Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Xu Chen Date

Comparison of mean residual life between mental disease patients and healthy population in a national survey

By

Xu Chen
Master of Science in Public Health

Department of Biostatistics and Bioinformatics

_____

Amita K. Manatunga
Thesis Advisor

_____

Jose N. Binongo
Reader

Comparison of mean residual life between mental disease patients and healthy population in a national survey

By

Xu Chen

Bachelor of Science
Beijing University of Technology
2007

Thesis Committee Chair: Amita Manatunga, PhD
Thesis Committee Member: Jose Binongo, PhD

An abstract of
A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Science in Public Health
in the Department of Biostatistics and Bioinformatics
2013

Abstract

Comparison of mean residual life between mental disease patients and healthy population in a national survey

By Xu Chen

Mental illness has a shockingly high prevalence across the globe and in the United States, which has become a heavy public health burden. Many studies have shown that people with mental illness are at high risk of mortality and morbidities. Understanding the magnitude of premature mortality and underlying risk factors among patients with mental disease has important implications for decreasing the burden of mental disease.

To quantifying the social and economic impact of mortality caused by mental disease in a society, researchers and health policymakers need index that are easy to understand for most people. One of the most commonly used methods measuring premature mortality is years of potential life lost (YPLL). YPLL is an estimate of the years a person would have lived if he or she had not died. Despite of its nice interpretation, the estimate of YPLL is a biased estimate because censored data are not included. Another method to measure the risk of premature death is Cox proportional hazard model. But the estimated risk is not enough to portray the actual degree of year gap between people with and without mental illness. We demonstrate an alternative method to examine the premature mortality of mental illness patients in terms of mean or median residual life. The objective of the thesis is to illustrate the use of mean or median residual life methods in describing the two populations: individuals with and without mental disease. We show that how policymakers can use the method of mean or median residual life to summarize the data with censoring and to give nice and straightforward interpretation.

Data are obtained from the 1989 National Health Interview Survey mental health supplement, with mortality data through 2006. There are 80,850 participants in total, among which 16,435 are dead during the follow-up time. Mean residual lifetimes and median residual lifetimes are estimated and compared over the age range 24 to 100 between the populations with and without mental disease, using the empirical likelihood test. In addition, gender, race and mental health insurance coverage modify the effect of mental disease on premature mortality.

Comparison of mean residual life between mental disease patients and healthy population in a national survey

By

Xu Chen

Bachelor of Science
Beijing University of Technology
2007

Thesis Committee Chair: Amita Manatunga, PhD
Thesis Committee Member: Jose Binongo, PhD

A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Science in Public Health
in the Department of Biostatistics and Bioinformatics
2013

Acknowledgements

I wish to express my very great gratitude to my advisor Dr. Amita Manatunga for her valuable and constructive guidance throughout this work. It would not have been possible without her kind support and help for me to complete this thesis. Her willingness to give her time so generously has been very much appreciated.

My grateful thanks are also extended to Dr. Mai Zhou at the University of Kentucky for his advice on statistical problems within his areas of expertise.

Dr. Jose Binongo is kind enough to be the reader of this work and offer criticism, for which I am appreciative.

I would also like to extend my thanks to all the professors and staff in the Department of Biostatistics and Bioinformatics. They have offered the greatest environment for me to learn and enjoy my two years' study here.

Lastly, I thank my parents and friends for their constant encouragement and assistant.

Table of Contents

1. Introduction

1.1 Burden and Causes of Mental Illness

Millions of people worldwide are affected by mental disease. Statistics in 2004 show that mental disorders accounted for 13% of the global burden of disease, which is defined as premature death and years lived with disability [1]. In the United States, one in four adults experience mental health disorders, and mental disorders are among the most common causes of disability [2]. The high prevalence of this disease raises the question about understanding its impact on patients' life length and advocacy for mental health.

Excess mortality and comorbidity among persons with mental illness has been reported in a plethora of studies [3-5]. Although the excess deaths associated with mental illness are partially attributed to unnatural causes, such as suicide, homicide and accidents, increasing evidence has verified that general medical disorders explains more premature deaths among mental disease patients. Research show that patients with psychiatric disorder are at higher risk of comorbidities such as cardiovascular diseases, respiratory illness, substance abuse and other physical illness, compared to general population [6-9]. A literature review on bipolar spectrum disorders reviewing 17 large studies states that individuals with bipolar disorders are at significantly higher risk of premature death from natural causes, such as cardiovascular, respiratory, cerebrovascular, and endocrine illnesses [10]. Quantitative analysis results are given in a meta-analysis on excess mortality of schizophrenia: "28% of the excess mortality is attributable to suicide and 12% to accidents. The rest of the excess mortality is from the same broad range of conditions

which cause deaths in the general population." [11] According to the 2001-2003 National Comorbidity Survey Replication, a nationally epidemiological survey, "more than 68 percent of individuals with a mental disorder reported having at least one general medical disorder" [12].

The fact that general medical disorders are major contributor to premature mortality among mental illness patients underscores the importance of quantifying its effect and identifying potential risk factors. Previous studies have presented many risk factors, such as chronic stress, socioeconomic factors, smoking, excessive alcohol and drug consumption, sedentary lifestyle, lack of nutrition and poor quality of health services [12]. Other critical determinants of mental illness worthy of attention include gender, race and health-care system factors such as health insurance coverage.

Gender is strongly associated with mental illness, due to the long existing disparities between men and women in terms of social economic status, power to control their own life, access to health-related resources and other important determinant to mental health. Striking gender differences have been found in mental disease patterns [13]. Race, unlike gender, affects mental health in rather complex ways. Inconsistent results were found in studies of race and mental health, because the patterns of racial disparity in mental health vary by looking at different indicators of mental health status [14]. As for mental health insurance, it provides financial protection for mental disease patients to receive sufficient treatment and other mental health services. Those who uncovered with insurance are at higher risk of low economic status, physical illness and subsequently bad life quality.

1.2 Problem Statement

Mathematical methods have been important tools in analyzing the magnitude of premature death among mental health patients and the effect of risk factors. One of the measures of premature death is years of potential life lost (YPLL).

YPLL is an estimate of the years a person would have lived if he or she had not died [15]. It is calculated based on the life expectancy estimate for each age, which is normally acquired by data from the Centers for Disease Control and Prevention (CDC). The YPLL is defined as: the difference between a person's age at death and the mean survival age for a living cohort at that age [16, 17]. For example, if a person dies at age 50 in the year of 2012, and the statistics from CDC shows that the general cohort in year 2012 at age 50 are expected to live up to 75 years old, then the year of life lost for this individual is 25. According to the concept, YPLL only can be calculated for decedents.

The YPLL has been brought into attention since 1982, when CDC included YPLL measure in its standard set of tables of reported disease. It is promoted as an important index in quantifying the social and economic impact of mortality in a society. We can know about how many years the individual dies before the "natural" time, which leads to loss of life length and possible contribution to the society. The advantage of this approach is that this measurement is quite easy to comprehend for most people, in contrast to usual mortality statistics, such as mortality rate, which are less intuitive. By focusing on a specific cause of death or a certain group of population, YPLL can also highlights causes of premature mortality or subgroups in great need of health intervention. Its nice

interpretation permits policymakers and caregivers to comprehend the causes and demographical distribution of premature death burden and prioritize people most at risk. However, this measure will easily cause bias on estimating premature death when the study population includes both decedents and people who are alive at the end of study. To calculate YPLL for this population, one has to exclude subjects alive, and calculate each decedent's YPLL. The exclusion of the living cohort results in not only loss of information, but also biased estimation of excess death impact on the whole population.

