Examining the Effects of Differential Coverage of Three Mortality Record Sources Analyses from Project Talent

Jizhi Zhang[1] & Jamie Kim[2], Sandy Eyster[1], Michael Cohen[1]

[1]American Institutes for Research, 1990 K Street, NW, Washington, DC 20006
[2] Department of Public Health and Preventive Medicine, University of Toledo, Toledo, OH 43606

Abstract

Key to research on mortality is determining mortality status via linking to administrative records. Several methods for ascertaining mortality status exist, including the Social Security Administration's Death Master File (DMF), the National Center for Health Statistics' National Death Index (NDI), and consumer and commercial credit bureau databases. This study examines the effect of differential coverage of three mortality record sources for a random subsample of 4,159 Project Talent participants. Project Talent is a longitudinal study that began in 1960 and measured the cognitive abilities, interests, personality, and demographics of approximately 440,000 9th-12th grade students. In this study we summarize findings on coverage rates, mortality rates, and potential bias related to the different mortality record sources. Furthermore, we investigate whether and to what extent relying on different mortality record sources affects the results of survival analyses. In addition to key demographics, we examine the relationship among early life personality and cognitive ability factors and mortality, which are rarely available for a large, nationally-representative sample.

Key Words: administrative records, mortality, survival analysis, bias

1. Introduction

Accurate determination of mortality status is integral to studying mortality. Investigators must be able to accurately determine study participants' current status based on previously gathered data. There are a variety of vital record databases available to ascertain status. Some are easily accessible on the internet and are freely available under the Freedom of Information Act while others are available for a fee. Some of the more widely used sources include the Social Security Administration's Death Master File (DMF), National Death Index (NDI) administered by the National Center for Health Statistics (NCHS), and Lexis-Nexis (LN), a consumer and commercial credit bureau database.

Because of the differences in coverage rates, costs, and time involved, some sources may be better suited than others. The DMF contains over 89 million death records that have been reported to the Social Security Administration (National Technical Information Service, 2012) since 1936. Information is updated weekly. The NDI is a centralized computer index of death record information submitted by state vital records offices (Office of Information Services, 2011). The NCHS updates the NDI once a year. Data collection started in 1979, and currently, the NDI covers years 1979-2009. LN provides an "Identity Verification and Authentication" (Lexis-Nexis, 2012) service utilizing a proprietary search formula that searches commercial and credit databases.

This study examines the effects of different coverage rates by three mortality record sources using a random subsample of 4,159 Project Talent (PT) participants. Project Talent is a nationally representative longitudinal study of high school students in 1960 that collected extensive cognitive, personality and background information from 400,000 9^{th}-12^{th} graders. Follow-ups were conducted at 1, 5, and 11 years after each grade cohort graduated high school (at approximate ages 19, 23, and 30, respectively). PT sample members are now 66 to 70 years old. The purpose of this study is to investigate potential bias related to coverage rates by different mortality record sources. It reviews whether and to what extent relying on different mortality record sources affects the results of analyses conducted to predict the likelihood of mortality based on a number of factors. In addition to key demographic variables, we examine early life personality and cognitive ability measures that are rarely available from a large, nationally-representative sample of individuals. Previous research studies comparing different mortality records sources have focused primarily on coverage, cost, efficiency, and other logistics issues. Our study intends to provide insight for future research on mortality by elucidating the potential bias present in mortality records sources within the context of predicting the likelihood of mortality.

1.1 Mortality Records Sources

There are few sources of centralized and publicly accessible mortality information in the United States. Investigators often turn to either, the Social Security Administration's (SSA) DMF or the NDI (Schnorr & Steenland, 1997). There are other record sources such as Veteran's Administration data or pension information from private corporations, but they only contain records pertaining to specific cohorts. Another source of mortality records are private companies such as Equifax or LN. LN is a consumer and commercial credit bureau database. It provides a number of services including records matching to a variety of databases. This study examines the differences between three sources of mortality records: the DMF, NDI, and LN.

The data available in each record source varies, resulting in the possibility of differential results when comparing mortality from different sources. The DMF is a database only, and does not offer matching services. It is up to the investigator to match and verify mortality records. LN and the NDI utilize their own matching algorithms and report the results of matching to the investigator. Single or multiple criteria may be used in matching records. For example, the NDI has a list of 7 criteria or matching conditions of which, at least one must be satisfied to make a positive match (National Center for Health Statistics, 2011). Such criteria can range from a single social security number to a combination of date of birth with first and last name. The sensitivity of each source varies due to the different coverage and matching techniques. Sensitivity is defined as the number of correctly identified decedents (Cowper, Kubal, Maynard, & Hynes, 2002).

The DMF was created by the Social Security Administration in 1988, and provides weekly updates on the vital status of all individuals enrolled in the Social Security program since 1936. Information contained in the DMF includes name, Social Security number, last known residence, dates of birth and death, and the state where the individual first enrolled for a social security card (Hill & Rosenwaike, 2002). There are a range of values reported for the sensitivity of the DMF. Schisterman and Whitcomb (2004) reported a sensitivity of 92.2% for American born men and women. Schall, Buchanich, Marsh, & Bittner (2001) reported an overall sensitivity of 83%. The DMF has proven utility, but it is not without its limitations. Reporting deaths to the SSA is voluntary, and since the data are collected for purposes other than maintaining vital status records (e.g., maintaining records of payments to social security beneficiaries), researchers have investigated the completeness of records over time and for different

groups of individuals (Schnorr & Steenland, 1997). Several studies have reported that deaths are underestimated in the DMF. Ninety-five percent of individuals 65 and older are in the DMF, but it is only 74% complete for individuals 25-54 years of age and only 42% complete for decedents under the age of 25 (Social Security Administration; Office of the Inspector General, 2003). Other studies (Cowper, Kubal, Maynard, & Hynes, 2002; Hill & Rosenwaike, 2002; Schnorr & Steenland, 1997) have suggested that many deaths were not recorded in SSA files that existed prior to the creation of the DMF, thereby decreasing the completeness of the pre-1988 death records. Additionally, an audit by the SSA's Inspector General found about 1.3 million deaths recorded in their Master Beneficiary Record that should have been included in the DMF (Hill & Rosenwaike, 2002).

