Gender-Related Individual Differences and Mortality in the Terman Longitudinal Study: Is Masculinity Hazardous to Your Health?

Article *in* Personality and Social Psychology Bulletin · December 2000

DOI: 10.1177/01461672002612010

Leslie R Martin
La Sierra University

Gender-Related Individual Differences and Mortality in the Terman Longitudinal Study: Is Masculinity Hazardous to Your Health?

Richard A. Lippa
California State University, Fullerton
Leslie R. Martin
La Sierra University
Howard S. Friedman
University of California, Riverside

Data were collected, refined, and analyzed on 654 men and 210 women in Lewis Terman's "gifted children" longitudinal study to examine links between masculinity and mortality. Masculinity measures included gender diagnosticity (GD) scores, measuring the male- or female-typicality of occupational preferences in 1940 and masculinity-femininity (M-F) scores from the Strong Vocational Interest Blank (SVIB). Hazard analyses showed GD was significantly related to mortality for both men and women (interquartile relative hazard 1.25 for men and 1.62 for women), with masculine women and masculine men more likely to die at any given age. SVIB M-F was similarly related to mortality for both men and women (respective interquartile relative hazards = 1.26 and 1.36). The effects remained significant after controlling for certain health behaviors and Big Five traits.

On average, men die sooner than women (Nathanson, 1984; Niven & Carroll, 1993; Strickland, 1988; Verbrugge, 1985; Wingard, 1984)—a difference that shows itself for most major causes of death, including heart disease, cancer, car accidents, homicide, and suicide. Hypothesized factors leading to higher male mortality include biological differences between men and women (Hazzard, 1990), social roles and the differing stresses they place on men and women (Waldron, 1995b), gender differences in behavioral risk factors (Waldron, 1995a, 1995b), and gender differences in personality (Helgeson, 1995).

The robust relationship between gender and mortality suggests a closely related question: Are within-sex, gender-related individual differences similarly related to mortality?

The data to be reported here will provide evidence that the answer to this question is, yes. Specifically, they will show that men and women who are more male-typical in their occupational preferences are more likely to die at any given age than individuals who are less male-typical. The data come from Lewis Terman's longitudinal study of "gifted children" (e.g., Terman & Oden, 1947), which has served in recent years as a major source of information about psychological factors influencing health and mortality (see Friedman, Tucker, Schwartz, Tomlinson-Keasey, et al., 1995, for an overview).

Personality, Illness, and Longevity

There is a long history of research documenting associations between personality and illness. Much early research on this topic, however, considered only single disease entities and used unreliable measures and unconvincing correlational designs. After an early emphasis on Type A behavior and coronary heart disease, empirical investigations of personality and illness/mortality gradually broadened in scope, focusing

Authors' Note: This project was supported by a research grant (No. AGO8825) from the National Institute of Aging. Part of the data were made available from the Terman Life-Cycle Study, begun by Lewis Terman. Special thanks go to Eleanor Walker and Al Hastorf for their assistance in collecting Strong Vocational Interest Blank data from the Terman archives and to Joseph Schwartz for his assistance in data analysis and the preparation of Figure 1. Correspondence concerning this article should be addressed to Richard A. Lippa, Department of Psychology, California State University, Fullerton, CA 92834; e-mail: rlippa@fullerton.edu.

PSPB, Vol. 26 No. 12, December 2000 1560-1570

increasingly on personality traits other than the Type A syndrome and on illnesses other than heart disease. Research suggests that traits such as anxiety and depression show equally strong relations with heart disease as the Type A syndrome. Furthermore, these traits are linked with diseases other than coronary heart disease as well (Friedman & Booth-Kewley, 1987).

A series of recent studies using the Terman longitudinal data have provided new evidence about links between personality and mortality (see Friedman, Tucker, Schwartz, Martin, et al., 1995; Friedman, Tucker, Schwartz, Tomlinson-Keasey, et al., 1995; Friedman et al., 1993). Using archival data collected at regular intervals from 1921 through 1991, Friedman and his colleagues investigated psychosocial variables and mediating risk factors that influence longevity. For example, hazard analyses indicated that childhood Conscientiousness was related to greater adult longevity (considered up to 1986), whereas childhood cheerfulness, contrary to expectation, was related to reduced longevity.

No comparable study has investigated links between masculinity and mortality. Still, evidence is beginning to accrue that gender-related traits may be linked to important health outcomes. Helgeson (1990, 1995) conducted programmatic research on the relationship between various measures of masculinity and coronary heart disease. Specifically, she assessed masculine instrumentality and feminine expressiveness, as well as negative masculinity (socially undesirable male-typical personality traits such as arrogance, hostility, and egotism) (see Helgeson, 1994; Spence, Helmreich, & Holahan, 1979). In a study of 90 patients who had suffered heart attacks, Helgeson (1990) reported evidence that negative masculinity was associated with Type A behavior and furthermore that it was associated with a composite index known to predict subsequent heart disease. In a later study of angioplasty patients, Helgeson (1995) found that negative masculinity was additionally associated with negative social support, psychological distress, and at least one significant risk factor—smoking.

This past research, however, did not directly address the issue of masculinity and mortality. Furthermore, although Helgeson's research included measures of masculine instrumentality, feminine expressiveness, and negative masculinity, it did not include other measures of gender-related traits, such as traditional bipolar masculinity-femininity scales (e.g., Strong, 1943; Terman & Miles, 1936) or gender diagnosticity measures (Lippa, 1991, 1995b; Lippa & Connelly, 1990).

Measures of Masculinity and Femininity

In seminal research, Terman and Miles (1936) presented a bipolar conception of masculinity-femininity (M-F). In essence, Terman and Miles held M-F to be a single dimension, with masculinity and femininity as mutually exclusive endpoints. Terman and Miles, and their many successors, created M-F scales from self-report items that showed reliable and strong sex differences in normative populations.

The bipolar approach to M-F waned by the early 1970s in the face of conceptual and empirical critiques (e.g., Block, 1973; Constantinople, 1973), which argued, in part, that supposedly unidimensional M-F scales were, in fact, multidimensional measures. The bipolar approach was supplanted by a two-dimensional conception of masculinity and femininity, with masculinity (M) defined in terms of instrumental personality traits (e.g., aggressive, dominant, independent) and femininity (F) defined in terms of expressive traits (e.g., warm, sensitive, nurturant) (Bem, 1974, 1981; Spence & Helmreich, 1978; Spence, Helmreich, & Stapp, 1974). A large empirical literature now exists on the psychometric properties and correlates of M and F (see Ashmore, 1990; Cook, 1985; Lenney, 1991).