Due to the shortcoming of YPLL, alternative methods are needed to study the population with both dead and living individuals. This is the typical situation where survival analysis is employed. Seeing the occurrence of death as event, and individuals alive as censored observation, we are able to compare difference of mortality between two populations using methods such as log-rank test or Cox proportional hazard regression model. Log-rank test is a commonly used nonparametric hypothesis test to examine the equivalence of two groups' overall survival curves, which requires no assumption on the two groups' hazard distribution. Another method, Cox regression model, enables to take explanatory variables into account. The only assumption for this model is that the hazard ratio between two populations is constant over time. In spite of their wide applicability, these two methods are limited in that they fail to provide estimation of expected lifetime and make comparison between groups at any time points.

A myriad of health service research have investigated how mental illness affects life length and possible ways to address this problem utilizing the two statistical methods

mentioned above. One example is Druss et al. [18], which examined the magnitude and complex causes of premature death using national representative data from 1989 National Health Interview Survey mental health supplement. Years of potential life lost were calculated for decedents to quantify the year gap between general population and mental disorder patients. Cox proportional hazard model was used as well to examine the relative risk of mortality for persons with mental illness, and potential risk factors contributing to mortality. The paper found a two-fold increase in the risk of mortality for mental disease patients, and 82% of the excess mortality can be explained by socioeconomic status, health care accessibility and baseline physical conditions, adjusted for demographics.

Although this study provides important insights about excess rates of mortality under a national context, the results of YPLL and the age gap between decedents with and without mental illness might be biased. It has been pointed out that since the whole population has both decedents and subjects alive, the YPLL calculated from only dead cohort provide incomplete information about the total study population. In addition, the estimate of excess mortality risk from Cox proportional hazard model fails to portray the actual degree of the year gap between population with and without mental illness. For example, when we state that persons with mental disease have twice the risk of death than those without mental disease, we cannot know exactly the difference of number of years they can live. This method is not enough to provide a complete picture of the burden that excess death imposed on this national population. Thus, a new measure of premature mortality is needed to gauge its magnitude.

1.3 Purpose Statement

There has been a lot of effort focused on improving mental health status in the nation. However, the quality and efficiency of mental health care is still far from satisfaction. To promote actions aiming to reduce unnecessary premature mortality caused by mental illness and to evaluate competing claims for allocation of health resources, we need a better understanding of the impact and the mechanism of mental illness.

Achieving the goals of reflecting magnitude of mental illness caused excess death for a general population requires a quantitative measure that is applicable to data subject to censoring. YPLL, though has nice interpretation, excludes subjects alive and tends to conclude to biased result. Proportional hazard model fails to provide the actual year gap between people with and without mental illness. In this case, mean residual life or median residual life is a good choice.

The mean residual lifetime for a person at a certain age estimates the expected remaining life length he or she has given that person's age. The mean residual life or median residual life function have been promoted by many researchers as practical summary of the lifetime distribution and residual life expectancy. We can calculate the mean/median residual lifetimes for population with and without mental illness at a given age, then compare them to examine how many years the gap is. This method is desirable for reflecting excess death in a general population in that it incorporates information about death from both censored and uncensored subjects, and that it has good and

straightforward interpretation. The comparison of mean/median residual lifetimes provides a more accurate and clear picture of premature mortality by describing the number of years that mental disease patients have short of general people at any age. These results are easier to be understood by policy makers and the general public than other mortality measures, which enables public health practitioners to call more attention to mental health problem and promote reforms on mental health care.

1.4 Objectives

The purpose of this article is to demonstrate utilization of an alternative method measuring excess mortality of individuals with mental disease. We estimated and compared the mean residual lifetimes of persons with mental illness with general population in the United States during selected follow-up years to examine whether there is significant difference in their life length at given ages. In addition, we examined some potential risk factors about their effects on premature mortality. The nationally representative sample provides the generalizability of our study findings.

2. Methods

2.1 Study Setting

The National Health Interview Survey (NHIS) has monitored the health of the United States since 1957. NHIS data on a broad range of health topics are collected through personal household interviews. In 1989, the National Institute of Mental Health in collaboration with the National Center for Health Statistics developed a special mental health supplement to the NHIS for the purpose of estimation the prevalence of serious mental illness in the United States[19]. The supplement survey got a response rate of 97%[19]. Mental diseases listed in the survey include schizophrenia, paranoid or delusional disorder, manic episodes, manic depression, major depression and personality disorder.

The death records for the survey subjects were obtained from National Death Index (NDI) mortality data with a range from 1989 to 2006. Then it was linked to 1989 NHIS mental health supplement. Therefore, duration of follow-up for decedents was calculated as the difference between the year of NHIS interview and the year of death; for respondents alive by 2006, follow-up time was the duration from NHIS interview to the end of 2006.

The NDI includes data both for overall mortality and cause-specific mortality for individuals older than age 18. Data met the following inclusion criteria were collected for analysis: eligible for mortality follow-up, sample weight greater than 0 and with no missing record on mental diagnosis[18]. The study population includes 80,850

participants, within which 16,435 are decedents.

Cause of death was determined using the underlying cause of death as documented by ICD-10 code groups. Record of death cause was available for 99.5% of decedents[18]. Unnatural causes of death include suicides, homicides and accidents.

2.2 Mean Residual Life

Because of the nature of the survival data, it is important to be able to describe or predict the residual life distribution of the subjects in the study. For example, at a diagnosis of cancer, patients may wish to know how many years they are expected to survive given their ages and how long there life may be prolonged if they undergo some cancer treatment. In social science, researchers are interested in the relationship between people's duration stay on a job and their willingness to move to a new job.

Mean residual life is introduced to answer these questions. Let X be the random variable with a life distribution F with finite first moment. Let the survival function be S(t)=1-F(t). The mean residual life function $m(t)$ at a given time t is defined as

$$m(t) = \begin{cases} E[X - t | X > t] & for\ S(t) > 0 \\ 0 & for\ S(t) = 0 \end{cases}$$

Where $E(.)$ is the expectation. It can be interpreted as the expected additional life length given a subject survived up to time t.[20]

In the situation of cancer survival prediction, for a patient with age 50, for example, $m(50)$ provides the expected remaining life length given that the patient has reached age 50. In the social science, increasing mean residual life is frequently found in job mobility

studies: the longer a person has been on a job, the less possible he or she wants to move to a new job [20].

Mean residual life function has many attractive properties. It is a good alternative to hazard function. While hazard function provides information about a small interval right after a time point t, the mean residual life function at time t considers the whole interval after that time point[20]. From the practical standpoint, the mean residual life function allows researchers and clinicians understand advantages of a specific treatment regime in terms of the expected remaining life length of patients, in contrast of hazard function, which requires substantial knowledge of statistical concepts.

2.3 Median Residual Life

Median residual life function is the quantile counterpart of mean residual life function. Median residual life at a time point is defined as the median of the remaining lifetimes among survivors beyond that particular time point[21]. Let X be the random variable with a life distribution F, then the median residual life function at a given time t is $Med(t) = median[X - t|X > t]$. In other words, median residual life $Med(t)$ is the time point φ that solves the equation $\frac{1-F(t+\varphi)}{1-F(t)} = 0.5$ [22].

