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:

_____ _____

Yi Wang Date

The Role of Vitamin D in the All-Cause Mortality of General Population

By

Yi Wang

MPH

Emory University

Rollins School of Public Health

Department of Biostatistics

_____ [Chair's signature]

Limin Peng

Committee Chair

_____ [Member's signature]

Vin Tangpricha

Committee Member

The Role of Vitamin D in the All-Cause Mortality of General Population

By

Yi Wang

BMgt, Fudan University

2011

MPH, Emory University

Rollins School of Public Health

2015

Thesis Committee Chair: Limin Peng, 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 Public Health

in Department of Biostatistics

2015

Abstract

The Role of Vitamin D in the All-Cause Mortality of General Population

By Yi Wang

This thesis is focused on investigating the association between vitamin D deficiency and all-cause mortality. We conducted analyses among a retrospective cohort of 1159 patients, who were admitted to Emory hospital from 2008-2012 and had vitamin D tests. Of the1159 subjects, 77% had vitamin D deficiency (25-hydroxyvitamin D (25(OH)D) <30 ng/mL); 44% had chronic kidney disease (CKD) (defined as estimated glomerular filtration rate (eGFR) <=60 mL/min/1.73 m2); and 380 died during an average follow-up of 427.3 days. We conducted survival analysis with the primary time-to-event outcome specified as time from admission to death. Based on multivariate Cox regression, we find that older age and lower eGFR are significantly associated with a higher risk of mortality (P<0.001 for age; P<0.001 for eGFR). Although no statistically significant association between vitamin D deficiency and mortality is found after adjusting for age and eGFR (Hazard ratio, 1.049; 95% confidence interval, 0.822-1.337), our multivariate analysis detects a significant interaction effect between vitamin D deficiency and eGFR on mortality risk (P=0.0208). A further analysis stratified on vitamin D deficiency status suggests that a decrease in eGFR in patients with vitamin D deficiency has a significant negative impact on mortality risk (Hazard ratio, 0.997; 95% confidence interval, 0.994-1.000), while the effect of eGFR diminishes in patients with normal vitamin D level.

The Role of Vitamin D in the All-Cause Mortality of General Population

By

Yi Wang

BMgt, Fudan University

2011

MPH, Emory University

Rollins School of Public Health

2015

Thesis Committee Chair: Limin Peng, 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 Public Health

in Department of Biostatistics

2015

Acknowledgements

I would like to express my deepest gratitude to my advisor, Professor Limin Peng, who provided me the opportunity to experience the research of survival analysis in the field and practical issues beyond the textbooks. Also a thanks for her excellent guidance, patience and support to help me write the thesis. Special thanks goes to Dr. Vin Tangpricha, who was willing to take the time to help me understand the medical background of my thesis and read my thesis. I would never have been able to finish my thesis without their help.

I would also like to thank the faculty and staff of the Biostatistics Department at Rollins School of Public Health for the amazing pasted two years of learning. This thesis is only a sample of the vast knowledge that was attained and applied through my two years here at Rollins.

Finally I would like to give a special thanks to my family, my dear friends Lei Yang and Yue Chu. They motivated me to pursue a degree at the graduate level and always stand by me with their loving support and encouragement.

Table of Contents

I Introduction

1.1 Vitamin D Function Mechanism

It is generally understood that vitamin D has both classic and non-classic functions in the body. The classic function of vitamin D is via the endocrine system. In this system, vitamin D is converted to 25-hydroxyvitamin D (25(OH)D) in the liver, with subsequent conversion of 25(OH)D to calcitriol (1,25(OH)2D) in the kidney, to maintain normal calcium homeostasis by promoting intestinal calcium absorption and resorption of calcium from bone. The non-classic function of vitamin D is via the autocrine/paracrine system to facilitate expression of genes not related to calcium homeostasis. Through the local production of 1,25(OH)2D from 25(OH)D in non-calcium regulating tissues such as the colon, prostate, and breast, vitamin D is able to regulate various tissue responses, including regulating cell growth and preventing cancer progression. [1] [2]

1.2 Source of Vitamin D

The major natural source of vitamin D is exposure to sunlight, whereas only a small part of vitamin D intake comes from daily-dietary intake since few foods contain vitamin D naturally.[3] Therefore, vitamin D deficiency is mainly associated with factors that impair or prevent cutaneous production of vitamin D [4] [5] Previous studies also shows that aging is associated with the result of decreased concentration of vitamin D in the skin;[6] Obesity is also a common cause of deficiency due to the fat sequestration of vitamin D.[7]