Although M and F scales continue to be widely used, they have been subject to both psychometric and conceptual critiques. Spence and Helmreich (1980) argued soon after publishing the Personal Attributes Questionnaire (PAQ) that its M and F scales are, in fact, instrumentality and expressiveness scales, which show at best weak and inconsistent relationships to other kinds of gender-related behaviors and attitudes (see Spence & Buckner, 1995). M and F scales suffer from additional problems as well (Lippa, in press; Lippa & Connelly, 1990). Although they adequately describe the gender stereotypic personality traits that are components of most cultures' conceptions of masculinity and femininity (e.g., see Williams & Best, 1990), they fail to embrace a host of other characteristics that are highly relevant to everyday conceptions of masculinity and femininity, such as gender-related appearances, nonverbal behaviors, hobbies, and interests (Lippa, in press).

The gender diagnosticity approach was developed to address some of these problems (Lippa, 1991, 1995a, 1995b; Lippa & Connelly, 1990). In brief, gender diagnosticity (GD) refers to the Bayesian probability that an individual in a given population is male or female based on sets of gender-related indicators (such as occupational preference ratings). GD is formally computed from sets of indicators by applying discriminant analyses (see Lippa & Connelly, 1990).

Previous research has shown that GD can be measured reliably within the sexes from self-report interest data such as occupational preference ratings and that GD measures are factorially distinct from M and F as assessed by the PAQ and Bern Sex Role Inventory (BSRI) (Lippa, 1995b, 1991; Lippa & Connelly, 1990). Furthermore, GD measures are largely independent of the Big

Five personality superfactors, whereas M and F are not (Lippa, 1991, 1995b). GD measures often predict various kinds of gender-related behaviors and attitudes within the sexes better than do M and F (Lippa, 1991, 1995a, 1995b, 1997, 1998a, 1998b, 1998c, 1999a).

Lippa (in press) has proposed that various measures of gender-related traits (M, F, GD) can be conceptualized in terms of broader structural models of individual differences. For example, M and F scales define masculinity and femininity in terms of Big Five factors and facets that show gender differences favoring men or women. In contrast, GD measures define M-F in terms of the vocational circumplex (Holland, 1992; Lippa, 1998b; Prediger, 1982).

In Terman's longitudinal study, the Strong Vocational Interest Blank (SVIB) was administered to 627 men and 200 women in 1940 and to an additional 27 men and 10 women in the 2 years preceding and following 1940. As a result, for most of these participants, scores on a traditional bipolar M-F scale were available as part of the Terman data archive. Furthermore, the occupational preference ratings that Terman's participants made in responding to the Strong test allowed us to compute GD measures for most of the men and women who completed the SVIB. Participants also completed a number of self-report personality items in 1940 (see Martin, 1996; Martin & Friedman, 2000a), which allowed us also to compute measures for four of the Big Five personality superfactors (Extraversion, Agreeableness, Conscientiousness, and Neuroticism), some of which could serve as proxies for masculine instrumentality and feminine expressiveness. Thus, in many ways the Terman data were ideal for a systematic examination of possible links between within-sex, gender-related individual differences and mortality.

METHOD

The Sample

The Terman sample has been extensively described in previous publications (see Friedman et al., 1993). In brief, the original Terman "gifted children" sample comprised 856 boys and 672 girls, most of whom were recruited from schools in California in 1921-1922. At the start of the study in the early 1920s, the children were on average 11 years old. The Terman participants were bright (e.g., Stanford-Binet of 135 or higher) and homogeneous in social class and ethnicity (mostly middle class and White).

In recent studies of predictors of mortality among Terman participants (see Friedman, Tucker, Schwartz, Tomlinson-Keasey, et al., 1995), analyses have been restricted to those born between 1904 and 1915 to achieve a reasonably homogeneous cohort of those in school in 1922. This qualification eliminated 155 of the participants. Of the 1,373 remaining participants, 104 were lost to follow-up at various points in the study; thus, no record exists if or when they died. Finally, 16 participants were excluded who died before 1940.

Measures

The SVIB. In 1940, when the average Terman participant was 30 years old, the SVIB was given "to 627 men and 200 women. It was given only to those women who expressed a desire to take it" (Terman & Oden, 1947, p. 196). The Strong test was administered to an additional 27 men and 10 women in the 2 years preceding and following 1940. Thus, the Strong test was administered to most of the Terman men but only to some of the Terman women. As we shall see in the Results section, the women who completed the Strong test do not seem to have differed from the larger sample of women in their assessed personality or risk factors. Furthermore, the women who took the Strong scale were quite diverse in their occupational choices. In the 1940 questionnaire, 100 of these women listed themselves as employed, 84 described themselves as homemakers, and 25 did not classify themselves. Because the Terman women were relatively young in 1940 (late 20s, early 30s), many of the women who described themselves as employed in 1940 would later give up their jobs upon marrying or having children.

Men and women completed different forms of the Strong test. Men received the Vocational Interest Blank for Men (the SVIB, Form M, 1938) and women received the Vocational Interest Blank for Women (SVIB, Form WA, 1933) (for exact copies of these tests, see Campbell, 1971, Appendix D). Each of these tests had an M-F scale. The M-F scale for men was scored so that high scores were masculine, whereas the M-F scale for women was scored so that high scores were feminine.

Because of differences between the male and female forms of the Strong test, women who completed the 1933 SVIB for women rated their degree of preference (like, indifferent, or dislike) for 128 occupations, whereas men who completed the 1939 men's form rated their degree of preference for 100 occupations. Sixty-one occupations on the men's and women's forms were identical and 1 additional item corresponded closely enough (aviator/aviatrix) to be considered functionally the same (see the appendix for a list of items). These 62 rated occupations were used to compute participants' gender diagnosticity scores.