Though mean residual life function has many good properties, there exist some shortcomings with it. The estimation of mean residual life will be strongly affected by few outliers, which results in unstable results. In this case, median residual life is recommended as a better behaving alternative.

2.4 Empirical Likelihood Ratio Test

Empirical likelihood is a nonparametric method of statistical inference, which parallels the theory of parametric likelihood, but requires no assumption on the underlying distribution of the data in model[23]. First employed by Thomas and Grunkemeier[24], this method was further applied in survival analysis by Owen[25], who devised the method to construct confidence interval for mean and some other statistics using empirical likelihood ratio test. Empirical likelihood ratio test attracts great attention because it inherits the advantages of likelihood ratio test and can be adopted in quite general settings.

Suppose that $X_1, \dots, X_n$ are i.i.d. nonnegative random variables of interest from a continuous life distribution F_0. Independent of the X there are censoring time $C_1, \dots, C_n$ which are i.i.d. with a distribution G_0. We can only observe:

$T_i = \min(X_i, C_i)\,;\ \delta_i = I[X_i \leq C_i]$ for $i = 1,2, \dots, n$

where n is the number of observations, T_i is the observed survival time and δ_i is the censoring indicator.

Let p_i denote the probability mass function of the observation T_i, the empirical likelihood (EL) for the observed data (T_i, δ_i) is $EL = \prod_{i=1}^{n} [p_i]^{\delta_i} [1 - p_i]^{1-\delta_i}$. It has been shown that the Kaplan-Meier estimator computed from (T_i, δ_i) maximizes the above empirical likelihood with respect to p_i. Let us denote the maximum empirical likelihood value obtained by plugging in Kaplan-Meier estimates as EL(KM).

In this study, the null hypothesis for the empirical likelihood ratio test is

$H_0: \sum_{i=1}^{n} g(T_i)p_i = \theta$, where θ is the mean residual life or median residual life we wish to test, the quantity $\sum g(T_i)\, p_i$ is a general form for the definition of mean/median residual life. We will discuss the specified g-function for mean residual life and median residual life below.

So an extra constraint condition is added to the maximization of the likelihood: $\sum_{i=1}^{n} g(T_i)p_i = \theta$, where $g(t)$ is a given function such that $0 < Varg(T) < \infty$ and θ is the value that is of interest.

Zhou and Jeong[22] showed that once the constrained maximum is obtained, which we denote as EL(Constrain), then the empirical likelihood ratio statistic, -2log$\frac{EL(Constrain)}{EL(KM)}$, converges in distribution to a chi-square distribution.

According to the definition of mean residual life function,

$$m(t) = E[X - t|X > t] = E[X|X > t] - t = \frac{\int_t^{\infty} sdF(s)}{1 - F(t)} - t = \frac{\int_t^{\infty} 1 - F(s)ds}{1 - F(t)}$$

Then the hypothesis $H_0: m(t) = \mu$ can be written as

$$H_0: \frac{\int_t^{\infty} sdF(s)}{1 - F(t)} = t + \mu\,,$$

which is also equivalent to

$$H_0: \int_t^{\infty} [s - (t + \mu)]dF(s) = 0$$

if we write $\int_t^{\infty} dF(s) = 1 - F(t)$. So, the g-function for mean residual life testing is

$$g(s) = [s - (t + \mu)]I_{[s>t]}$$

For median residual life $Med(t)$, which has been defined as the time point φ that

solves the equation $\frac{1-F(t+\varphi)}{1-F(t)} = 0.5$, it can also be written as the solution to $F(t+\varphi) - 0.5 * F(t) = 0.5$. Then the hypothesis test for the median residual life $H_0: Med(t) = b$ is equivalent to the test $H_0: \int_0^\infty g_b(s)dF(s) = 0$, where $g_b(s) = I_{[s\leq(x+b)]} - 0.5 * I_{[s\leq x]} - 0.5$. $g_b(s)$ is the g-function for testing median residual life [22, 23, 26].

2.5 Two-Sample Mean Residual Life Comparison

A variety of methods have been proposed to estimate and make inference on the mean residual life function[27-29]. For this study, the estimation is conducted via the empirical likelihood ratio test for censored survival data. This method doesn't require estimation of underlying density function, and inherits good properties of the likelihood test[22].

The two-sample mean residual lifetimes comparison is developed from the test introduced in 2.4. If we are to test the equality of two mean residual life at age t_0, expressed as $H_0: \frac{m_1(t_0)}{m_2(t_0)} = 1$, where $m_i(t_0)$ $(i = 1,2)$ denote the mean residual time from sample i at age t_0, we shall first evaluate the auxiliary hypothesis: $H_{00}: m_1(t_0) = m_2(t_0) = \theta$, where θ is a randomly chosen time point within support values. This hypothesis can also be written as two hypotheses: $H_{01}: m_1(t_0) = \theta$ and $H_{02}: m_2(t_0) = \theta$. Two test statistics for those two hypotheses using empirical likelihood ratio test can be obtained, denoted as $W_1(\theta; t_0)$ and $W_2(\theta; t_0)$. Then the test statistic for the original

hypothesis $H_0: \frac{m_1(t_0)}{m_2(t_0)} = 1$ is $Q(\theta) = \min_\theta \frac{W_1(\theta;t_0)+W_2(\theta;t_0)}{\theta}$, which follows a chi-square distribution with 1 degree of freedom under the null hypothesis.

2.6 Cox Proportional Hazards Model

Cox proportional hazards model is one of the most commonly used methods in survival analysis. This method is able to take all available information into consideration, including those subjects on whom the event of interest doesn't happen. It is usually used to assess the importance of various covariates in the subjects' survival times. The basic assumption for proportional hazard model is that the hazard, or risk, for an individual is consist of two parts: the baseline hazard function describing how the hazard changes over time when the covariates all have no effect; and the effect parameters, describing how the hazard varies in response to explanatory covariates. The expression of the Cox regression model is $h(t;x) = h_0(t)\exp(\beta_1 x_1 + \cdots + \beta_k x_k)$, where $h(t;x)$ is the hazard function of a subject at time t with covariate values $x_1,\ldots, x_k$, $h_0(t)$ is the baseline hazard function, when the covariates are all zero, $\beta_1, \ldots, \beta_k$ are the regression coefficient for covariate $x_1, \ldots, x_k$, respectively [30]. This assumption implies that the ratio of the two hazard functions for two different levels of a covariate should be constant over time.

One big advantage of Cox proportional regression model is that there is no need to specify the baseline hazard function. So the effect of covariates on hazard can be estimated without making assumption on the baseline function form, which allows this regression model to be widely applied to various data and study.

However, the proportional hazard assumption should be tested to make sure that this regression model is appropriate for the data. P. Grambsch and T. Therneau [31] have developed diagnostics for the proportional hazard assumption based on weighted residuals. We used this method in the study to verify the proportionality.

2.7 Potential Risk Factors

Three potential risk factors associated with excess mortality among individuals with mental illness were selected: gender, race and mental health insurance coverage. Cox proportional hazard models were used to analyze the hazard ratio of death for persons with mental disorder over those without the disorder, adjusting for gender, race and mental health insurance, respectively. We also divided study subjects into subgroups based on their gender, race and health insurance coverage. Then comparisons of mean residual lifetimes were made within each of the subpopulation.