1.3 Vitamin D Deficiency

1.3.1 Vitamin D Status Classification

Generally, the serum concentration of 25(OH)D is used to classify the vitamin D status. [8] However, there is no general agreement on the required serum 25(OH)D for an adequate vitamin D status. Most investigators agree that serum 25(OH)D concentration of <= 74 nmol/L, or 29

ng/mL, is considered to indicate vitamin D deficiency, whereas concentration >= 75 nmol/L, or 30 ng/mL is considered to be sufficient.[9] Some UK studies also indicated that individuals with symptomatic osteomalacia or rickets have serum 25(OH)D concentrations of less than 25 nmol/L, or 10 ng/mL, reflecting profound vitamin D deficiency.[10] [11]

1.3.2 Vitamin D Deficiency in General Population

Nowadays, vitamin D deficiency has been an increasing public health problem with a high prevalence of general population over the world. [3] A study investigating the data from the 2005 to 2006 National Health and Nutrition Examination Survey (NHANES) showed that the overall prevalence rate of vitamin D deficiency among 4495 US adult participants was 41.6% (defined as serum 25(OH)D ≤20 ng/mL), with a mean 25(OH)D level of 49.8 nmol/L.[12] An earlier study using previous NHANES data showed a mean 25(OH)D level of 75 nmol/L in 1988–1994 (n=18883) and a mean 25(OH)D level of 60nmol/L in 2001–2004 (n= 13369), indicating the decrease of average 25(OH)D level in the US general population.[13] Considering the risk factors associated with the vitamin D status in the body, possible factors that have likely contributed to the observed increasing trend of vitamin D deficiency could be the increased sunscreen use, the decreased outdoor activity,[14] and the increased obesity rate in US population.[7] [15]

1.3.3 Consequences of Vitamin D Deficiency on Morbidity and Mortality

In general population, vitamin D deficiency is a well-known factor of osteoporotic diseases.[16] [17] Furthermore, vitamin D deficiency might contribute to development and progression of various other common chronic diseases such as hypertension,[18] cardiovascular disease,[19] cancers like prostate,[20] breast,[21] lung[22] and several autoimmune conditions.[23] An inverse association between vitamin D status and all-cause mortality has been established in prospective cohort studies.[24] [25] Studies focusing on the cause-specific mortality reported significant inverse associations between vitamin D status and death caused by cardiovascular diseases, cancer,

diseases of the respiratory system, the digestive system, and by endocrine, nutritional and metabolic diseases.[25] [26]

1.4 Vitamin D Status in Patients with Chronic Kidney Disease

Vitamin D deficiency is common in most patients with chronic kidney disease (CKD). CKD patients have a significantly higher prevalence of vitamin D deficiency, that is, serum 25(OH)D level < 30 ng/mL.[27] [28] Possible reasons resulting in this situation of CKD patients include increased catabolism of circulating 25(OH)D; reduced conversion of circulating 25(OH)D to the more active vitamin D metabolite (1,25(OH)2D); and limited UVB-induced vitamin D synthesis in the skin with decreased outdoor activity.[19, 29-31] On the other hand, vitamin D deficiency itself may contribute to impaired kidney function.[28] [32]

1.5 Chronic Kidney Disease (CKD) Definition and Stages

According to the National Kidney Foundation Kidney Disease Outcomes Quality Initiative, CKD is defined based on the combination of kidney damage (commonly quantified using albuminuria) and the level of kidney function (quantified as glomerular filtration rate (GFR) estimated from the serum creatinine concentration (S_{cr})). Individual either with kidney damage or with an estimated GFR (eGFR) <60 mL/min/1.73 m^2 for ≥3 months are classified as having CKD.