Personality measures. In 1940, the Terman participants responded to 53 items selected from Bernreuter's (1933) Personality Inventory and in addition rated themselves on the following traits: happiness of temperament, moodiness, impulsivity, self-confidence, emotionality, conformity, ease of getting along with others, enjoyment of social contacts, persistence, definite purposes in life, sensitivity, feelings of inferiority, vanity and egotism, and exclusivity of friendships. Using factor analytic techniques, Martin (1996; Martin & Friedman, 2000b) created from these items scales for four of the Big Five factors: Extraversion, Agreeableness, Conscientiousness, and Neuroticism. Reliabilities (alphas) for these four scales were respectively .65, .72, .65, and .85. The items loading most highly on each scale were as follows: Extraversion—"How much do you enjoy social contacts?" Agreeableness—"Are you always careful to avoid saying anything that might hurt anyone's feelings?" Conscientiousness—"How persistent are you in the accomplishment of your ends?" and Neuroticism—"Are you moody?"

To further demonstrate the validity of these scales, Martin (1996) administered scale items and the NEO-PI (Costa & McCrae, 1992) to a contemporary sample of 203 California adults and then correlated Big Five scales developed from the Terman items with the Big Five scales of the NEO PI-R. The correlations were substantial—.71 for Extraversion, .63 for Agreeableness, .67 for Conscientiousness, and .81 for Neuroticism.

Documentation of death and cause of death. To verify deaths and ascertain causes of death when possible, death certificates were obtained from appropriate agencies in states where Terman participants had died (see Friedman, Tucker, Schwartz, Martin, et al., 1995). Death certificates were coded for cause of death by a certified nosologist in accordance with the ninth revision of the International Classification of Diseases (ICD-9, U.S. Department of Health and Human Services, 1980). In a few cases where death certificates could not be obtained but information about death was available, a trained physician coded cause of death into one of five broad categories: cancer, cardiovascular disease, accident/injury, other, and unknown. In the current study, analyses were restricted to participants who died after the 1940 assessment.

Risk factors. Information was gathered on the following risk factors for participants: smoking, alcohol consumption, obesity, physical activity, and riskiness of hobbies (see Friedman, Tucker, Schwartz, Martin, et al., 1995; Friedman et al., 1993). Information about smoking was obtained from living participants when possible or from relatives of deceased Terman participants in 1991. Participants were classified into four categories, ranging from nonsmoker to heavy smoker, on the basis of computed "cigarette years"—the total number of years an individual had smoked multiplied by the average number of cigarettes smoked per day.[1] Alcohol consumption was assessed in 1950 and 1960 through self-reports. Based on these data, participants were classified into three categories: those who never or seldom drank, those who were moderate drinkers, and those who were heavy drinkers. Obesity was estimated from self-reports in 1940 questionnaires, in which participants reported their height and weight. Body mass index was computed as weight divided by height.[2] Physical activity was rated from participants' self-reported hobbies and activities. Participants reporting at least one hobby or activity requiring high physical activity were rated high on physical activity, whereas those not reporting such a hobby or activity were rated low. Similarly, riskiness of hobbies and activities reported between 1922 and 1940 was rated by trained judges on overall riskiness.

RESULTS

Computing Gender Diagnosticity Measures

We computed gender diagnostic probabilities from participants' occupational preference ratings as described in Lippa (1991, 1995b). In brief, we conducted multiple discriminant analyses on discrete subsets of the 62 Strong occupations rated by both men and women, using sex of participant as the grouping variable. We conducted six discriminant analyses, each on a distinct subset of 10 or 11 rated occupations. Each discriminant analysis yielded the Bayesian probability that a given participant was male or female based on his or her occupational preferences, squared. A participant's overall GD score was simply the mean of the six computed probabilities. In commonsense terms, this score indicates how male- or female-typical an individual's occupational preferences are. The computation of multiple probabilities for each participant allowed us to assess their reliability (see Lippa, 1991, 1995a; Lippa & Connelly, 1990). The reliability (alpha) of the mean GD score was .82 for all participants, .69 for men, and .68 for women.

Intercorrelations of Personality Measures

Table 1 shows the intercorrelations of Extraversion, Agreeableness, Conscientiousness, Neuroticism, GD, and Strong M-F, computed separately for men and women.[3] Strong M-F is reversed for women here so that GD and Strong M-F are keyed in the same direction, with higher scores being more masculine. Table 1 indicates that GD and Strong M-F show little relationship to Big Five factors. However, GD and Strong M-F correlated .71 for men and .57 for women (both ps < .001). This substantial degree of correlation is not unexpected given that both measures were based on the same item domain and both were assessing within-sex, gender-related individual differences.[4]

TABLE 1: Intercorrelations of Personality Measures for Men (above diagonal) and Women (below diagonal)

	Personality Measures					
	Extraversion	*Agreeableness*	*Conscientiousness*	*Neuroticism*	*GD*	*Strong M-F*
Extraversion	—	–.12**	–.10*	.09*	–.02	–.16***
Agreeableness	–.02	—	.13***	.21***	–.03	.06
Conscientiousness	–.08	.11	—	.28***	.09*	.03
Neuroticism	.11	.14*	.20**	—	.17***	.18***
GD	–.20**	.12	.14*	.15*	—	.71***
Strong M-F	–.16*	–.10	.08	.04	.57***	—

NOTE: GD = gender diagnosticity, M-F = masculinity-femininity. For both men and women, Strong M-F scores are keyed here so that higher scores are more masculine.
*$p < .05$. **$p < .01$. ***$p < .001$.

Was the Strong Subsample Comparable to the Larger Sample?

To check whether the subsample of the Terman participants we studied was comparable to the larger Terman sample of women and men, we compared the mean personality scores of men and women in the subsample with the corresponding means of men and women in the total sample (see Table 2). In general, this comparison shows that the subsample was quite similar to the larger sample. Indeed, *t* tests indicated that none of the Big Five factors showed significant differences between the subsample and members of the larger population not in the subsample. Furthermore, patterns of intercorrelation among the personality variables were very similar in the subsamples of men and women and in the larger samples of men and women. Because the subsample of women who took the Strong test was considerably smaller than the total sample of women, we examined whether these women differed from women who did not complete the Strong on the following additional variables: adult adjustment, self-reported health, smoking, alcohol consumption, and obesity (body mass index). *t* tests indicated no significant differences between groups.