2.8 Statistical Analysis

Demographic analyses adjusted for complex survey design were conducted in Survey Data Analysis (SUDAAN). All the means and standard deviations were calculated with survey weights taken into account. Hazard ratios with their 95% confidence intervals were calculated in SAS. First we examined the relative hazard rate of mortality for subjects with mental disease. Then explanatory variables were added into the model. Mean/median residual life estimation and two-sample mean residual lifetimes

comparisons were conducted in R using package “emplik”[32]. R code for the analysis is included in the Appendix C.

3. Results

Table 1 presents descriptive information for the study population. 1725 survey subjects, out of a total of 80,850 participants, reported mental disorder in 1989. Mean age of individuals with mental disorder was 42.4, slightly younger than the population without mental disease (mean age 43.8). Gender and race composition was similar in these two groups. Individuals without mental illness had significantly larger percentage of mental health insurance coverage ($p<0.001$).

There were a total of 463 decedents with mental illness and 15,972 decedents without mental illness up until 2006(Table 2). Mental illness patients showed higher proportion of death (26.8% vs. 20.2%) during the 17-year follow-up period. The median survival time for those with mental illness is 77.5±1.2, which means at age of 77.5, half of the mental illness patients have died. While the median survival time for those without mental illness is 84.4±0.1, a difference of 7 years. The higher median age for both populations indicates that most death happens around the age 70 and older. Among the 463 decedents with mental disorder, only 25 persons (5.4%) died from unnatural causes (that is, suicide, homicide and accidents). This proportion is similar among the general population: 714 out of 15,972 deaths (4.7%) are caused by unnatural causes.

The Kaplan-Meier estimation for the two groups shows that mental disease patients

are likely to live shorter than persons without mental disease, especially after age of 40 (Fig. 1). Table 3 summarizes the survival probability at several ages with the 95% confidence intervals obtained from the Kaplan-Meier curve for people with and without mental illness. A majority of mental illness patients die between the age of 60 and 90, with the biggest reduction of survival probability happens after age 70. Dissimilarly, the survival probability for general population drops fast after 70, especially after age 80. This is consistent with the different mean survival age of two populations. At early ages, mental illness patients also tend to have smaller survival probability compared to people without mental disease. The difference of survival probability increases between the two groups as the age grows. The largest difference, as much as 20%, shows at the age 80. The survival trajectories are significantly different for these two populations ($p<0.001$ for log-rank test). It indicates that mental disease has significant impact on the patients' survival time.

Figure 2 shows the mean residual lifetime estimations for mental disease and non-mental disease groups over the age 24 to 100. We chose these two cut points because of the scarcity of death before 24 and after 100. Overall, the expected remaining life is always shorter for persons with mental disorder at all ages, compared with those who don't have mental disorder. The difference between two curves becomes smaller as the age goes larger. The tail performs unstable due to limited number of events. Table 3 reports the estimations of mean residual life and median residual life for people with and without mental illness at age 30 to 90, at every ten years. Significant differences are

shown at age 40 (p=0.044) and 60 (p=0.034). We can observe a decreasing year gap between the two populations in terms of mean residual lifetimes and median residual lifetimes. For example, a mental illness patient at age 40 is expected to have 36.4 remaining life length, while a 40-year-old person without mental disease has 43.7 additional years to live, a difference of 7 years. But at age 80, there is only 2 years gap between the two populations. Similar results can be found for median residual life. This trend is congruent with that reflected in the difference of survival probabilities over age.

Cox regression models were used to examine the effects of covariates gender, race and mental health insurance coverage. In the Cox proportional hazard model with the only covariate mental illness, the risk of mortality doubles [hazard ratio=2.09, 95% CI (1.90-2.29)] (Table 5.1). Existence of mental illness contributes significantly to the regression model ($p<0.001$). Other three univariate models, which includes just gender, race and mental insurance coverage, respectively, display statistically significant association between the explanatory variable and the outcome (Table 5.1). After controlling for gender, mental disease is still associated with more than 2-fold increase in death risk [hazard ratio=2.12, (1.93-2.33)]. The relative hazard rate of death adjusted for race is similar to that in unadjusted model [hazard ratio=2.08, (1.91-2.29)]. Adding mental health insurance as covariate, the hazard ratio for general people over mentally ill subjects stays roughly the same [hazard ratio=2.04, (1.86-2.23)] (Table 5.2). Adjusting for gender, race and mental insurance coverage, the existence of mental illness still doubles the risk of mortality [hazard ratio=2.07, (1.88-2.27)]. Mental disease always

statistically significantly elevates the risk of mortality no matter what gender, race or insurance status the patients are. However, graphical diagnostic for the proportional hazard assumption shows severe departure from the proportionality assumption for each covariate, indicating that proportional hazards model may be inadequate to describe the effect of these covariates on life length. (Figure 3)

An alternative way to examine the effect of a specific risk factor is to divide up the population according to the risk factor and examine the subpopulations in terms of their mean residual lifetimes. We split the population by gender, race and insurance coverage, respectively, and examine the difference of mean residual life length between those with and without mental illness. Results show that the effect of mental disease on excess mortality still exists, but becomes weaker in certain sub-groups and stronger in some other sub-groups. Table 6 summarizes the mean residual life comparison at six age points for mental disease and non-mental-disease population within male and female. For male and female, the differences between two populations are not statistically significant. Notice that female always have slightly longer expected residual lifetime than male no matter of their age and mental health status. Within race sub-groups, the mental health status doesn't make difference on the expected remaining life length (Table 7). No statistically significant result is shown for mean residual life comparison within white, black and others groups. Mental disease insurance coverage turns out to be a contributing factor to premature mortality (Table 8). For the population without mental disease insurance coverage, there are statistically significant differences of mean residual life

between mental illness patients and healthy individuals ($p=0.02$ at age 60, $p=0.004$ at age 70 and $p<0.001$ at age 80). The difference is not significant for those who are covered by mental disease insurance.

4. Discussion

The purpose of this study is to demonstrate an alternative measure of excess mortality of individuals with mental disease in a population-based, nationally representative sample. We compared mean residual lifetimes for people with and without mental illness at given ages, and examined the existence of significant difference. The finding in the study is consistent with previous research conclusions that persons with mental illness have shorter expected remaining life length compared with those in the general population [5, 16, 18]. We find 2-fold increase in the risk of mortality for individuals with mental illness from the proportional hazard model. From the perspective of mean residual lifetime, mentally ill people have as many as 7 years shorter residual life length compared to general population. Previous studies have also suggested that natural causes of deaths make up the majority of mortality among mental disease patients. This is true in the study presented here as well.

Druss et al. [18] has calculated the years of potential lost (YPLL) for those who died in this study population. Nevertheless, as we have pointed out, it is a biased result because it only includes information about decedents and ignores the censored observations. This point can be attested if we compare the YPLL and the mean residual

lifetimes. The YPLL for decedents with mental disorder is 11.3±0.6, for decedents without mental disorder is 5.9±0.1. The difference is 5.4 years. In terms of mean residual life, the difference between these two populations changes from 7.6 years at age 30 to 2 years at age 80.