The NDI is administered by the Center for Disease Control and Prevention's (CDC) NCHS. NCHS contracts with state health departments to collect mortality data. Reporting to the NDI is mandatory and covers everyone in the United States, Puerto Rico, and the U. S. Virgin Islands. It is available to researchers "solely for statistical purposes in medical and health research" (Office of Information Services, 2011). Those requesting records must complete an application process to verify that the information will be used appropriately. Matches are made between a possible record match and a particular user record, by a combination of the following: first and last name, middle initial, father's surname, social security number, month, day, and year of birth, race, sex, marital status, state of residence, and state of birth. Sensitivity for the NDI is high and it is considered the "gold standard" of mortality ascertainment (Cowper, Kubal, Maynard, & Hynes, 2002). Stampfer et al. (1984) reported a sensitivity of 96.5%. Fisher et al. (1995) reported 97.9% sensitivity. One main limitation of the NDI is that data collection started in 1979; thus, the NDI does not include deaths prior to 1979. Updates are made annually. Currently, data are available only until 2009 (Office of Information Services, 2011).

LN is a consumer and credit bureau information database. It provides a number of identity verification services. LN Batch Collection Solutions utilizes independent data sources comprised of more than four billion records. LN aggregates and reports data, as provided by public records and commercially available data sources (LexisNexis, 2011). Matching algorithms are proprietary.

1.2 The Relationship among Cognitive Ability, Personality, and Mortality

Cognitive ability and personality traits have an effect on mortality risk. The literature shows that individuals with higher IQ scores have lower mortality rates (Batty & Deary, 2004). The mechanism through which childhood IQ affects mortality is not well defined. A number of hypotheses are offered by Deary (2005) to explain the link. He suggests the following as IQ dependent mediators: socio-economic position, health behaviours and knowledge, developmental events, and "system integrity." For example, individuals with higher IQs tend to occupy safer, more professional occupations or IQ may affect one's knowledge and adherence to protective health behaviors. Childhood IQ may be a result of environmental and genetic influences on developmental events which may affect health outcomes later in life. IQ may also be a measure of the body's overall efficiency or system integrity which would be related to mortality outcomes. System integrity is related to prior developmental events, but has less emphasis on earlier development, i.e., IQ at any age may be indicative of the body's general health.

Personality is defined as a "set of cognitive-motivational and socio-emotional traits and behaviour patterns" (Martin, Friedman, & Schwartz, 2007). Current research places personality traits within the "Big Five Taxonomy of Traits" (Goldberg, 1993). The Big Five organize personality traits into the follow domains: extraversion (activity and

sociability), agreeableness (likability and friendliness), conscientiousness (dependability and will to achieve), emotional stability (adjustment vs. anxiety), and openness to experience (imaginativeness, broad-mindedness, and artistic sensibility) (Poropat, 2009).There are many studies on the link between personality and mortality, specifically, conscientiousness. A meta-analysis by Bogg and Roberts (2004) demonstrated a clear link between conscientiousness and mortality risk. Martin, Friedman and Schwartz (2007) found that higher scores of childhood conscientiousness led to longer life spans. Schwartz et al. (1995) theorized that "psychosocially stable and socially responsible people" may practice health protective habits that reduce risk of mortality, such as adherence to medical treatments or avoiding dangerous situations. Whereas individuals who are more impulsive maybe practice more risky behaviors (Pharo, Sim, Graham, Gross, & Hayne, 2011).

The current study focuses on comparing different mortality records sources to identify possible bias associated with a particular source. To do so, we examine whether and to what extent relying on different mortality record sources affects predictions of the likelihood of mortality using early life cognitive and personality measures.

2. Methods

2.1 Sample

The data used in this study was from the 2011 Project Talent Pilot Study. The pilot study was conducted in 2011 and 2012 to evaluate the feasibility of a follow up of Project Talent (PT) participants who are currently between 66 and 70 years of age. The pilot study sample was created by first randomly subsampling 10% of schools from the original sample, and then randomly selecting 10% of students from within those schools. The total sample size used in these analyses was 4,159 individuals.

2.2 Measures

Mortality status was determined by matching PT records to the DMF, the NDI, and LN. Matching to these records was done through batch-matching using individual identifiers, including name, date of birth, location, and social security number information, the latter when available. SSNs were available for only about 50 percent of PT Pilot Study sample members. Two rounds of matching were conducted for all records. For the first-round, we send all sample records separately to the DMF, NDI, and LN for matching using available personal identifiers. Then, we took additional individual identifiers obtained from the three sources after the first round of matching and conducted a second matching of the DMF, NDI and LN. Three sources had an equal chance to be searched with the same individual identifiers.

NDI collected death records starting in 1979. Thus, in order to analyze comparable time ranges for all three sources, we limited our study to mortality records found from 1979 to 2009.

