state definitively that there aren't individuals with other chronic conditions in our healthy aging sample, but we believe the impact of such individuals would be slight. Analyses using the narrower definition of wellness, which included hypertension, led to slightly higher heritability estimates. We did not exclude persons who met the physical wellness definition but rated their health as fair or poor. There may be other reasons to give low self-ratings (recent death of close family members or recent injury or recent flu or severe colds that seemed to hang on) that would not reflect poor aging as we have defined it. Additionally, the substantial estimates of genetic variance for wellness in our sample of approximately 75-year-old men may differ from estimates obtained with other selection procedures in the general population. This group of men because of selection has been healthier than others their age in the general population. The veterans had to be part of a pair who both completed the induction physical examination in World War II and both survived their military service. Subsequently, both agreed to become part of the panel in the 1960s, and then both survived into their 70s and agreed to participate again by both completing Q8. This sample is a magnification of the so-called healthy worker effect whereby those who work and also volunteer for studies are healthier than their cohorts in the general population. Many of the potential causes of mental un-wellness (dementia, schizophrenia, severe depression) would also be selected out through this process. Thus, our definition of wellness focuses on healthy physical aging. Despite these selection biases, the genes leading to wellness in this cohort should be similar to those in less selected cohorts in the general population.

In conclusion, healthy aging, as measured among elderly males using questionnaire assessment, exhibits good validity and significant heritability. The broader definition of wellness, excluding hypertension, exhibited better psychometric properties, and significantly increased the number of individuals who met the wellness criteria. This suggests that this phenotype may be a strong candidate for use in analyses aimed at identifying genes involved in healthy aging. We are currently recruiting concordant DZ (sib) pairs from this cohort for initial linkage studies of wellness.

Acknowledgments

Supported by grant R01AG18736 (T.R.) and NRSA F32AA113358 (D.M.D.). We wish to thank Drs. William F. Page and Joe C. Christian for helping to make the Q8 questionnaire survey possible. Thanks to Peggy A. Knople, Shannon A. Rinehart, and Tiffany S. Robinson for their technical help and Chris J. Williams PhD for his helpful suggestions using Proc MIXED.

References

Bak, S., Gaist, D., Sindrup, S. H., Skytthe, A., & Christensen, K. (2002). Genetic liability to stroke: A long-term follow-up study of Danish twins. *Stroke, 33*, 769–774.

Bowlin, S. J., Morrill, B. D., Nafzinger, A. N, Jenkins, P. L., Lewis, C., & Pearson, T. A. (1993). Validity of cardiovascular disease risk factors assessed by telephone survey: The behavioral risk factor survey. *Journal of Clinical Epidemiology, 6*, 561–571.

Browne, M. W., & Cudek, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen, & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Newbury Park: SAGE Publication.

Edwards, W. S., Winn, D. M., Kurlnatzick, V., Sheridan, S., Berk, M. L., Retchin, S., & Collins, J. G. (1994). *Evaluation of national health interview survey diagnostic reporting.* Hyattsville, MD: National Center for Health Statistics.

Engstad, T., Bonaa, K. H., & Viitanen, M. (2000). Validity of self-reported stroke. The Tromso study. *Stroke, 31,* 1602–1607.

Feinleib, M., Garrison, R. J., Fabsitz, R., Christian, J. C., Hrubec, Z., Borhani, N. O., Kannel, W. B., Rosenman, R., Schwartz, J. T., & Wagner, J. O. (1977). The NHLBI twin study of cardiovascular disease risk factors: Methodology and summary of results. *American Journal of Epidemiology, 106*, 284–295.

Fleiss, J. L. (1981). *Statistical methods for rates and proportions* (2nd ed., pp. 212–236). New York: John Wiley & Sons.

Fogarty, D. G., Rich, S. S., Hanna, L., Warram, J. H., & Krolewski, A. S. (2000). Urinary albumin excretion in families with type 2 diabetes is heritable and genetically correlated to blood pressure. *Kidney International, 57*, 250–257.

Grant, J. M. (1991). The fetal heart rate trace is normal, isn't it: Observer agreement of categorical assessments. *Lancet, 337*, 215–218.

Haapanen, N., Miilunpalo, S., Pasanen, M., Oja, P., & Vuori, I. (1997). Agreement between questionnaire data and medical records of chronic diseases in middle-aged and elderly Finnish men and women. *American Journal of Epidemiology, 145*, 762–769.