Inconsistent results are found from the two statistical methods regarding the effects of potential risk factors--gender, race and mental disease insurance. Proportional hazard models show that relative hazard rate of mortality stays approximately same before and after adjusting for these risk factors. Mental disease always doubles the risk of death, no matter what gender, race and health insurance status the patients are. Whereas the mean residual life comparison between mental disordered people and general population becomes not significant when we only look at the male/female and white/black/others subpopulation. This incongruity may be caused by the differential underlying assumption of these two methods: proportional hazard model assumes the hazard functions for the two subgroups are proportional over time, while mean residual life comparison only looks at a certain time point. The mean residual life lengths comparison makes more sense in this study setting, because the proportional hazard assumption can hardly stand over the whole life course. The only congruency appears in the subgroup where people are uncovered by mental insurance. Those who don't have insurance on mental disease are at higher risk of mortality and have shorter residual life at age 50 or older. This finding suggests that lack of financial protection and access healthcare is one reason of the mortality gap for persons with mental disorders.

Most studies examining excess mortality have study populations that are typically from a small number of communities or states. Studies have examined premature mortality within community-based samples, which are mental disease patients seeking treatment at community mental health centers[16, 33]. The study presented here expanded previous work by measuring excess mortality in a nationally representative sample. The use of treated samples may present bias in the analysis because these populations have more access to health services than the general population. It would be difficult to assess the role of healthcare in these studies. These subjects also tend to be poor than patients in hospitals residence, making it challenging to fully adjust for socioeconomic status factors. Multistate studies also lack nationally representativeness because the properties of states such as population composition and economic level would obscure possible influence factors.

This study also makes contribution by estimating the expected remaining life and comparing between groups at any given age. Instead of looking at premature mortality over all ages, our method is able to quantify the difference of mean residual life at any age of interest. Thus, if we are interested in examining any age-specific effect of mental disease on excess death, the mean residual life comparison method is a good choice. These advantages enable researchers and policy makers to better understand the impact of mental illness and the age trend of the impact, which will help to establish public health priorities at patients' different ages.

Several limitations exist of this study. First, all two-sample mean residual life

comparisons are conducted at a specific age. Bathke et al. [34] proposed a method to combine multiple tests in censored data using empirical likelihood method. Future researchers could combine the two-sample tests at several ages to form an overall two-sample test. Second, it would be of interest to have the proportional mean residual life model analogous to proportional hazards model, which can examine the difference of mean residual lifetimes between the two populations adjusting for covariates. There is some limited work about the proportional mean residual life model [35, 36], but we didn't implement it due to lack of software. Third, the method will fail to compute median residual life if the survival curve does not fall down to 0.5 (50%). In this case, only mean residual lifetime can be utilized. Fourth, the power of two-sample mean residual life test within subgroups is probably weak owing to smaller sample size. When we divide up the population and only examine the effect of mental illness in specific subgroups, we may have insufficient sample size to detect the difference of mean residual life between group with and without mental disease.

5. Conclusions

We find that individuals with mental illness have shorter remaining life expectancy compared to those without mental illness in a nationally representative population. Some potential risk factors, such as mental health insurance coverage, are related to the life expectancy gap. The study finding suggests that the mean residual life can be adopted as

a desirable measure to quantify premature mortality and better understand the burden of mental disease at various ages and within specific subgroups. It is important for policymakers to prioritize resources to those at most risk.

Reference

1. WHO. *The global burden of mental disorders and the need for a comprehensice, coordinated response from health and social sectors at the country level*.
2. Kessler, R.C., et al., *Severity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication Adolescent Supplement.* Arch Gen Psychiatry, 2012. **69**(4): p. 381-9.
3. Babigian, H.M. and C.L. Odoroff, *The mortality experience of a population with psychiatric illness.* Am J Psychiatry, 1969. **126**(4): p. 470-80.
4. Brown, A.S. and J. Birthwhistle, *Excess mortality of mental illness.* Br J Psychiatry, 1996. **169**(3): p. 383-4.
5. Dembling, B.P., D.T. Chen, and L. Vachon, *Life expectancy and causes of death in a population treated for serious mental illness.* Psychiatr Serv, 1999. **50**(8): p. 1036-42.
6. Davidson, S., et al., *Cardiovascular risk factors for people with mental illness.* Aust N Z J Psychiatry, 2001. **35**(2): p. 196-202.
7. Kendrick, T., *Cardiovascular and respiratory risk factors and symptoms among general practice patients with long-term mental illness.* Br J Psychiatry, 1996. **169**(6): p. 733-9.
8. Sartorius, N., *Physical illness in people with mental disorders.* World Psychiatry, 2007. **6**(1): p. 3-4.
9. Newcomer, J.W. and C.H. Hennekens, *Severe mental illness and risk of cardiovascular disease.* JAMA, 2007. **298**(15): p. 1794-6.
10. Roshanaei-Moghaddam, B. and W. Katon, *Premature mortality from general medical illnesses among persons with bipolar disorder: a review.* Psychiatr Serv, 2009. **60**(2): p. 147-56.
11. Brown, S., *Excess mortality of schizophrenia. A meta-analysis.* Br J Psychiatry, 1997. **171**: p. 502-8.
12. Benjamin G. Druss, E.R.W., *Mental disorders and medical comorbidity.* Research Synthesis Report, 2011. **No. 21**.
13. *Gender and women's mental health*. Available from: http://www.who.int/mental_health/prevention/genderwomen/en/.
14. David T. Takeuchi, D.R.W., *Race, ethnicity and mental health: Introduction to the special issue.* Journal of Health and Social Behavior, 2003. **Vol 44**: p. 233-236.
15. John W. Gardner, J.S.S., *Years of Potential Life Lost (YPLL)-- what does it measure.* Epidemiology, 1990. **1**: p. 322-329.
16. Piatt, E.E., M.R. Munetz, and C. Ritter, *An examination of premature mortality among decedents with serious mental illness and those in the general population.* Psychiatr Serv, 2010. **61**(7): p. 663-8.
17. Colton, C.W. and R.W. Manderscheid, *Congruencies in increased mortality*

rates, years of potential life lost, and causes of death among public mental health clients in eight states. Prev Chronic Dis, 2006. **3**(2): p. A42.

18. Druss, B.G., et al., *Understanding excess mortality in persons with mental illness: 17-year follow up of a nationally representative US survey.* Med Care, 2011. **49**(6): p. 599-604.
19. Barker, P.R., et al., *Serious mental illness and disability in the adult household population: United States, 1989.* Adv Data, 1992(218): p. 1-11.
20. F Guess , F.P., *Mean Residual Life: Theory and Applications*, 1985.
21. Bandos, H., *Regression on median residual life function for censored survival data*, in *Department of Biostatistics*2007, University of Pittsburgh.
22. Zhou, M. and J.H. Jeong, *Empirical likelihood ratio test for median and mean residual lifetime.* Stat Med, 2011. **30**(2): p. 152-9.
23. A, O., *Empirical Likelihood*. 2001, London: Chapman & Hall.
24. Thomas DR, G.G., *Confidence interval estimation of survival probabilities for censored data.* Journal of the American Statistical Association, 1975. **70**: p. 865-871.
25. A, O., *Empirical likelihood ratio confidence intervals for a single functional.* Biometrika, 1988. **75**: p. 237-249.
26. Pan, X.-R. and M. Zhou, *Empirical Likelihood Ratio in Terms of Cumulative Hazard Function for Censored Data.* Journal of Multivariate Analysis, 2002. **80**(1): p. 166-188.
27. Jeong, J.H., S.H. Jung, and J.P. Costantino, *Nonparametric inference on median residual life function.* Biometrics, 2008. **64**(1): p. 157-63.
28. Zhao, Y. and G. Qin, *Inference for the Mean Residual Life Function via Empirical Likelihood.* Communications in Statistics - Theory and Methods, 2006. **35**(6): p. 1025-1036.
29. McLain, A.C. and S.K. Ghosh, *Nonparametric estimation of the conditional mean residual life function with censored data.* Lifetime Data Anal, 2011. **17**(4): p. 514-32.
30. Kleinbaum, D.G., Klein, Mitchel, *Survival Analysis*. 3rd ed. 2011: Springer.
31. P. Grambsch, T.T., *Proportional hazards tests and diagnostics based on weighted residuals.* Biometrika, 1994. **81**: p. 515-526.
32. Zhou, M., *emplik: Empirical likelihood ratio for censored/truncated data. R package version 0.9-8-2.* http://CRAN.R-project.org/package=emplik, 2012.
33. Bruce, M.L. and P.J. Leaf, *Psychiatric disorders and 15-month mortality in a community sample of older adults.* Am J Public Health, 1989. **79**(6): p. 727-30.
34. Arne Bathke, M.-O.K., Mai Zhou, *Combined multiple testing by censored empirical likelihood.*
35. David Oakes, T.D., *A note on residual life.* Biometrika, 1990. **77(2)**: p. 409-410.
36. Chen, Y.Q., et al., *Semiparametric estimation of proportional mean residual life*