The CKD stages are based on the level of kidney function, irrespective of diagnosis, as follows: stage 1, persistent albuminuria with an eGFR higher than 90 mL/min/1.73 m^2; stage 2, persistent albuminuria with an eGFR of 60 to 89 mL/min/1.73 m^2; stage 3, an eGFR of 30 to 59 mL/min/1.73 m^2; stage 4, an eGFR of 15 to 29 mL/min/1.73 m^2; and stage 5, an eGFR less than 15 mL/min/1.73 m^2.[33]

1.6 Previous Studies Related to the Impact of Vitamin D deficiency on the Mortality of CKD Patients

A meta-analysis of randomized controlled trials showed using vitamin D supplementation resulting in decreases in total mortality rates among frail elderly patients.[34] Previous observational studies showed, but not consistently, that higher serum 25(OH)D level are associated with significantly decreased mortality risk of CKD patients.[28, 35-43] A meta-analysis of 10 prospective studies (1 case-control and 9 cohort studies) concluded a summary estimated dose-response relative risk for the association of 25(OH)D level and mortality risk in CKD patients is 0.86 (95% CI: 0.82-0.91).[44]

1.7 An Overview

In this thesis, we investigate the association between vitamin D deficiency and all-cause mortality through analyzing a retrospective cohort study, which included 1159 patients who were admitted to Emory hospital from 2008-2012 with vitamin D test records. In addition to vitamin D lab results, the dataset contains information on vital status, including eGFR results based on the Modification of Diet in Renal Disease (MDRD) formula, age, gender and relative disease diagnosis records. Of the 1159 subjects, 866(77%) had vitamin D deficiency (25(OH)D<30 ng/mL) and 512(44%) had CKD (defined as eGFR<=60 mL/min/1.73 m^2). Given 44% of the study subjects had CKD, we are interested in elucidating whether vitamin D deficiency are associated with all-cause mortality as well as whether the association is significant and modified among CKD patients. By separately examining each individual risk factor on mortality and then fitting a multivariate Cox regression model on mortality with vitamin D deficiency and other covariates, we investigate how vitamin D deficiency impacts all-cause mortality, We also use this dataset to evaluate the potential interacted effect between vitamin D deficiency and CKD status on mortality based on a multivariate Cox regression model and stratified Cox regression analyses.

II Methods

2.1 Data Description

The data were collected from 1159 patients admitted to Emory hospital with a vitamin D test from 2008 to 2012. The actual laboratory test of 25(OH)D level (ng/mL) was used as the measurement of vitamin D result. Our primary outcome of interest is all-cause mortality. Emory medical records and Social Security Death Index (SSDI) file were linked to our study cohort to obtain the decease status of subjects in our study. Survival times were calculated from the date of hospitalization date either until the date of death or until January 2012. The subjects who were alive after January 2012 were considered as censored subjects in the survival analysis.

The eGFR (mL/min/1.73 m^2) was calculated based on the Modification of Diet in Renal Disease (MDRD) formula, which was developed by Levey et al. in 1999 using 1628 subjects with CKD from the MDRD study group and was re-expressed for use with IDMS traceable creatinine assay in 2006.[45] The equation is shown as below.

$$\text{eGFR} = 175 \times S_{cr} - 1.154 \times (\text{Age}) - 0.203 \times (0.742 \times I_{\text{female}}) \times (1.212 \times I_{\text{African American}})$$

$$I_{\text{female}} = 1 \text{ if female}, I_{\text{African American}} = 1 \text{ if African American}$$

Demographics, including age, gender were extracted from the electronic health records. The major diagnosis of disease present on admission, and the index for diabetes showed were also obtained from the records.

2.2 Correlates of Vitamin D Analyses

To evaluate the correlates of vitamin D deficiency, we first classified the patients into two groups by 25(OH)D level: the normal vitamin D group with 25(OH)D level >=30 ng/mL and vitamin D deficiency group with 25(OH)D level <30 ng/mL. As for the CKD status, we categorized patients into two groups: No CKD group (eGFR >=60 mL/min/1.73 m^2) and CKD group (eGFR <60 mL/min/1.73 m^2). We compared baseline characteristics (gender, age, diabetes, eGFR and CKD status) between patients with normal vitamin D and those with vitamin D deficiency using Pearson Chi-square Test and analysis of variance (ANOVA) test for categorical and continuous

variables, respectively. The correlation analyses were conducted based on observed data; the missing data were not included into analyses.