For the analyses examining the relationship among cognitive ability, personality and mortality, five individual background variables were employed: age in 1960, social economic status (SES), self-reported health before 10 years old, and high school class rank. SES was created as a composite, consisting of students' self-reports of home value, parent's income, parent's education, parent's occupation, and other items such as the number of books in the home. High school class rank, shown by Hauser and Palloni (2011) to be a proxy measure of academic behavior, was created by comparing sample members' self-reported grades to the self-reported grades of all students in the same grade in his/her school. This relative ranking was then classified into five quintiles.

A composite measure of cognitive ability was created based on participants' scores on the 1960 tests of reading comprehension, abstract reasoning, and math. Cognitive ability was categorized into five quintiles.

Ten scales were used to assess personality traits in 1960. They were sociability, social sensitivity, impulsiveness, vigour, calmness, tidiness, culture, leadership, self-confidence, and mature personality. Roberts (2012) mapped the Project Talent personality item-level data to the Big Five and found that personality traits measured in Project Talent were highly correlated with the Big Five.

2.3 Statistical Methods

Basic descriptive statistics are reported to compare the coverage rates and mortality rates across the three different mortality record sources. Coverage rates were calculated by dividing the number of deceased each record source identified by the total number of deceased identified for the sample overall based on pooling information from all morality record sources, then multiplying the number by 100. Likewise, mortality rates were computed by dividing the number of deceased people that were identified by each source by the total number of records in the overall sample. Besides comparing coverage and mortality rates, descriptive statistics regarding demographic background, cognitive ability, and personality traits were analyzed to capture the general characteristics of identified decedents.

2.3.1 Survival Analyses

The major focus of the current study was to investigate whether and to what extent relying on mortality record sources with different coverage rates results in different findings regarding the relationship among cognitive ability, personality and mortality. Three sets of variables are examined in the current study, including demographic background, cognitive ability, and personality traits. A series of Cox proportional hazards regression models were conducted to evaluate how these three sets of variables are associated with mortality. The Cox model predicts the probability that a case will be deceased at time (*t*). In our study, the models predicted the mortality rate at time (*t*). At time (*t*) for an individual with a vector of explanatory variables $x = (x_1 ... x_k)$, the Cox model defines the hazard for mortality (h) at time (t) as

$$h_i(t) = \lambda_0(t)\exp(\beta_1 x_{i1} + ... + \beta_k x_{ik}).$$

In this equation, the term $\lambda_0(t)$ represents the baseline hazard that may vary over time; it is the hazard for mortality at time *t* when all independent variables values are equal to zero. The factor $\exp(\beta_1 x_{i1} + ... + \beta_k x_{ik})$ is time independent. $\beta = (\beta_1 ... \beta_k)$ is a vector of regression coefficients reflecting the effects of the vector of explanatory variables on mortality.

In order to examine how the likelihood of mortality risk was affected by demographic, cognitive ability, and personality traits, we specified four models:

(1) $h(t) = \lambda_0(t)\exp(\beta_j Background_j)$

(2) $h(t) = \lambda_0(t)\exp(\beta_j Background_j + \beta_1 CognitiveAbility)$

(3) $h(t) = \lambda_0(t)\exp(\beta_j Background_j + \beta_k Personality_k)$

(4) $h(t) = \lambda_0(t)\exp(\beta_j Background_j + \beta_1 CognitiveAbility + \beta_k Personality_k)$

The first model was estimated to approximate the total effects of a vector of *(j)* background information including age at base year (1960), SES, class rank, and self-reported health before age of 10 on mortality. The second model extended the Cox model by taking into account the effects of cognitive ability to assess how early life cognitive ability affects mortality when controlling for demographic background. The third model added a vector of (*k*) personality trait variables to investigate whether personality traits influence mortality with demographic background information as covariates. For the fourth model, we incorporated both cognitive ability and personality traits to examine how early life cognitive ability and personality were related to mortality when controlling for demographic background.

Due to different mortality rates for men and women, as well as our expectation that personality and cognitive ability would operate differently for men and women (Stone et al., 2011), we conducted separate analyses for males and females. Further, we conducted separate Cox regressions for each mortality records source. Our analysis focused on whether the same sets of predictors would be identified by the Cox models for each of the sources of mortality records.

A key assumption of using the Cox models is proportional hazards. That is, the Cox models assume that the effect of each covariate is the same at all points in time. We tested the proportional hazards assumption for all covariates included in our models. All variables displayed stable mortality risks at all points of time, except age for the male Cox models. Therefore, we introduced a time-dependent covariate (age by time) to represent the interactions between age and time in our Cox models for male.

3. Results

3.1 Comparisons of Coverage Rates and Mortality Rates

The original sample size included 4,159 records. After restricting analyses to deaths that occurred between the years of 1979-2009, the DMF and LN identified 108 deaths that were outside this range. These cases were dropped from all analyses resulting in an analysis sample size of 4,051 records. Of those, 510 people were confirmed as deceased in at least one source, for an overall mortality rate of 13% (Table 1). This rate of 13% is lower than the mortality rate of approximately 16% that would be expected for this cohort given the constraints applied.[1] Males had a higher mortality rate than females (16% vs. 9% respectively). LN and the DMF had very similar mortality rates, but the NDI showed a higher mortality rate than either LN or DMF for both men and women.