model in presence of censoring. Biometrics, 2005. **61**(1): p. 170-8.

Appendix A. Tables

Table 1. Demographic and Clinical Characteristics

Characteristics	Mental Disorder (n=1725)	No Mental Disorder (n=79125)	P❖
Age, years	42.4±0.5	43.8±0.2	0.007
Male sex, n(%)	772(45.9)	36,655(47.6)	0.21
Race			0.24
White, n(%)	1399(84.2)	64,890(84.2)	
Black, n(%)	262(11.8)	10,576(10.9)	
Other, n(%)	64(4.0)	3659(4.9)	
Mental health insurance, n(%)	610(64.1)	28,773(71.7)	<0.001

❖p-value for design-adjusted statistical test

Table 2. Decedents Descriptions and Cause of Death

	Mental Disorder (n=463)	No Mental Disorder (n=15,972)
Number of decedents between 1989 and 2006, n(% of the whole population)	463(26.8% of 1725)	15,972(20.2% of 79,125)
Unnatural cause of death, n(%)	25(5.4%)	714(4.7%)
Median survival age, years	77.5±1.2	84.8±0.1

Note: In Druss et al., the years of potential life lost (YPLL) is calculated for decedents. For those with mental disorder, the YPLL is 11.3±0.6; for those have no mental disorder, the YPLL is 5.9±0.1. These are not reported in the table because they are biased measure.

Table 3. Probability of Survival based on Kaplan-Meier Curve, with 95% Confidence Interval

Age	Mental Disorder (n=1725)	No Mental Disorder (n=79125)
30	0.998(0.996-1)	0.999(0.998-0.999)
40	0.986(0.981-0.992)	0.994(0.994-0.995)
50	0.945(0.933-0.957)	0.982(0.981-0.983)
60	0.868(0.848-0.888)	0.952(0.950-0.954)
70	0.686(0.653-0.720)	0.870(0.867-0.874)
80	0.420(0.377-0.468)	0.661(0.655-0.668)
90	0.149(0.111-0.201)	0.294(0.286-0.302)

Table 4. Residual Life Comparison Between Individuals With and Without Mental Disorders

Age	Mean Residual Life			Median Residual Life		
	Mental Disorder (n=1725)	No Mental Disorder (n=79125)	P❖	Mental Disorder (n=1725)	No Mental Disorder (n=79125)	P❖
30	45.9	53.5	0.069	47.5	55.0	0.176
40	36.4	43.7	0.044	37.5	45.0	0.125
50	27.7	34.1	0.051	28.0	35.0	0.090
60	19.7	25.0	0.034	19.5	25.5	0.123
70	13.3	16.8	0.063	12.5	16.5	0.160
80	8.4	10.3	0.197	7.7	9.2	0.080
90	4.2	6.4	0.258	3.3	4.3	0.310

❖p-value for two-sample mean residual lifetimes statistical test

Table 5.1. Univariate Effect on Survival, Hazards Ratio with 95% Confidence Interval

Covariates	Univariate
Mental illness	2.09 (1.90-2.29)
Female sex	0.62 (0.60-0.64)
Race	
Black	1.32 (1.26-1.38)
Others	0.90 (0.82-0.99)
No mental insurance coverage	1.72 (1.61-1.83)

❖ All covariates contribute significantly to the Cox regression model

Table 5.2. Multivariate Effect of Mental Illness on Survival, Hazards Ratio with 95% Confidence Interval

Covariates	Adjusted for sex	Adjusted for race	Adjusted for insurance	Adjusted for all covariates
Mental illness	2.12 (1.93-2.33)	2.08 (1.91-2.29)	2.04 (1.86-2.23)	2.07 (1.88-2.27)
Female sex	0.62 (0.60-0.64)			0.63 (0.61-0.65)
Race				
Black		1.32 (1.26-1.38)		1.31 (1.25-1.36)
Others		0.90 (0.82-0.99)		0.84 (0.76-0.93)
No mental insurance coverage			1.70 (1.60-1.81)	1.67 (1.56-1.77)

❖ All covariates contribute significantly to the Cox regression model

Table 6. Mean Residual Life Comparison Between Individuals With and Without Mental Disorders, by Gender

	Male			Female		
Age	Mental Disorder (n=772)	No Mental Disorder (n=36655)	P	Mental Disorder (n=953)	No Mental Disorder (n=42470)	P
30	42.7	50.8	0.129	48.7	55.4	0.238
40	33.2	41.1	0.096	39.1	45.6	0.210
50	24.9	31.6	0.093	30.0	35.9	0.191
60	16.8	22.6	0.079	22.1	26.7	0.191
70	10.2	14.6	0.109	15.7	18.2	0.262
80	6.3	8.6	0.285	9.5	11.1	0.318

❖p-value for two-sample mean residual lifetimes statistical test

Table 7. Mean Residual Life Comparison Between Individuals With and Without Mental Disorders, by Race

	White			Black			Others		
Age	Mental Disorder (n=1399)	No Mental Disorder (n=64890)	P	Mental Disorder (n=262)	No Mental Disorder (n=10576)	P	Mental Disorder (n=64)	No Mental Disorder (n=3659)	P
30	46.1	53.8	0.10	44.5	51.0	0.50	45.9	55.0	0.57
40	36.6	43.9	0.06	35.0	41.4	0.46	35.9	45.3	0.52
50	28.0	34.3	0.08	26.3	32.2	0.44	28.1	35.7	0.53

60	19.8	25.1	0.05	18.9	23.5	0.45	19.8	26.7	0.50
70	13.1	16.8	0.10	14.7	16.1	0.55	13.1	18.7	0.45
80	8.2	10.2	0.24	8.9	10.4	0.56	8.3	11.6	0.07

❖p-value for two-sample mean residual lifetimes statistical test

Table 8. Mean Residual Life Comparison Between Individuals With and Without Mental Disorders, by Mental Health Insurance Coverage

	Mental Illness Insured			Mental Illness Uninsured		
Age	Mental Disorder (n=610)	No Mental Disorder (n=28773)	P	Mental Disorder (n=1115)	No Mental Disorder (n=50352)	P
30	46.3	52.9	0.194	40.2	50.3	0.09
40	36.6	43.1	0.154	30.8	40.8	0.06
50	27.5	33.5	0.128	22.4	31.7	0.05
60	18.6	24.4	0.10	14.4	23.5	0.02
70	12.4	16.4	0.105	8.7	17.1	0.004
80	7.8	10.1	0.175	3.2	12.8	<0.001