Hadley, E. C., Rossi, W. K., & Albert, S. (2000). Genetic epidemiologic studies on age-specified traits. *American Journal of Epidemiology, 152*, 1003–1008.

Halabi, S., Zurayk, H., Awaida, R., Awaida, R., Darwish, M., & Saab, B. (1992). Reliability and validity of self and proxy reporting of morbidity data: A case study from Beirut, Lebanon. *International Journal of Epidemiology, 21*, 607–612.

Harlow, S. D., & Linet, M. S. (1989). Agreement between questionnaire data and medical records: The evidence for accuracy of recall. *American Journal of Epidemiology, 129*, 233–248.

Hebert, P. L., Geiss, L. S., Tierney, E. F., Engelgau, M. M., Yawn, B. P., & McBean, A. M. (1999). Identifying persons with diabetes using Medicare claims data. *American Journal of Medical Quality, 14*, 270–77.

Heliovaara, M., Aromaa, A., Klaukka, T., Knekt, P., Joukamaa, M., & Impivaara, O. (1993). Reliability and validity of interview data on chronic diseases: The mini-Finland health survey. *Journal of Clinical Epidemiology, 46*, 181–191.

Herskind, A. M., McGue, M., Holm, N. V., Sorensen, T. I., Harvald, B., & Vaupel, J. W. (1996). The heritability of

human longevity: A population-based study of 2872 Danish twin pairs born 1870–1900. *Human Genetics, 97*, 319–323.

Horner, R. D., Cohen, H. J., & Blazer, D. G. (2001). Accuracy of self-reported stroke among elderly veterans. *Aging and Mental Health, 5*, 275–281.

Hrubec, Z., & Neal, J. V. (1978). The national academy of sciences-national research council twin registry: Ten years of operation. *Progress in Clinical and Biological Research, 24B*, 153–172.

Hrubec, Z., & Omenn, G.S. (1981). Evidence of genetic predisposition to alcoholic cirrhosis and psychosis: Twin concordances for alcoholism and its biological end points by zygosity among male veterans. *Alcoholism, Clinical and Experimental Research, 5,* 207–215.

Hsueh, W. C., Mitchell, B. D., Aburomia, R., Pollin, T., Sakul, H., Gelder Ehm, M., et al. (2000). Diabetes in the older order Amish: Characterization and heritability analysis of the Amish family diabetes study. *Diabetes Care, 23*, 595–601.

Jablon, S., Neel, J. V., Gershowitz, H., & Atkinson, G. F. (1967). The NAS-NRC twin panel: Methods of construction of the panel, zygosity diagnosis, and proposed use. *American Journal of Human Genetics, 19,* 133–161.

Joreskog, K. G., & Sorbom, D. (1993). *New features in PRELIS2*. Chicago, IL.

Katzmarzyk, P. T, Perusse, L., Rice, T., Gagnon J., Skinner, J. S., Wilmore, J. H., Leon, A. S., Rao, D. C., & Bouchard, C. (2000). Familial resemblance for coronary heart disease risk: The HERITAGE family study. *Ethnicity & Disease, 10*, 138–147.

Kehoe, R., Wu, S. Y., Leske, M. C., & Chylack, L. T. (1994). Comparing self-reported and physician-reported medical history. *American Journal of Epidemiology, 139*, 813–818.

Kendler, K. S. (1993). Twin studies of psychiatric illness: Current status and future directions. *Archives of General Psychiatry, 50,* 905–915.

Kriegsman, D. M., Penninx, B. W. J. H., vanEijk, J. T. M., Boeke, A. J. P., & Deeg, D. J. H. (1996). Self-reports and general practitioner information on the presence of chronic diseases in community dwelling elderly: A study on the accuracy of patients' self-reports and on determinants of inaccuracy. *Clinical Epidemiology, 49*, 1407–1417.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*, 159–174.

Lichtenstein, P., Holm, N. V., Verkasalo, P. K., Iliadou, A., Kaprio, J., Koskenvuo, M., Pukkala, E., Skytthe, A., & Hemminki, K. (2000). Environmental and heritable factors in the causation of cancer — Analyses of cohorts of twins from Sweden, Denmark, and Finland. *New England Journal of Medicine, 343*, 78–85.

Ljungquist, B., Berg, S., Lanke, J., McClearn, G. E., & Pedersen, N. L. (1998). The effect of genetic factors for longevity: A comparison of identical and fraternal twins in the Swedish twin registry. *Journal of Gerontology: Series A Biological Sciences Medical Sciences, 53*, 441–446.