Table 1: Mortality Rate by Source and Gender, 1979-2009

Mortality Records Sources	ALL		Male		Female	
	Number	%	Number	%	Number	%
Lexis-Nexis (LN)	434	10.7	288	14.9	146	6.9
Death Master File (DMF)	424	10.5	280	14.5	144	6.8
National Death Index (NDI)	504	12.4	313	16.2	191	9.0
All Sources	510	12.6	317	16.4	193	9.1
Total	4051		1937		2114	

As shown in Table 2, the NDI had the highest coverage rate among the three sources. The NDI found almost 99% of the total cases confirmed as dead in our study,

[1] Derived from SSA life tables, 16% of individuals alive in 1960 would be dead in the years 1979-2009.

compared to 85% for LN and 83% for the DMF. In particular, LN and the DMF had much lower coverage rates for women, where the NDI still found 99% of deaths. Coverage rates were computed by dividing the number of deceased each source identified by the total number of deceased identified for the sample overall based on pooling information from all morality record sources, then multiplying the number by 100.

Table 2: Coverage Rate by Source and Gender, 1979-2009

Mortality Source	All		Male		Female	
	Number	%	Number	%	Number	%
Lexis-Nexis (LN)	434	85.1	288	90.9	146	75.7
Death Master File (DMF)	424	83.1	280	88.3	144	74.6
National Death Index (NDI)	504	98.8	313	98.7	191	99.0
All Sources	510		317		193	

3.2 Comparisons of Demographics, Cognitive Ability, and Personality Traits

We next compared the demographic characteristics of individuals identified as deceased in LN, the DMF and the NDI, including age, socioeconomic status, high school class rank, and self-reported health status before age 10. The purpose of this analysis was to see if the demographic makeup of deceased individuals differed across the records sources based on descriptive statistics. When comparing across sources (Table 3), relatively few differences were found. LN and the DMF had similar values for the demographic characteristics examined, and results for both LN and the DMF tended to differ slightly from the results of the NDI. Within sources, LN and the DMF had a higher proportion of females with higher SES and class rank than the NDI.

Table 3: Demographic Variables by Source and Gender, 1979-2009

	Male				Female			
	LN	DMF	NDI	All	LN	DMF	NDI	All
	%	%	%	%	%	%	%	%
SES Quintiles								
Lowest Quintile	23.7	24.0	23.0	22.7	22.0	22.3	25.0	24.7
2nd Quintile	24.7	24.4	23.4	23.4	22.7	23.0	21.2	21.0
3rd Quintile	18.3	18.1	18.1	18.8	22.0	21.6	21.2	22.0
4th Quintile	16.1	16.2	17.1	16.9	15.6	15.1	16.9	16.7
Highest Quintile	17.2	17.3	18.4	18.2	17.7	18.0	15.8	15.6
High School Class Rank								
Lowest Quintile	24.5	24.4	23.4	23.8	20.9	21.2	21.7	22.1
2nd Quintile	22.6	22.6	22.4	22.4	20.9	21.2	19.4	19.3
3rd Quintile	20.4	20.3	20.4	20.1	17.3	16.8	20.0	19.9
4th Quintile	16.4	16.5	16.4	16.5	24.5	24.8	25.0	24.9
Highest Quintile	16.1	16.2	17.4	17.2	16.6	16.1	13.9	13.8
Health Before Age 10								
Excellent	54.8	54.3	55.0	54.9	53.0	53.4	54.4	54.3
Average	38.2	38.5	38.6	38.7	37.1	37.4	36.8	36.4
Poor	7.1	7.3	6.5	6.4	9.9	9.2	8.8	9.3

Across sources, there was little difference in mean personality scores (Table 4). Individuals identified by the three records sources displayed similar personality characteristics.

Table 4: Personality Means by Source and Gender; 1979-2009

Personality Variable	LN		DMF		NDI		All	
Male	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev
Sociability	5.8	2.8	5.8	2.8	5.7	2.8	5.7	2.8
Social Sensitivity	3.6	2.1	3.6	2.1	3.6	2.1	3.6	2.1
Impulsiveness	2.0	1.7	2.0	1.7	2.0	1.6	2.0	1.6
Vigor	3.2	2.0	3.3	2.0	3.2	2.0	3.2	2.0
Calmness	3.5	2.4	3.6	2.4	3.5	2.4	3.5	2.4
Tidiness	4.5	2.6	4.5	2.6	4.6	2.6	4.6	2.6
Culture	4.2	2.2	4.2	2.2	4.2	2.2	4.2	2.2
Leadership	1.1	1.3	1.1	1.3	1.1	1.3	1.1	1.3
Self Confidence	4.8	2.5	4.8	2.5	4.8	2.5	4.8	2.5
Mature Personality	9.7	4.7	9.7	4.8	9.7	4.8	9.7	4.8
Female								
Sociability	7.1	2.9	7.1	2.9	7.0	2.9	7.0	2.9
Social Sensitivity	5.4	2.3	5.3	2.3	5.2	2.3	5.2	2.3
Impulsiveness	2.0	1.7	2.0	1.7	2.0	1.7	2.0	1.7
Vigor	3.4	2.2	3.4	2.2	3.5	2.2	3.5	2.2
Calmness	4.3	2.7	4.2	2.6	4.1	2.6	4.1	2.6
Tidiness	6.1	2.7	6.1	2.7	6.0	2.7	6.0	2.7
Culture	5.7	2.3	5.7	2.2	5.7	2.2	5.7	2.2
Leadership	1.5	1.5	1.4	1.5	1.4	1.4	1.4	1.4
Self Confidence	5.3	2.7	5.3	2.7	5.1	2.7	5.1	2.7
Mature Personality	11.4	5.2	11.3	5.2	11.1	5.2	11.1	5.2

As shown in Table 5, all three sources showed very similar distributions of cognitive ability by quintile for men. However, sizable differences in cognitive ability based on different sources were observed for women. For LN and the DMF, 37% of identified deaths were in the lowest cognitive ability groups (who were in the lowest and 2nd quintiles groups); however, the percentage for the NDI was 44%. The NDI seemed to contain more records for women with relatively lower cognitive ability.