TABLE 2: Mean Big Five Scores for Men and Women in the Subsample and in the Larger Terman Population

	Men		*Women*	
Big Five Factor	*Subsample* (N = *612)*	*Full Population* (N = *693)*	*Subsample* (N = *203)*	*Full Population* (N = *542)*
Extraversion	2.96	2.96	3.07	3.05
Agreeableness	2.90	2.91	3.14	3.12
Conscientiousness	3.07	3.06	2.97	2.92
Neuroticism	3.06	3.07	2.92	2.91

Analyses of Personality and Mortality

Statistical methods. The main statistical method we used to investigate links between personality and mortality was hazard regression analysis, a form of survival analysis (for detailed accounts of the application of such analyses to the Terman data, see Friedman, Tucker, Schwartz, Martin, et al., 1995; Friedman et al., 1993; for an introduction to survival analysis geared to psychologists, see Morita, Lee, & Mowday, 1989). In brief, hazard analyses estimate the probability of death as a function of age. This relationship is expressed through a hazard function. Hazard analyses can estimate how the hazard function is influenced by various predictor variables.

Cox's widely used (nonparametric) proportional hazards model makes no assumption about the functional form of the underlying hazard function but does assume that the effect of each predictor variable is multiplicative and constant across age. In the context of the current study, the Cox model would assume that the effect of personality predictors is multiplicative (i.e., personality level increases or decreases death rate by some multiplicative factor) and constant across age (the increase or decrease in death rate due to a given personality factor is the same at all ages). In contrast, the Gompertz (parametric) model makes the more limiting assumption that the underlying hazard function is exponential, but this restriction is offset by the model's ability to test other questions, for example, whether the effects of predictors vary depending on age. We used both the Cox and Gompertz methods, as appropriate, in the following analyses.

Because the various personality variables studied here were scaled quite differently, we present the effects of personality on mortality in terms of the interquartile relative hazard (rh)—the computed ratio of mortality risk for individuals at the 25th versus 75th percentile of a given personality dimension.

Relations between within-sex, gender-related individual differences and mortality. Previous analyses of the Terman data have shown expected gender differences in

TABLE 3: Cox Proportional Hazards Models Predicting Mortality From GD and From Strong M-F, Controlling for Extraversion, Agreeableness, Conscientiousness, and Neuroticism

	Predicting Mortality From GD		*Predicting Mortality From Strong M-F*	
Predictor Variables	*Males (N = 612)*	*Females (N = 202)*	*Males (N = 604)*	*Females (N = 205)*
Extraversion				
β	.03	–.12	.06	–.12
rh	1.03	.89	1.06	.89
Agreeableness				
β	–.17†	–.14	–.19*	–.09
rh	.85	.87	.83	.92
Conscientiousness				
β	–.17*	–.25	–.14*	–.28
rh	.84	.78	.87	.75
Neuroticism				
β	.09	.02	.09	.11
rh	1.09	1.02	1.09	1.12
GD				
β	.23**	.48**		
rh	1.26	1.61		
Strong M-F				
β			.24**	.31**
rh			1.26	1.36

NOTE: GD = gender diagnosticity, rh = relative hazard.
†$p < .1$. *$p < .05$. **$p < .01$.

mortality, with men 48% more likely to die than women at any given age (Friedman et al., 1993). Accordingly, Cox survival analyses were conducted on men and women separately to determine if there were relationships between gender-related individual differences (Strong M-F, GD) and mortality. For women, Strong M-F was significantly related to mortality ($p < .01$, $n = 205$, interquartile rh = 1.36). GD also was significantly related to mortality among women ($p < .01$, $n = 203$, interquartile rh = 1.62). For men, too, Strong M-F was significantly related to mortality ($p < .01$, $n = 639$, interquartile rh = 1.26). GD also was related to men's mortality rates ($p < .01$, $n = 612$, interquartile rh = 1.25).

Hazard analyses that controlled for the four assessed Big Five factors (Extraversion, Agreeableness, Conscientiousness, and Neuroticism) found that GD and Strong M-F continued to be significantly associated with men's and women's mortality and that their associated relative hazards remained virtually unchanged (see Table 3 for rhs and beta weights).

Gender-related individual differences and risk factors. In an attempt to identify possible risk factors mediating observed links between masculinity (as measured by Strong M-F and GD) and mortality, we computed correlations between gender-related individual differences and assessed risk factors, and we also computed correlations between gender-related individual differences and two variables that served as rough indices of socioeconomic status—earned income in 1946 and cumulative level of education achieved by 1950 (both self-reports).[5] Table 4 presents these correlations, computed separately for men and women. Because of missing data for some measures, reported *N*s are sometimes considerably smaller than those reported in the previous hazard analyses.

In general, the correlations in Table 4 suggest at best modest relationships between gender-related individual differences, risk factors, and socioeconomic variables. For men, there were small but significant relationships between GD and body mass, physical activity in 1950, riskiness of hobbies, and earned income, and similarly, there were relationships between Strong M-F and body mass, physical activity in 1940 and 1950, riskiness of hobbies, and education. For women, there were significant correlations between GD, Strong M-F, and earned income (but in a much reduced sample, which probably resulted from the fact that by the mid- to late-1940s, most of the Terman women did not engage in income-producing work). Women's Strong M-F also showed a weak correlation with alcohol consumption in 1950.

Follow-up hazard analyses exploring possible mediators. As just reported, men's body mass, activity in 1950, riskiness of hobbies, and income measures were slightly but reliably correlated with GD. Accordingly, we conducted hazard analyses for GD in which these risk variables were entered as controls. This had only a small effect on the hazard associated with GD (change in rh from 1.35, $p < .01$, to 1.40, $p < .01$).

As also noted before, body mass, physical activity in 1940 and 1950, riskiness of hobbies, and level of education were significantly correlated with men's Strong M-F scores. However, physical activity levels in 1940 and 1950 were not risk factors for men in this subsample, and although level of education was slightly protective for men (rh = .94, $p < .05$), entering this variable and the other four risk variables as controls did not substantially change the risk associated with Strong M-F (change in rh from 1.41, $p < .01$, to 1.38, $p < .05$).

For women, GD and Strong M-F were correlated with earned income, but because of missing data and a reduced sample size, this variable could not usefully serve as a potential mediator. Alcohol use in 1950 was a risk factor for women in our subsample ($n = 183$, rh = 1.87, $p < .01$) and correlated significantly with Strong M-F. Controlling for this slightly decreased the risk associated with Strong M-F (change in rh from 1.59, $p < .01$, to 1.49, $p < .01$).

In sum, controlling, when possible, for risk and demographic factors that were associated with GD and Strong M-F did not have major effects on the mortality risk associated with them.