Maclure, M., & Willett, W. C. (1987). Misinterpretation and misuse of the kappa statistic. *American Journal of Epidemiology, 126*, 161–169.

Martin, L. M., Leff, M., Calonge, N., Garrett, C., & Nelson, D. E. (2000). Validation of self-reported chronic conditions and health services in a managed care population. *American Journal Preventative Medicine, 18*, 215–218.

McGue, M., Vaupel, J. W., Holm, N., & Harvald, B. (1993). Longevity is moderately heritable in a sample of Danish twins born 1870–1880. *Jounal of Gerontology, 48*, 237–244.

Midthjell, K., Holmen, J., Bjorndal, A., & Lundlarsen, P. G. (1992). Is questionnaire information valid in the study of a chronic disease such as diabetes? The Nor-Trondelag diabetes study. *Journal of Epidemiology and Community Health, 46*, 537–542.

Mitchell, B.D., Kammerer, C.M., Blangero, J., Mahaney, M.C., Rainwater, D.L., Dyke, B., et al. (1996). Genetic and environmental contributions to cardiovascular risk factors in Mexican Americans: The San Antonio family heart study. *Circulation, 94*, 2159–2170.

Neale, M. C., Boker, S. M., Xie, G., & Maes, H. H. (1999). *Mx: Statistical modeling* (5th ed.). VCU Box 900126, Richmond VA, 23298: Department of Psychiatry.

O'Donnell, C. J., Larson, M. G., Feng, D., Sutherland, P. A., Lindpaintner, K., Myers, R. H., et al. (2001). Genetic and environmental contributions to platelet aggregation: The Framingham heart study. *Circulation, 103*, 3051–3056.

Olsson, L., Svardsudd, K., Nilsson, G., Ringqvist, I., & Tibblin, G. (1989). Validity of a postal questionnaire with regard to the prevelance of myocardial infarction in a general population sample. *European Heart Journal, 10*, 1011–1016.

Paganini-Hill, A., & Ross, R. K. (1982). Reliability of recall of drug usage and other health related information. *American Journal of Epidemiology, 116*, 114–22.

Page, W. F., Braun, M. M., Partin, A. W., Caporaso, N., & Walsh, P. (1997). Heredity and prostate cancer: A study of World War II veteran twins. *Prostate, 33*, 240–245.

Pearl, R. (1931). Studies on human longevity. IV. The inheritance of longevity, preliminary report. *Human Biology, 3*, 245–269.

Perls, T. T., Wilmouth, J., Levenson, R., Drinkwater, M., Cohen, M., Bogan, H., et al. (2002). Life-long sustained mortality advantage of siblings of centenarians. *Proceedings of the National Academy of Sciences, 99,* 8442–8447.

Puca, A. A., Daly, M. J., Brewster, S. J., Matise, T. C., Barrett, J., Shea-Drinkwater, M., et al. (2001). A genome-wide scan for linkage to human exceptional longevity identifies a locus on chromosome 4. *Proceedings of the National Academy of Sciences, 98*, 10505–10508.

Quam, L., Ellis, L. B. M., Venus, P., Clouse, J., Taylor, C. G., & Leatherman, S. (1993). Using claims for epidemiologic research: The concordance of claims-based criteria with the medical record and patient survey for identifying a hypertensive population. *Medical Care, 31*, 498–507.

Reed, T., Quiroga, J., Selby, J. V., Carmelli, D., Christian, J. C., Fabsitz, R. R., et al. (1991). Concordance of ischemic heart disease in the NHLBI twin study after 14–18 years of follow-up. *Journal of Clinical Epidemiology, 44*, 797–805.

Robinson, J. R., Young, T. K., Roos, L. L., & Gelskey, D. E. (1997) Estimating the burden of disease: Comparing administrative data and self-reports. *Medical Care, 35*, 932–47.

Walker, M. K., Whincup, P. H., Shaper, A. G., Lennon, L. T., & Thomson, A. G. (1998). Validation of patient recall of doctor-diagnosed heart attack and stroke: A postal questionnaire and record review comparison. *American Journal of Epidemiology, 148*, 355–361.

Ware, J. Jr. (1997). *SF-36 Health Survey, manual and interpretation.* Boston: The Health Institute, New England Medical Center.

Wienke, A., Holm, N. V., Skytthe, A., & Yashin, A. I. (2001). The heritability of mortality due to heart diseases: A correlated frailty model applied to Danish twins. *Twin Research, 4*, 266–274