Table 5: Cognitive Measures by Source and Gender, 1979-2009

	Male				Female			
Variable	LN	DMF	NDI	All	LN	DMF	NDI	All
Cognitive Quintile	%	%	%	%	%	%	%	%
Lowest Quintile	28.3	28.4	27.6	27.3	24.7	24.3	28.5	28.7
2nd Quintile	22.2	22.1	23.4	23.1	12.7	12.9	15.6	15.4
3rd Quintile	18.3	18.5	17.4	17.5	21.8	22.1	22	21.8
4th Quintile	15.8	15.9	16.8	17.2	26.8	27.1	21.5	21.3
Highest Quintile	15.4	15.1	14.8	14.9	14.1	13.6	12.4	12.8

3.4 Survival Analyses

Appendix A includes significant predictors in each of 4 models with associated Hazard Ratio trend by source and sex. Model 1 shows the association between mortality and demographic controls. Model 2 includes demographic controls and cognitive ability. Model 3 consists of the demographic variables plus 10 personality variables. Model 4 is the inclusion of all variables.

Males. For males, the models showed similar results for all three sources and the overall source which combined all three sources (see Appendix A). In model 1, the demographic characteristics only model, the pattern of results was the same across sources – greater age, poorer health before age 10 were associated with greater mortality.

In model 2, where cognitive ability was added, higher cognitive ability (quintiles 4 & 5) was associated with reduced mortality risk across sources.

In model 3, where personality items were added to the demographic variables in model 1, impulsiveness was the only personality variable associated with mortality risk and was found in all sources, with mortality increasing as impulsiveness increased. Poor health before age 10 remained a significant predictor of mortality risk for the LN and DMF, but not the NDI. Individuals who reported poor health before age 10 were more likely than individuals who did report excellent health before age 10 to have higher rates of mortality.

In model 4 which included demographic, cognitive, and personality items, significant predictors of mortality risk were generally similar across sources (see Appendix B). Higher impulsivity across sources was significantly associated with a 9% increase in mortality risk (HR=1.09). For those with higher cognitive ability, there was a significantly reduced risk of mortality across sources. The only difference across sources in model 4 was in the significance of health before age 10. Relative to those who reported excellent health before age 10, average and poor health were significant predictors for increased risk in LN and the DMF. Only average health was associated with increased risk in all deaths, and no health variables were significantly associated in the NDI.

Females. Similar significant factors were extracted by Cox models across three sources and the overall source, as shown in Appendix A. In model 1 age and poor health before age 10 were significantly associated with increased risk for mortality in three sources and the overall source. Females reporting poor health before age 10 had higher risk of death compared to those who reported excellent health before age 10. Class rank was found to predict mortality based on using mortality information in the NDI, but not the LN and the DMF. With the NDI as a source of mortality, higher class rank was found to be associated with lower mortality.

In model 2, age and poor health were significant predictors of increased risk. High age and poorer health before age 10 were associated with increased risk of death. High cognitive ability was also associated with reduced risk in the LN and NDI sources.

In model 3, only age was significantly associated with increased risk in all sources. Poor health before age 10 was associated with increased risk in LN, NDI, and the overall sources, but not in the DMF. Social sensitivity was associated with increased risk in LN and the DMF, but not in the NDI or overall. Females with higher level of social sensitivity tended to have increased risk for mortality. High class rank had a protective effect, but was only found in the NDI and all sources; it was not present in LN or the DMF.

Females also had generally similar predictors across sources in the final model (see Appendix C). Age was significantly associated with about a 25% increase in mortality risk across sources. High cognitive ability was significantly associated with a decreased risk of mortality across sources. For personality variables, greater social sensitivity was associated with an increased risk of mortality across sources. The only difference was health before age 10 in the DMF. Compared to those who reported excellent health before age 10, those who reported poor health were at higher risk for mortality in all sources except the DMF, were health before age 10 was not found to be a significant predictor of mortality.

4 Discussions

This study compared coverage and mortality rates of LN, DMF, and NDI from 1979-2009 by using data from the 2011 Project Talent Pilot Study. We found that NDI is the "gold standard" of mortality records source, which echoed previous researchers' conclusions (Cowper, Kubal, Maynard, & Hynes, 2002). Matching results were comparable for LN and DMF. Both LN and DMF identified approximately 89% of the total cases confirmed as dead in our study for men and 75% for women. NDI had the highest coverage rate among the three sources, 99% for men and women. Compared to LN and DMF, the NDI is a more comprehensive source for determining mortality status, particularly for women.

Our descriptive statistics showed that records identified through LN, DMF, and NDI are similar for men pertaining to demographic background, cognitive ability, and personality traits. However, sizable differences in cognitive ability based on different sources were detected for women. For LN and DMF, about 37% of identified definite female deaths accounted for the first two quintiles; however, the percentage for the NDI was 44%. The NDI appears to contain more records for women with relatively lower cognitive ability. On the other hand, both LN and DMF appear to be biased in regard to their data sources. LN is based on whether individuals are found in consumer and commercial credit bureau databases – individuals more likely to be active in the economy. These individuals will have credit histories, employment histories, and other financial information. DMF relies solely on death benefits, which makes it vulnerable to self-reported bias. That is to say, individuals with 10 years of employment history are more likely to be reported. The results indicated that the NDI might be a more comprehensive source for identifying female mortality status compared to the LN and DMF.