TABLE 4: Correlations Between GD, Strong M-F, and Health Risk Factors

	Men		*Women*	
Risk Factor	*GD*	*Strong M-F*	*GD*	*Strong M-F*
Smoking (pack years)	.06 (356)	.00 (350)	.02 (125)	.05 (120)
Alcohol consumption (1950)	.02 (589)	–.05 (578)	.08 (188)	.15* (183)
Alcohol consumption (1960)	.03 (515)	–.07 (504)	.04 (172)	.04 (168)
Obesity (body mass index)	.12** (609)	.11** (601)	.09 (202)	.03 (197)
Physical activity (1940)	.08† (544)	.10* (536)	.05 (168)	.07 (164)
Physical activity (1950)	.10* (592)	.11** (581)	.09 (189)	–.06 (184)
Physical activity (1960)	.05 (586)	.06 (574)	–.03 (184)	–.03 (179)
Riskiness of hobbies	.09* (654)	.08* (639)	.08 (209)	.05 (205)
Earned income (1946)	.09* (545)	–.01 (535)	.29** (86)	.33** (86)
Level of education (1950)	–.04 (582)	–.16*** (571)	.12 (187)	.05 (182)

NOTE: Numbers in parentheses give *N*s for correlations. GD = gender diagnosticity, M-F = masculinity-femininity. Strong M-F is keyed so higher scores are more masculine.
†$p < .1$. *$p < .05$. **$p < .01$. ***$p < .001$.

Causes of death analyses. Because GD was associated with increased risk of death in both women and men, we conducted Gompertz hazard analyses to determine whether GD was predictive of death particularly from certain causes or whether GD was equally predictive of all causes of death. The causes of death considered were cardiovascular disease, cancer, accident/injury, other, and unknown. Specifically, we created three Gompertz models and then compared their goodness of fit. The first model predicted mortality from sex only. The second model predicted from sex and GD while holding the effects of GD constant across all causes of death. The third model also predicted from sex and GD but allowed GD to predict differentially to various causes of death. The results of these analyses are presented in Table 5, as well as chi-square tests for comparisons of models. These model comparisons also are distributed as chi-square, with their degrees of freedom equal to the differences in the degrees of freedom between the compared models.

As can be seen from Table 5, both models that included GD as a predictor of mortality (Models 2 and 3) performed better than the first model, which included only sex as a predictor. However, the difference in chi-square between Models 2 and 3 was not significant, indicating that GD did not display a significant differential prediction to cause of death.

Similarly, we conducted three Gompertz analyses to determine whether Strong M-F was predictive of death particularly from certain causes or whether Strong M-F was equally predictive of death from all causes. Because the original Strong M-F scores were keyed in opposite directions for men and women, we converted these scores to z scores, separately for men and women, and reversed the signs of women's z scores. The results of these analyses are presented in Table 6. As was the case for the corresponding analyses for GD, analyses indicated that Strong M-F did not display a differential prediction to cause of death.

DISCUSSION

Within-sex, gender-related individual differences, as assessed by GD measures and Strong M-F, were significantly predictive of mortality risk. For both men and women, individuals who were more male-typical in their occupational preferences tended to show higher mortality rates than did individuals who were more female-typical. These associations were not due to a specific cause of death.

The links between gender-related individual differences (GD and Strong M-F) and mortality were substantial and remained when four Big Five traits were entered as control variables. As described earlier, an association between childhood Conscientiousness and adult mortality has been documented in previous studies (Friedman, Tucker, Schwartz, Martin, et al., 1995; Friedman et al., 1993). The link between Agreeableness and mortality has been less well documented but work in progress (Martin, 1996; Martin & Friedman, 2000a) adds to previous evidence linking Type A behavior and negative masculinity (e.g., disagreeable behaviors and dispositions) to illness and mortality (Helgeson, 1990, 1995).[6] Indeed, many of the 1940 personality items used to assess Agreeableness (e.g., "Do you lose your temper easily?" "With the opposite sex do you tend to be dominant and to have your own way?" "Do you consider yourself as vain and egotistic?") seem to overlap strongly with components of negative masculinity.

As noted before, Agreeableness overlaps with feminine expressiveness and components of Extraversion overlap with masculine instrumentality. Thus, the fact that GD and Strong M-F showed significant links to mortality, even when controlling for Agreeableness and

TABLE 5: Three Gompertz Models Testing Whether GD Is Equally Predictive of All Causes of Death

	Chi-Square	df	p
Model 1			
(Mortality predicted from gender only)	566.66	10	<.001
Model 2			
(Mortality predicted from gender and GD, with effect of GD constant across causes of death)	581.30	11	<.001
Model 3			
(Mortality predicted from gender and GD, with GD allowed to predict differentially across causes of death)	587.14	15	<.001
Comparisons of models			
Model 2 vs. Model 1	14.64	1	<.01
Model 3 vs. Model 1	20.48	5	<.01
Model 3 vs. Model 2	5.84	4	*ns*

NOTE: GD = gender diagnosticity.

TABLE 6: Three Gompertz Models Testing Whether Strong M-F Is Equally Predictive of All Causes of Death

	Chi-Square	df	p
Model 1			
(Mortality predicted from gender only)	563.57	10	<.001
Model 2			
(Mortality predicted from gender and Strong M-F with effect of Strong M-F constant across causes of death)	580.46	11	<.001
Model 3			
(Mortality predicted from gender and GD, with Strong M-F allowed to predict differentially across causes of death)	583.95	15	<.001
Comparisons of models			
Model 2 vs. Model 1	16.89	1	<.001
Model 3 vs. Model 1	20.38	5	<.01
Model 3 vs. Model 2	3.49	4	*ns*

NOTE: M-F = masculinity-femininity, GD = gender diagnosticity.

Extraversion, suggests that these links are independent of any relationships that might exist between instrumentality, expressiveness, and mortality risk. One limitation of the current results is that personality scale reliabilities were not as high as might be achieved without the constraints of archival data. Nonetheless, reliabilities were clearly high enough to demonstrate links between some traits (e.g., GD, Conscientiousness, Agreeableness) and mortality risk.

The relationships between gender-related personality traits (GD and Strong M-F) and mortality documented here are comparable in magnitude with associations documented for physiological risk factors such as systolic blood pressure and serum cholesterol levels (see Barrett-Connor, Suarez, Khaw, Criqui, & Wingard, 1984) and for high versus low levels of social support (see Berkman & Syme, 1979; House, Robbins, & Metzner, 1982; Schoenbach, Kaplan, Fredman, & Kleinbaum, 1986). For women, the association between masculinity and mortality was as strong as the association between gender and mortality for the sexes combined.