❖p-value for two-sample mean residual lifetimes statistical test

Appendix B. Figures

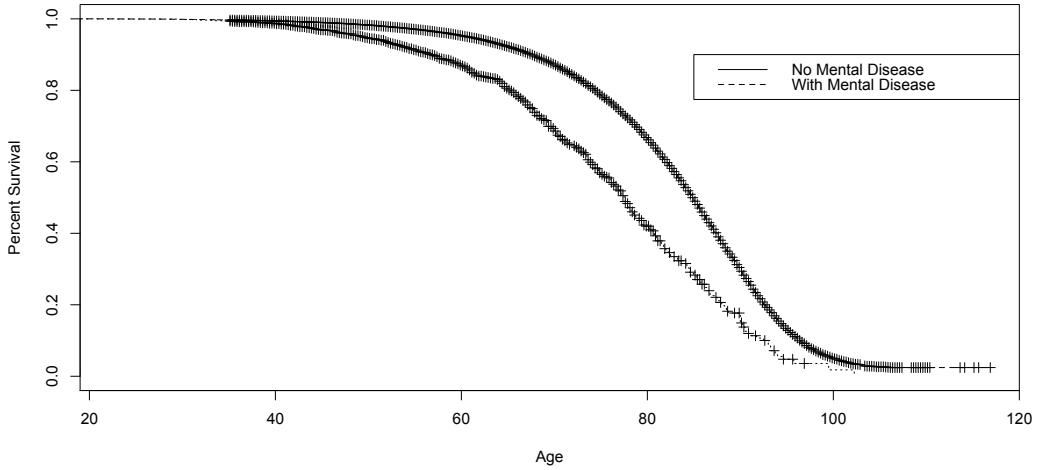


Figure 1. Comparison of survival estimates among individuals with and without mental disease

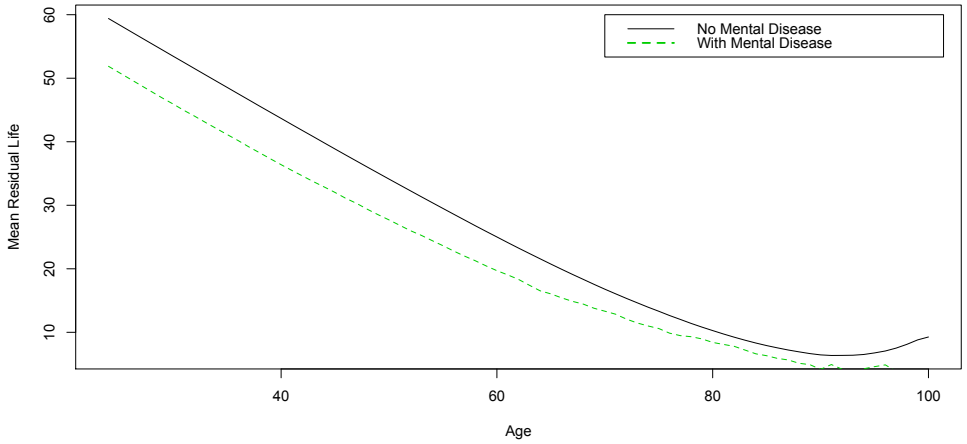
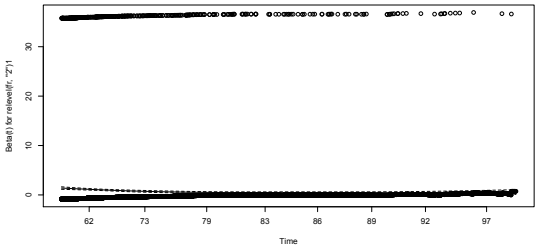
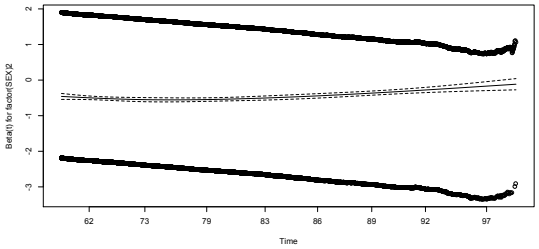


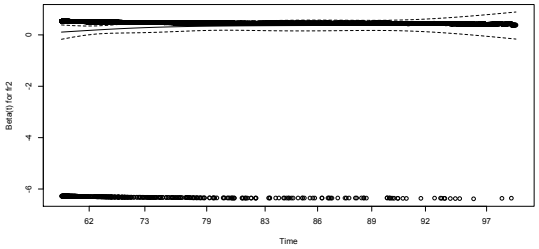
Figure 2. Mean residual life estimation for populations with and without mental disease



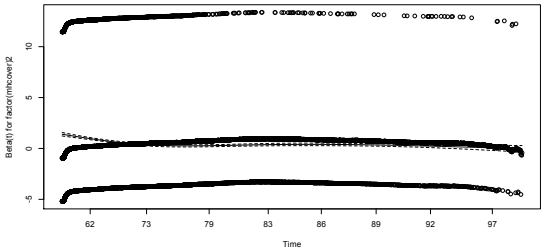
Mental illness



Female gender



Black race



No mental insurance

Figure 3. Graphical diagnostic of proportional hazards assumption for each covariate