Results for a series of Cox proportional hazard models displayed similar patterns of predictors for the likelihood of morality across the three sources for both men and women. Models from LN and the DMF extracted exactly the same sets of variables

associated with mortality, including self-reported health before age 10, impulsiveness, and cognitive ability. Survival analyses conducted using the NDI database showed slightly different results compared to LN and the DMF. The model showed that self-reported health before age 10 years does not significantly influence the probability of mortality risk. Overall, our study demonstrated that the results of survival analyses by each mortality records source were similar. Relying on one source of mortality records to predict the risk of morality is unlikely to be biased. The greater inclusion of women with lower cognitive ability in the NDI probably makes it the better of the three records sources. While these findings suggest that concerns about the DMF suggested by other researchers may not apply in this case, we have several cautions. First, our analyses were limited to the years (1979 to 2009) covered by the NDI, and we might have found differences (e.g. personality traits might have different impacts on mortality risk across the human life span) had we not limited the sample in such a manner. Further, with other cohorts, there may be potential for bias that we did not observe. Research suggests that the completeness of records in the DMF is a function of period (DMF year) and age at death (with older decedents more likely to be reported), and it's also likely that these factors interact to affect coverage. Finally, it's possible that bias may occur in other types of analyses, particularly if researchers include factors that are likely related to completeness of records in the different sources. Our analyses suggest that bias is not necessarily an outcome of differential completeness of records across the sources, but nonetheless, researchers should carefully consider the relative strengths and weaknesses of the different sources based on the unique characteristics of their analyses.

References

Batty, G. D., & Deary, I. J. (2004). Early Life Intelligence and Adult Health. *British Medical Journal* , 585-586.

Bogg, T., & Roberts, B. W. (2004). Conscientiousness and Health-Related Behaviours: A Meta-analysis of the Leading Behavioral Contributors to Mortality. *Psychological Bulletin* , 887-919.

Cowper, D. C., Kubal, J. D., Maynard, C., & Hynes, D. M. (2002). A Primer and Comparative Review of Major U.S. Mortality Databases. *Annals of Epidemiology* , 462-468.

Deary, I. J. (2005). Intelligence, Health, and Death. *The Psychologist* , 610-613.

Fisher, S. G., Weber, L., Goldberg, J., & Davis, F. (1995). Mortality Ascertainment in the Veteran Population: Alternatives to the National Death Index. *American Journal of Epidemiology* , 242-250.

Goldberg, L. R. (1993). The Structure of Phenotypic Personality Traits. *American Psychologist* , 26-34.

Hauser, R. M., & Palloni, A. (2011). Adolescent IQ and Survival in the Wisconsin Longitudinal Study. *The Journals of Gerontology, Series B: Psychological Sciences and Social Sciences* , i91-i101.

Hill, M. E., & Rosenwaike, I. (2001/2002). The Social Security Administration's Death Master File: The Completeness of Death Reporting at Older Ages. *Social Security Bulletin* , pp. 45-51.

Lexis Nexis. (2011). *Lexis Nexis Batch Collections Solutions.* Retrieved June 28, 2012, from Lexis Nexis Risk Solutions: http://www.lexisnexis.com/risk/downloads/literature/batch-collections-solutions.pdf+

Lexis Nexis. (2012). *Lexis Nexis Identiy Verification and Authentication*. Retrieved June 22, 2012, from Lexis Nexis: http://www.lexisnexis.com/risk/identity-verification-authentication.aspx

Martin, L. R., Friedman, H. S., & Schwartz, J. E. (2007). Personality and Mortality Risk Across the Life Span: The Importance of Conscientiousness as a Biopsychosocial Attribute. *Health Psychology* , 428-436.

National Center for Health Statistics. (2011, March 17). *Matching Criteria.* Retrieved June 27, 2012, from National Death Index: http://www.cdc.gov/nchs/data/ndi/NDICriteria_Front.pdf

National Technical Information Service. (n.d.). *Social Security Administration's Death Master File*. Retrieved June 22, 2012, from NTIS.gov: http://www.ntis.gov/products/ssa-dmf.aspx

Office of Information Services. (2011, December 19). *About the National Death Index*. Retrieved June 22, 2012, from Centers for Disease Control and Prevention: http://www.cdc.gov/nchs/data_access/ndi/about_ndi.htm

Pharo, H., Sim, C., Graham, M., Gross, J., & Hayne, H. (2011). Risky Business: Executive Function, Personality, and Reckless Behavior During Adolescence and Emerging Adulthood. *Behavioral Neuroscience* , 970-978.

Poropat, A. E. (2009). A Meta-Analysis of the Five-Factor Model of Personality and Academic Performance. *Psychological Bulletin* , 322-338.

Roberts, B. (2012). Personality in Project Talent. [PowerPoint Slides.]

Schall, L. C., Buchanich, J. M., Marsh, G. M., & Bittner, G. M. (2001). Utilizing Multiple Vital Status Tracing Services Optimizes Mortality Follow-up in Large Cohort Studies. *Annals of Epidemiology* , 292-296.

Schisterman, E. F., & Whitcomb, B. W. (2004). Use of the Social Security Administration Death Master File for Ascertainment of Mortality Status. *Population Health Metrics* .

Schnorr, T. M., & Steenland, K. (1997). Identifying Deaths before 1979 Using the Social Security Administration Death Master File. *Epidemiology* , 321-323.