To illustrate graphically the effects of GD and gender on mortality, we used Gompertz hazard analyses to draw survival curves, separately for men and women, that show the predicted probability of a low- and high-GD 20-year-old surviving to a given age (see Figure 1). These curves clearly show that the relative risk of high versus low levels of GD was comparable in magnitude to the relative risk of gender, particularly for women. Furthermore, Figure 1 makes obvious a conclusion that is implicit in our results—the most advantaged group in terms of longevity consisted of feminine women, and the

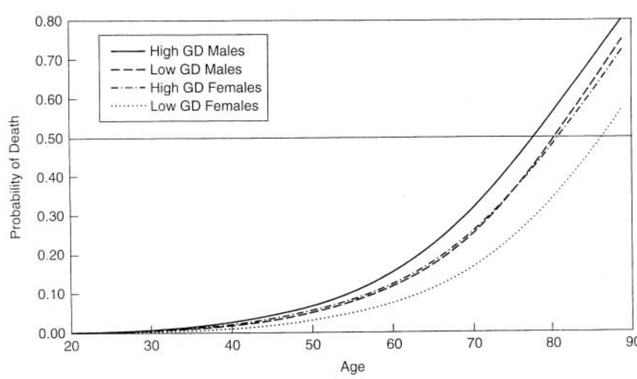

Figure 1 Probability of a 20-year-old dying by a given age, by sex and gender diagnosticity.

most disadvantaged group consisted of masculine men. Figure 1 therefore suggests that the gender difference in mortality observed in the overall Terman sample may be due particularly to differences in mortality between masculine men and feminine women.

In general, our analyses of risk factors were unable to identify the mediating routes by which GD and Strong M-F came to be associated with mortality. However, these null findings should be accompanied by several caveats. First, reduced sample sizes (particularly for women) and missing values for many potential mediating variables limited the power of our mediational analyses. Second, some of the mediating variables studied here may have suffered from unreliability, restrictions in range, or selective attrition in the Terman sample. One of the risk variables—physical activity—was dichotomous, and this undoubtedly reduced its correlation with gender-related traits. Finally, not all plausible mediating variables (e.g., social support) were considered in our analyses.

The four factors we listed at the start to explain links between gender and mortality—biology, gender roles, behavioral risks, and personality—seem prime candidates for understanding the relationship between within-sex, gender-related individual differences and mortality. There may be biological third variables, for example, that influence both gender-related individual differences and mortality. Some have hypothesized, to offer a speculative conjecture, that sex hormone levels may be related both to temperament and longevity (Nyborg, 1994). Despite the possibility that biological third variables influence both masculinity and mortality, we suspect that mediating links are more likely to be found in gendered roles, gendered behaviors, and their accompanying risk factors. For example, masculine individuals may cultivate different kinds and different degrees of social support than feminine individuals. Masculine people may, in certain circumstances, engage in behaviors (e.g., driving fast, smoking, drinking, eating poorly) that lead to poorer health and higher mortality rates and they may seek out medical care differently from feminine individuals and maintain medical regimes differently.[7] All of these possibilities seem worthy of further research attention given the strength of the current findings. Because gender itself is linked to a variety of health risk factors, it may not be surprising if within-sex, gender-related individual differences similarly prove to be linked to a variety of health risk factors.

It is important to note in conclusion that the Terman data represent a sample of people from a particular era in history. The Terman women, for example, came of age during the Great Depression and embarked on their careers as wives, mothers, and sometimes workers as the world faced the cataclysm of World War II. The modern women's movement would not emerge as a potent social force until the Terman women were entering their seventh decade of life. Men's roles have undoubtedly also changed over time. Certainly, then, it would be interesting to study if masculinity has similar ties to mortality in contemporary cohorts.

Despite their unique historical context, the Terman data yield strong, if preliminary, evidence that gender-related traits can show a powerful relationship to mortality risk. In fact, the relative homogeneity of Terman participants on ethnicity and intelligence suggests that typically robust societal influences on mortality risk are not able to account for the current findings. Although we do not yet know why, we can conclude with some confidence that masculinity—in both women and men—can indeed be hazardous to your health.

APPENDIX

Correlations of Gender With Rated Preferences for Strong Occupation Items

.33***	Manufacturer
.31***	Mechanical engineer
.28***	Factory manager
.26***	Governor of a state
.23***	Author of a technical book
.22***	Inventor
.21***	Aviator/aviatrix
.20***	Judge
.20***	Chemist
.20***	Wholesaler
.19***	Athletic director
.19***	Stockbroker
.18***	Lawyer, corporation
.17***	Scientific research worker
.15***	Statistician
.15***	Politician
.13***	Factory worker
.13***	Life insurance salesman
.13***	Lawyer, criminal
.12***	Dentist
.10**	Laboratory technician
.10**	Surgeon
.10**	College professor
.09**	Sales manager
.08*	Secret service man/woman
.08*	Farmer
.08*	Traveling salesman
.06	Real estate salesman
.06	Draftsman
.05	Cartoonist
.04	Physician
.03	Architect
.03	Pharmacist
.03	Reporter, general
.02	Office manager
.02	Retailer
.01	Editor
.01	Bank teller
–.01	Hotel keeper or manager
–.01	Playground director
–.02	Foreign correspondent
–.02	Advertiser
–.03	Bookkeeper
–.09*	Civil service employee
–.10**	Author of novel
–.104**	Magazine writer
–.13**	Artist
–.13***	Sculptor
–.14***	Employment manager
–.14***	Specialty salesman/woman

–.14***	Musician
–.15***	Buyer of merchandise
–.15***	Landscape gardener
–.16***	Typist
–.17***	Office clerk
–.20***	Poet
–.21***	Librarian
–.23***	Interpreter
–.28***	Social worker
–.30***	Private secretary
–.41***	Florist
–.42***	Interior decorator

NOTE: Positive correlations occur when men prefer occupation more than do women. Negative correlations occur when women prefer occupations more than do men.

*$p < .05$. $p < .01$. $p < .001$.

NOTES

1. Smoking effects in the Terman data are likely underestimated because participants who smoked very heavily and died very young are more likely to be missing from the sample.