Appendix C. R Code Samples

1. R Codes for mean residual lifetime estimation and two-sample comparison, generating Table 4, Figure 2 and Table 5

```
nhisdata<-read.csv("/Users/Tracy/Desktop/study/Thesis/data_9var.csv",he
 ader=T)
healthsub<-subset(nhisdata,anymental==2)
illsub<-subset(nhisdata,anymental==1)

#######    for the whole population, Table 4    ##########
library(emplik)
Age<-seq(24,100)
MRLhealth<-rep(1,77)
MRLill<-rep(1,77)
p_value<-rep(1,77)
mygfun<-function(s,age,muage){as.numeric(s>=age)*(s-(age+muage))}
for(AGE in 24:100){
     temp<-WKM(healthsub$end_age,healthsub$MORTSTAT)
     tivec<-temp$times
     pivec<-temp$jump
     pivec[tivec<AGE]<-0
     Sage<- sum(pivec)
     fenzi<-sum((tivec-AGE)*pivec)
     MRtime1<-fenzi/Sage
     MRLhealth[AGE-23]=MRtime1

 temp<-WKM(illsub$end_age,illsub$MORTSTAT)
     tivec<-temp$times
     pivec<-temp$jump
     pivec[tivec<AGE]<-0
     Sage<- sum(pivec)
     fenzi<-sum((tivec-AGE)*pivec)
     MRtime2<-fenzi/Sage
     MRLill[AGE-23]=MRtime2

     if (AGE%%10==0 & AGE!=100){
     samemr=(MRtime1+MRtime2)/2
     theta<-seq(samemr-5,samemr+5)
     W1<-rep(1000,length(theta))
     W2<-rep(1000,length(theta))
```

```
     U<-rep(1000,length(theta))
     for(i in 1:length(theta)){

     W1[i]<-el.cen.EM2(x=healthsub$end_age,d=healthsub$MORTSTAT,fun=mygf
 un,mu=0,age=AGE,muage=theta[i])$`-2LLR`

     W2[i]<-el.cen.EM2(x=illsub$end_age,d=illsub$MORTSTAT,fun=mygfun,mu=
 0,age=AGE,muage=theta[i])$`-2LLR`
          U[i]<-(W1[i]+W2[i])/theta[i]
          }
     Ustat=min(U)
    p_value[AGE-23]=pchisq(Ustat,1,lower.tail=F)
     }
}
table<-data.frame(cbind(Age,MRLhealth,MRLill,p_value))


##########    mean residual life plot, Figure 2    ######
plot(Age,MRLhealth,xlab="Age",ylab="Mean Residual Life",type="l",lty=1)
lines(Age,MRLill,type="l",col=3,lty=2)
legend(70,60,c("No Mental Disease","With Mental
 Disease"),lty=c(1,2),lwd=c(1,2,1.5,2.5),col=c(1,3))

#########  by sex, Table 5  ##################
menhealthsub<-subset(healthsub,SEX==1)
womenhealthsub<-subset(healthsub,SEX==2)
menillsub<-subset(illsub,SEX==1)
womenillsub<-subset(illsub,SEX==2)
Age<-seq(24,100)
MRLhealth<-rep(1,77)
MRLill<-rep(1,77)
p_value<-rep(1,77)
mygfun<-function(s,age,muage){as.numeric(s>=age)*(s-(age+muage))}
for(AGE in 24:100){
     temp<-WKM(menhealthsub$end_age,menhealthsub$MORTSTAT)
     tivec<-temp$times
     pivec<-temp$jump
     pivec[tivec<AGE]<-0
     Sage<- sum(pivec)
     fenzi<-sum((tivec-AGE)*pivec)
     MRtime1<-fenzi/Sage
```

```
    MRLhealth[AGE-23]=MRtime1

temp<-WKM(menillsub$end_age,menillsub$MORTSTAT)
    tivec<-temp$times
    pivec<-temp$jump
    pivec[tivec<AGE]<-0
    Sage<- sum(pivec)
    fenzi<-sum((tivec-AGE)*pivec)
    MRtime2<-fenzi/Sage
    MRLill[AGE-23]=MRtime2

    if (AGE%%10==0 & AGE!=100){
    samemr=(MRtime1+MRtime2)/2
    theta<-seq(samemr-5,samemr+5)
W1<-rep(1000,length(theta))
    W2<-rep(1000,length(theta))
    U<-rep(1000,length(theta))
    for(i in 1:length(theta)){

    W1[i]<-el.cen.EM2(x=menhealthsub$end_age,d=menhealthsub$MORTSTAT,fu
n=mygfun,mu=0,age=AGE,muage=theta[i])$`-2LLR`

    W2[i]<-el.cen.EM2(x=menillsub$end_age,d=menillsub$MORTSTAT,fun=mygf
un,mu=0,age=AGE,muage=theta[i])$`-2LLR`
         U[i]<-(W1[i]+W2[i])/theta[i]
         }
    Ustat=min(U)
   p_value[AGE-23]=pchisq(Ustat,1,lower.tail=F)
    }
}
table_men<-data.frame(cbind(Age,MRLhealth,MRLill,p_value))

MRLhealth<-rep(1,77)
MRLill<-rep(1,77)
p_value<-rep(1,77)
mygfun<-function(s,age,muage){as.numeric(s>=age)*(s-(age+muage))}
for(AGE in 24:100){
    temp<-WKM(womenhealthsub$end_age,womenhealthsub$MORTSTAT)
    tivec<-temp$times
    pivec<-temp$jump
    pivec[tivec<AGE]<-0
```

```
    Sage<- sum(pivec)
    fenzi<-sum((tivec-AGE)*pivec)
    MRtime1<-fenzi/Sage
    MRLhealth[AGE-23]=MRtime1

temp<-WKM(womenillsub$end_age,womenillsub$MORTSTAT)
    tivec<-temp$times
    pivec<-temp$jump
    pivec[tivec<AGE]<-0
    Sage<- sum(pivec)
    fenzi<-sum((tivec-AGE)*pivec)
    MRtime2<-fenzi/Sage
    MRLill[AGE-23]=MRtime2

    if (AGE%%10==0 & AGE!=100){
    samemr=(MRtime1+MRtime2)/2
    theta<-seq(samemr-5,samemr+5)
W1<-rep(1000,length(theta))
    W2<-rep(1000,length(theta))
    U<-rep(1000,length(theta))
    for(i in 1:length(theta)){

    W1[i]<-el.cen.EM2(x=womenhealthsub$end_age,d=womenhealthsub$MORTSTA
T,fun=mygfun,mu=0,age=AGE,muage=theta[i])$`-2LLR`

    W2[i]<-el.cen.EM2(x=womenillsub$end_age,d=womenillsub$MORTSTAT,fun=
mygfun,mu=0,age=AGE,muage=theta[i])$`-2LLR`
         U[i]<-(W1[i]+W2[i])/theta[i]
         }
    Ustat=min(U)
   p_value[AGE-23]=pchisq(Ustat,1,lower.tail=F)
    }
}
table_women<-data.frame(cbind(Age,MRLhealth,MRLill,p_value))
```

2. R Codes for median residual lifetime estimation and two-sample comparison, generating Table 4

```
library(emplik)
Age<-seq(24,100)
MedRhealth<-rep(1,77)
```

```
MedRill<-rep(1,77)
p_value<-rep(1,77)
mygfun2<-function(s,age,mdage){as.numeric(s<=(age+mdage))-0.5*as.numeri
 c(s<=age)-0.5}
for(AGE in 24:100){
     temp<-WKM(healthsub$end_age,healthsub$MORTSTAT)
     tivec<-temp$times
     pivec<-temp$jump
     pivec[tivec<AGE]<-0
     Sage<- sum(pivec)
 Ptheta<-Sage/2
 Cprob<-cumsum(pivec)
 posi<-sum(Cprob<Ptheta)
 theta1<-tivec[posi+1]
 MedRhealth[AGE-23]=theta1-AGE

 temp<-WKM(illsub$end_age,illsub$MORTSTAT)
     tivec<-temp$times
     pivec<-temp$jump
     pivec[tivec<AGE]<-0
     Sage<- sum(pivec)
     Ptheta<-Sage/2
 Cprob<-cumsum(pivec)
 posi<-sum(Cprob<Ptheta)
 theta2<-tivec[posi+1]
 MedRill[AGE-23]=theta2-AGE

     if (AGE%%10==0 & AGE!=100){
     samemr=(theta1+theta2)/2-AGE
     theta<-seq(samemr-5,samemr+5,0.25)
 W1<-rep(1000,length(theta))
     W2<-rep(1000,length(theta))
     U<-rep(1000,length(theta))
     for(i in 1:length(theta)){

     W1[i]<-el.cen.EM2(x=healthsub$end_age,d=healthsub$MORTSTAT,fun=mygf
 un2,mu=0,age=AGE,mdage=theta[i])$`-2LLR`

     W2[i]<-el.cen.EM2(x=illsub$end_age,d=illsub$MORTSTAT,fun=mygfun2,mu
 =0,age=AGE,mdage=theta[i])$`-2LLR`
          U[i]<-(W1[i]+W2[i])/theta[i]
```

```
        }
    Ustat=min(U)
   p_value[AGE-23]=pchisq(Ustat,1,lower.tail=F)
    }
}
   table_median<-data.frame(cbind(Age,MedRhealth,MedRill,p_value))
```