Schwartz, J. E., Friedman, H. S., Tucker, J. S., Tomlinson-Keasey, C., Wingard, D. L., & Criqui, M. H. (1995). Sociodemographic and Psychosocial Factors in Childhood as Predictors of Adult Mortality. *American Journal of Public Health* , 1237-1245.

Social Security Administration; Office of the Inspector General. (2003). *The Social Security Administration's Efforts to Process Death Records and Improve its Death Master File.* Baltimore: Office of the Inspector General.

Stampfer, M. J., Willett, W. C., Speizer, F. E., Dyster, D. C., Lipnick, R., & Rosner, B. (1984). Test of the National Death Index. *American Journal of Epidemiology* , 837-839.

Appendix A: Cox Model Hazard Ratio Summary by Source

Parameter	Male				Female			
	LN	DMF	NDI	All	LN	DMF	NDI	All
Age (Base year age)	-	-	-	-	+	+	+	+
SES								
Health before 10 (ref.=Excellent)								
Average Health	+	+		+				
Poor Health	+	+			+		+	+
Class Rank								
Follow-up Status								
Age*Death	+	+	+	+				
Cognitive Ability (Lowest Performing)								
Cognitive Ability (2nd Quintile)					-	-	-	-
Cognitive Ability (3rd Quintile)								
Cognitive Ability (4th Quintile)	-	-	-	-				
Cognitive Ability (Highest Performing)	-	-	-	-	-	-	-	-
Sociability								
Social Sensitivity					+	+	+	+
Impulsiveness	+	+	+	+				
Vigor								
Calmness								
Tidiness								
Culture								
Leadership								
Self Confidence								
Mature Personality								

+ indicates increased hazard for significant ($p<.05$) predictor, - indicates decreased hazard

Appendix B: Cox Models Results by Source, Males only

Cox Models Results by Source: Males only

	LN			DMF			NDI			All		
Parameter	**Parameter Estimate**	**Pr > ChiSq**	**Hazard Ratio**	**Parameter Estimate**	**Pr > ChiSq**	**Hazard Ratio**	**Parameter Estimate**	**Pr > ChiSq**	**Hazard Ratio**	**Parameter Estimate**	**Pr > ChiSq**	**Hazard Ratio**
Age (Base year age)	-1.05	0.01	0.35*	-1.07	0.01	0.34*	-1.10	0.00	0.33*	-1.12	0.00	0.33*
SES (Lowest Quintile)												
SES (2nd Quintile)	-0.04	0.84	0.96	-0.08	0.68	0.92	-0.02	0.91	0.98	-0.03	0.89	0.97
SES (3rd Quintile)	-0.14	0.51	0.87	-0.15	0.50	0.86	-0.10	0.62	0.90	-0.05	0.82	0.96
SES (4th Quintile)	-0.18	0.43	0.84	-0.19	0.41	0.83	-0.05	0.80	0.95	-0.06	0.78	0.94
SES (Highest Quintile)	0.16	0.48	1.18	0.16	0.49	1.18	0.32	0.15	1.37	0.30	0.17	1.35
Health before 10 (ref.=Excellent)												
Average Health	0.28	0.05	1.33*	0.31	0.04	1.37*	0.26	0.07	1.29	0.27	0.05	1.32*
Poor Health	0.54	0.04	1.71*	0.58	0.03	1.79*	0.44	0.09	1.56	0.44	0.10	1.55
Class Rank (Lowest Performing)												
Class Rank (2nd Quintile)	0.30	0.14	1.35	0.31	0.12	1.37	0.30	0.12	1.35	0.29	0.13	1.34
Class Rank (3rd Quintile)	0.22	0.31	1.24	0.22	0.32	1.24	0.27	0.18	1.32	0.25	0.22	1.28
Class Rank (4th Quintile)	0.21	0.36	1.23	0.23	0.33	1.25	0.27	0.22	1.30	0.26	0.23	1.30
Class Rank (Highest Performing)	0.23	0.34	1.26	0.25	0.31	1.28	0.32	0.16	1.37	0.29	0.20	1.34
Follow-up Status	0.04	0.80	1.04	0.08	0.57	1.08	0.04	0.75	1.04	0.02	0.86	1.02
Age*Death	0.02	0.00	1.02*	0.02	0.00	1.02*	0.02	0.00	1.02*	0.02	0.00	1.02*
Cognitive Ability (Lowest Performing)												
Cognitive Ability (2nd Quintile)	-0.05	0.81	0.95	-0.08	0.69	0.92	0.03	0.89	1.03	0.02	0.90	1.02
Cognitive Ability (3rd Quintile)	-0.26	0.23	0.77	-0.28	0.21	0.76	-0.30	0.15	0.74	-0.30	0.15	0.74
Cognitive Ability (4th Quintile)	-0.53	0.02	0.59*	-0.57	0.01	0.57*	-0.49	0.02	0.62*	-0.44	0.04	0.64*
Cognitive Ability (Highest Performing)	-0.54	0.02	0.58*	-0.61	0.01	0.55*	-0.63	0.01	0.53*	-0.59	0.01	0.55*
Sociability	0.02	0.53	1.02	0.02	0.51	1.02	-0.01	0.83	0.99	0.00	0.95	1.00
Social Sensitivity	-0.01	0.73	0.99	-0.01	0.83	0.99	0.00	0.93	1.00	0.00	0.92	1.00
Impulsiveness	0.09	0.04	1.09*	0.09	0.05	1.09*	0.09	0.03	1.09*	0.09	0.04	1.09*
Vigor	-0.04	0.35	0.96	-0.03	0.51	0.97	-0.06	0.11	0.94	-0.06	0.14	0.94
Calmness	0.00	0.98	1.00	0.01	0.89	1.01	0.00	0.99	1.00	0.00	0.95	1.00
Tidiness	-0.02	0.47	0.98	-0.03	0.42	0.97	-0.02	0.55	0.98	-0.02	0.53	0.98
Culture	0.00	0.99	1.00	-0.01	0.88	0.99	0.03	0.48	1.03	0.02	0.52	1.03
Leadership	-0.05	0.48	0.96	-0.07	0.32	0.94	-0.08	0.21	0.92	-0.08	0.19	0.92
Self Confidence	0.01	0.66	1.01	0.02	0.64	1.02	0.01	0.85	1.01	0.00	0.88	1.01
Mature Personality	0.00	0.84	1.00	0.00	0.87	1.00	0.00	0.84	1.00	0.00	0.88	1.00