2. In essence, discriminant analyses weight highly those items that strongly correlate with gender. In general, correlations between gender and occupational preferences for Terman participants (which are listed in the appendix) were similar to correlations found in contemporary samples of men and women. Men, on average, were more interested than women in "thing-oriented," realistic occupations, whereas women were more interested than men in "people-oriented," artistic and social occupations (see Lippa, 1998b). These results suggest that gender diagnosticity (GD) assessed much the same dimension in the Terman sample as it does in contemporary populations.

3. In these and subsequent analyses, sample sizes sometimes vary because of missing values for various personality and risk factor variables.

4. Despite the substantial correlation between GD and Strong masculinity-femininity (M-F), it is worth noting that these two measures are conceptually distinct. Items of the Strong M-F scales were chosen because they discriminated men and women in prior normative samples. In contrast, items were weighted highly in GD measures because they discriminated men and women in the Terman sample. Subsequent analyses show that for women, GD showed a stronger relationship to mortality than did Strong M-F.

5. The Terman data codebook indicates that many participants reported sources of income (such as room and board and inheritances), which were not counted as earned income. As a result, earned income may be an unreliable measure of socioeconomic status for some Terman participants.

6. On the surface, findings of a negative relationship between Agreeableness and mortality might appear inconsistent with previous demonstrations of a positive relationship between childhood cheerfulness and mortality. However, Agreeableness and cheerfulness assess quite different constructs. Childhood cheerfulness was assessed from parent and teacher ratings of cheerfulness/optimism and sense of humor. Cheerfulness was a risk factor for mortality (Friedman et al., 1993), possibly because the children so identified included fun/thrill seekers and jokers. Adult Agreeableness was based on self-ratings and closely approximates Agreeableness as assessed by contemporary Big Five scales. It is this latter construct that is positively linked to longevity. Childhood cheerfulness was not significantly correlated with adult Agreeableness in the Terman sample.

7. Recent unpublished data show that young men who suffer brain injuries, largely due to traumatic accidents, are higher on GD than are control men without brain injuries (Shears, 1996) and that GD is associated with self-reported thrill seeking, risky driving, and alcohol consumption in young men (Arad, 1998). Both of these findings point to a cluster of health risk factors associated with masculinity in young men. They also raise the possibility that the health risks associated with masculinity may vary somewhat over the life span.

REFERENCES

Arad, S. (1998). *The Big Five, gender-related traits, and health risk factors.* Master's thesis, California State University, Fullerton.

Ashmore, R. D. (1990). Sex, gender, and the individual. In L. A. Pervin (Ed.), *Handbook of personality: Theory and research* (pp. 426-486). New York: Guilford.

Barrett-Connor, E., Suarez, L., Khaw, K. T., Criqui, M. H., & Wingard, D. L. (1984). Ischemic heart disease risk factors after age 50. *Journal of Chronic Disease, 37*, 903-908.

Bem, S. L. (1974). The measurement of psychological androgyny. *Journal of Consulting and Clinical Psychology, 42*, 165-172.

Bem, S. L. (1981). *Bem Sex-Role Inventory professional manual.* Palo Alto, CA: Consulting Psychologists Press.

Berkman, L. F., & Syme, S. L. (1979). Social networks, host resistance, and mortality: A nine-year follow-up of Alameda County residents. *American Journal of Epidemiology, 109*, 186-204.

Bernreuter, R. G. (1933). The theory and construction of the Personality Inventory. *Journal of Social Psychology, 4*, 387-405.

Block, J. H. (1973). Conceptions of sex roles: Some cross-cultural and longitudinal perspectives. *American Psychologist, 28*, 512-526.

Campbell, D. P. (1971). *Handbook for the Strong Vocational Interest Blank.* Stanford, CA: Stanford University Press.

Constantinople, A. (1973). Masculinity-femininity: An exception to a famous dictum. *Psychological Bulletin, 80*, 389-407.

Cook, E. P. (1985). *Psychological androgyny.* New York: Pergamon.

Costa, P. T., Jr., & McCrae, R. R. (1992). *The revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual.* Odessa, FL: Psychological Assessment Resources.

Friedman, H. S., & Booth-Kewley, S. (1987). The "disease-prone" personality: A meta-analytic view of the construct. *American Psychologist, 42*, 539-555.

Friedman, H. S., Tucker, J. S., Schwartz, H. E., Martin, L. R., Tomlinson-Keasey, C., Wingard, D. L., & Criqui, M. H. (1995). Childhood conscientiousness and longevity: Health behaviors and cause of death. *Journal of Personality and Social Psychology, 68*, 696-703.

Friedman, H. S., Tucker, J. S., Schwartz, J. E., Tomlinson-Keasey, M.L.R., Wingard, D. L., & Criqui, M. H. (1995). Psychosocial and behavior predictors of longevity: The aging and death of the "Termites." *American Psychologist, 50*, 69-78.

Friedman, H. S., Tucker, J. S., Tomlinson-Keasey, C., Schwartz, J. E., Wingard, D. L., & Criqui, M. H. (1993). Does childhood personality predict longevity? *Journal of Personality and Social Psychology, 65*, 176-185.

Hazzard, W. R. (1990). A central role of sex hormones in the sex differential in lipoprotein metabolism, atherosclerosis, and longevity. In M. G. Ory & H. R. Warner (Eds.), *Gender, health, and longevity: Multidisciplinary perspectives* (pp. 87-108). New York: Springer.

Helgeson, V. S. (1990). The role of masculinity in a prognostic predictor of heart attack severity. *Sex Roles, 22*, 755-774.

Helgeson, V. S. (1994). Relation of agency and communion to well-being: Evidence and potential explanations. *Psychological Bulletin, 116*, 412-428.

Helgeson, V. S. (1995). Masculinity, men's roles, and coronary heart disease. In D. Sabo & D. F. Gordon (Eds.), *Men's health and illness: Gender, power, and the body* (pp. 68-104). Thousand Oaks, CA: Sage.

Holland, J. L. (1992). *Making vocation choices* (2nd ed.). Odessa, FL: Psychological Assessment Resources.

House, J. S., Robbins, C., & Metzner, H. L. (1982). The association of social relationships and activities with mortality: Prospective evidence from the Tecumseh community health study. *American Journal of Epidemiology, 116*, 123-140.

Lenney, E. (1991). Sex roles: The measurement of masculinity, femininity, and androgyny. In J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), *Measures of personality and social psychological attitudes* (Vol. 1, pp. 573-660). San Diego, CA: Academic Press.

Lippa, R. (1991). Some psychometric characteristics of gender diagnosticity measures: Reliability, validity, consistency across domains and relationship to the Big Five. *Journal of Personality and Social Psychology, 61*, 1000-1011.

Lippa, R. (1995a). Do sex differences define gender-related individual differences within the sexes? Evidence from three studies. *Personality and Social Psychology Bulletin, 21*, 349-355.

Lippa, R. (1995b). Gender-related individual differences and psychological adjustment in terms of the Big Five and circumplex models. *Journal of Personality and Social Psychology, 69*, 1184-1202.

Lippa, R. (1997). The display of masculinity, femininity, and gender diagnosticity in self-descriptive photo essays. *Journal of Personality, 65*, 137-169.

Lippa, R. (1998a). Gender-related individual differences and National Merit test performance: Girls who are "masculine" and boys who are "feminine" tend to do better. In L. Ellis & L. Ebertz (Eds.), *Males, females, and behavior: Toward biological understanding*. New York: Praeger.

Lippa, R. (1998b). Gender-related individual difference and the structure of vocational interests: The importance of the "People-Things" dimension. *Journal of Personality and Social Psychology, 74*, 996-1009.

Lippa, R. (1998c). The nonverbal judgment and display of extraversion, masculinity, femininity, and gender diagnosticity: A lens model analysis. *Journal of Research in Personality, 32*, 80-107.

Lippa, R. (1999a). *Gender-related traits in gay men, lesbian women, and heterosexual men and women: The virtual identity of homosexual-heterosexual diagnosticity and gender diagnosticity*. Unpublished manuscript, California State University, Fullerton.

Lippa, R. A. (in press). *On deconstructing and reconstructing masculinity-femininity. Journal of Research in Personality*.

Lippa, R., & Connelly, S. C. (1990). Gender diagnosticity: A new Bayesian approach to gender-related individual differences. *Journal of Personality and Social Psychology, 59*, 1051-1065.

Martin, L. R. (1996). *Consonance of archival and contemporary data: A comparison of personality scales reconstructed from the Terman data set with modern personality scales*. Doctoral dissertation, University of California, Riverside.

Martin, L. R., & Friedman, H. S. (2000a). *Personality predictors of longevity*. Unpublished manuscript.

Martin, L. R., & Friedman, H. S. (2000b). Comparing personality scales across time: An illustrative study of validity and consistency in a large archival data set. *Journal of Personality, 68*, 85-110.

Morita, J. G., Lee, T. W., & Mowday, R. T. (1989). Introducing survival analysis to organizational researchers: A selected application to turnover research. *Journal of Applied Psychology, 74*, 280-292.

Nathanson, C. A. (1984). Sex differences in mortality. *Annual Review of Sociology, 10*, 191-213.

Niven, C. A., & Carroll, D. (1993). *The health psychology of women*. Langhorne, PA: Harwood.

Nyborg, H. (1994). *Hormones, sex, and society: The science of physicology*. Westport, CT: Praeger.

Prediger, D. J. (1982). Dimensions underlying Holland's hexagon: Missing link between interests and occupations? *Journal of Vocational Behavior, 21*, 259-287.

Schoenbach, V. J., Kaplan, B. H., Fredman, L., & Kleinbaum, D. (1986). Social ties and mortality in Evans County, Georgia. *American Journal of Epidemiology, 123*, 577-591.

Shears, C. (1996). *Acquired brain injury and its relationship to gender-related traits*. Master's thesis, California State University, Fullerton.

Spence, J. T., & Buckner, C. (1995). Masculinity and femininity: Defining the undefinable. In P. J. Kalbfleisch & M. J. Cody (Eds.), *Gender, power, and communication in human relationships* (pp. 105-138). Hillsdale, NJ: Lawrence Erlbaum.

Spence, J. T., & Helmreich, R. L. (1978). *Masculinity and femininity: Their psychological dimensions, correlates, and antecedents*. Austin: University of Texas Press.

Spence, J. T., & Helmreich, R. L. (1980). Masculine instrumentality and feminine expressiveness: Their relationships with sex role attitudes and behaviors. *Psychology of Women Quarterly, 5*, 147-163.

Spence, J. T., Helmreich, R. L., & Holahan, C. K. (1979). Negative and positive components of psychological masculinity and femininity and their relationship to self-reports of neurotic and acting out behaviors. *Journal of Personality and Social Psychology, 37*, 1673-1682.

Spence, J. T., Helmreich, R. L., & Stapp, J. (1974). The Personal Attributes Questionnaire: A measure of sex role stereotypes and masculinity-femininity. *JSAS: Catalog of Selected Documents in Psychology, 4*, 43-44. (No. 617)

Strickland, B. R. (1988). Sex-related differences in health and illness. *Psychology of Women Quarterly, 12*, 381-399.

Strong, E. K. (1943). *Vocational interests of men and women*. Stanford, CA: Stanford University Press.

Terman, L. M., & Miles, C. C. (1936). *Sex and personality: Studies in masculinity and femininity*. New York: McGraw-Hill.

Terman, L. M., & Oden, M. H. (1947). *Genetic studies of genius: The gifted child grows up* (Vol. 4). Stanford, CA: Stanford University Press.

U.S. Department of Health and Human Services. (1980). *International classification of diseases* (9th revision, clinical modification, 2nd ed. DHHS Publication No. PHS 80-1260). Washington, DC: Government Printing Office.

Verbrugge, L. M. (1985). Gender and health: An update on hypotheses and evidence. *Journal of Health and Social Behavior, 26*, 156-182.

Waldron, I. (1995a). Contributions of biological and behavioural factors to changing sex differences in ischemic heart disease mortality. In A. Lopez, G. Caselli, & T. Valkonen (Eds.), *Adult mortality in developed countries* (pp. 161-178). New York: Oxford University Press.

Waldron, I. (1995b). Contributions of changing gender differences in behavior and social roles to changing gender differences in mortality. In D. Sabo & D. F. Gordon (Eds.), *Men's health and illness: Gender, power, and the body* (pp. 22-45). Thousand Oaks, CA: Sage.

Williams, J. E., & Best, D. L. (1990). *Measuring sex stereotypes: A multination study*. Newbury Park, CA: Sage.

Wingard, D. L. (1984). The sex differential in morbidity, mortality, and lifestyle. *Annual Review of Public Health, 5*, 433-458.

Received December 14, 1998

Revision accepted October 31, 1999