*p<.05

Appendix C: Cox Models Results by Source. Females only

Cox Models Results by Source: Females only

	LN			DMF			NDI			All		
Parameter	**Parameter Estimate**	**Pr > ChiSq**	**Hazard Ratio**	**Parameter Estimate**	**Pr > ChiSq**	**Hazard Ratio**	**Parameter Estimate**	**Pr > ChiSq**	**Hazard Ratio**	**Parameter Estimate**	**Pr > ChiSq**	**Hazard Ratio**
Age (Base year age)	0.22	0.00	1.24*	0.23	0.00	1.25*	0.22	0.00	1.25*	0.23	0.00	1.25*
SES (Lowest Quintile)												
SES (2nd Quintile)	0.02	0.95	1.02	0.01	0.97	1.01	-0.11	0.66	0.90	-0.11	0.66	0.90
SES (3rd Quintile)	0.18	0.53	1.20	0.13	0.65	1.14	0.11	0.66	1.11	0.13	0.60	1.14
SES (4th Quintile)	-0.22	0.49	0.80	-0.24	0.45	0.78	-0.13	0.63	0.88	-0.13	0.62	0.88
SES (Highest Quintile)	0.19	0.55	1.21	0.17	0.59	1.19	0.09	0.74	1.10	0.08	0.77	1.09
Health before 10 (ref.=Excellent)												
Average Health	0.25	0.20	1.28	0.25	0.21	1.28	0.24	0.16	1.27	0.24	0.17	1.27
Poor Health	0.68	0.03	1.98*	0.59	0.07	1.80	0.58	0.04	1.79*	0.65	0.02	1.92*
Class Rank (Lowest Performing)												
Class Rank (2nd Quintile)	-0.02	0.95	0.98	-0.02	0.94	0.98	-0.03	0.92	0.97	-0.06	0.81	0.94
Class Rank (3rd Quintile)	-0.36	0.23	0.70	-0.40	0.18	0.67	-0.14	0.59	0.87	-0.17	0.49	0.84
Class Rank (4th Quintile)	-0.07	0.82	0.94	-0.06	0.83	0.94	0.06	0.79	1.07	0.02	0.93	1.02
Class Rank (Highest Performing)	-0.45	0.18	0.64	-0.45	0.18	0.64	-0.44	0.14	0.64	-0.50	0.10	0.61
Follow-up Status	0.15	0.46	1.16	0.17	0.40	1.19	-0.01	0.97	0.99	0.00	1.00	1.00
Cognitive Ability (Lowest Performing)												
Cognitive Ability (2nd Quintile)	-0.94	0.00	0.39*	-0.91	0.01	0.40*	-0.84	0.00	0.43*	-0.84	0.00	0.43*
Cognitive Ability (3rd Quintile)	-0.40	0.15	0.67	-0.36	0.20	0.70	-0.42	0.08	0.66	-0.43	0.07	0.65
Cognitive Ability (4th Quintile)	-0.17	0.56	0.85	-0.13	0.65	0.88	-0.41	0.10	0.66	-0.41	0.10	0.66
Cognitive Ability (Highest Performing)	-0.83	0.02	0.44*	-0.80	0.02	0.45*	-0.90	0.00	0.41*	-0.84	0.01	0.43*
Sociability	0.00	0.95	1.00	0.00	0.97	1.00	-0.01	0.82	0.99	-0.01	0.82	0.99
Social Sensitivity	0.13	0.02	1.14*	0.13	0.02	1.14*	0.11	0.03	1.11*	0.11	0.03	1.11*
Impulsiveness	-0.04	0.55	0.97	-0.03	0.63	0.97	-0.02	0.68	0.98	-0.02	0.65	0.98
Vigor	-0.08	0.14	0.93	-0.08	0.13	0.92	-0.03	0.57	0.97	-0.03	0.50	0.97
Calmness	-0.05	0.28	0.95	-0.05	0.26	0.95	-0.07	0.12	0.94	-0.07	0.09	0.93
Tidiness	-0.03	0.45	0.97	-0.03	0.46	0.97	-0.02	0.54	0.98	-0.02	0.66	0.98
Culture	0.01	0.83	1.01	0.02	0.72	1.02	0.03	0.58	1.03	0.03	0.49	1.04
Leadership	0.09	0.25	1.10	0.08	0.34	1.08	0.08	0.27	1.08	0.08	0.26	1.08
Self Confidence	0.06	0.13	1.07	0.06	0.16	1.06	0.03	0.40	1.03	0.03	0.36	1.03
Mature Personality	-0.01	0.65	0.99	-0.01	0.62	0.99	-0.02	0.38	0.98	-0.02	0.35	0.98

*p<.05