2.3 Survival Analyses

First, we evaluated the marginal association of mortality with each potential risk factor. For the categorical covariates, including vitamin D level, gender, CKD status and presence of diabetes, we used the Kaplan-Meier estimator to estimate the survival function curves for different factor levels,[46] and then we used log-rank tests to assess whether mortality risk in the different factor levels significantly differs. The Kaplan-Meier estimator is written:

$$S_{KM}(t) = \prod_{i=0}^{t}(1 - \frac{d_i}{R(t_i)})$$

$$d_i = \text{number of events at time } t_i;\ R(t_i) = \text{number of events risk at time } t_i$$

For the continuous covariates, including age and eGFR, we used the Wald test from a univariate Cox regression model to evaluate the association between mortality and each covariate respectively.[47] A Cox regression model (Cox proportional hazards model) assumes an underlying baseline hazard function $\lambda_0(t)$ and holds an assumption of a constant shift of the baseline hazard by non-zero covariate values (i.e., proportional assumption). A Cox regression model takes the form,

$$\lambda(t|x) = \lambda_0(t)\exp(\beta'X), \qquad X = \text{covariates}$$

Then a multivariate Cox regression model was used to assess the association between vitamin D deficiency and mortality. Covariates which have shown significant associations with mortality in the marginal test were added to the multivariate model. In addition, 2-way interaction items between vitamin D deficiency and another candidate covariate were also considered when we built the multivariate Cox regression model.

Upon the detection of a significant interaction term, we further performed stratified Cox regression models with the strata defined by the vitamin D level. We also performed formal tests to compare the coefficients between strata. A significant difference would indicate the corresponding covariate plays a different role in influencing mortality risk for subjects with vitamin D deficiency versus those with normal vitamin D level.

Estimated hazard ratios with 95% confidence interval (95% CI) for the vitamin D deficiency and other covariates included in the final model were calculated and presented in the result section.

2.3 Model Diagnosis

We first tested the proportional hazards assumption for each covariate. This assumption means the hazard rates for two observations are proportional to one another and that proportionality is constant over time. We used a non-proportionality test based on the scaled Schoenfeld residuals to test the assumption.[48] At the jth event time t_{i_j} of the ith subject, the Schoenfeld residual is the difference between the ith subject covariate vector at t_{i_j} ($Z_i(t_{i_j})$) and the average of the covariates vectors over risk set at time t_{i_j}

$$\hat{U}_i\left(t_{i_j}\right) = Z_i\left(t_{i_j}\right) - \bar{Z}(\hat{\beta}, t_{i_j})$$

An alternative to proportional hazards model is a time-varying coefficient model

$$\lambda(t, Z) = \lambda_0(t)\exp(\beta(t)Z)$$

Testing $\beta(t) = \beta$, which indicates time-constant coefficients, is equivalent to testing for a non-zero slope in a generalized linear regression of the Schoenfeld residuals on functions of time. Grambsch and Therneau gave a scaled Schoenfeld residual definition as the product of the inverse of the covariance matrix of residuals times the vector of residuals. The test statistics for covariate specific test is as follows.

$$T_k = \frac{\{\sum(g_k - \bar{g})r_k^*\}^2}{dI_k \sum(g_k - \bar{g})^2}$$

where r_k is the scaled Schoenfeld residual for covariate k, g_k is the time scale (in this study we used log transformation of time) and $\bar{g}$ as the average time scale, d are the survival time, I_k is the information matrix elements for covariate k. This test statistic follows a χ^2 distribution with 1 degree of freedom for the covariate-specific version of the test. Test statistics that exceed 5% critical values are viewed as evidence to violate the proportional hazards assumption. The Schoenfeld residuals plot for each of the covariate in the model, including a Lowess smoothing curve versus time, were also generated to examine the proportional hazard assumption by looking at the slope. A plot of residuals versus time would be centered about zero if proportional hazards assumption holds.

We also used Score residuals method to examine the influential observations on the estimation of coefficients. $\hat{\beta}$ is the estimator of coefficient vector from fitting the Cox model based on all data; $\widehat{\beta_J}$ is the estimator of coefficient vector from the data with the jth observation removed. The approximation of difference of $\hat{\beta} - \widehat{\beta_J}$ based on the Score residuals is

$$\hat{\beta} - \widehat{\beta_J} \approx \Delta_j = I(\hat{\beta})^{-1}(S_{j1}, \dots, S_{jp})'$$

where $I(\hat{\beta})^{-1}$ is the covariance matrix of $\hat{\beta}$; S_{jk} is the Score residuals and defined as

$$\delta_j\{Z_{jk} - \overline{Z_K}(T_J)\} - \sum_{t_h \le T_j} \{Z_{jk} - \overline{Z_K}(t_h)\} \times exp(\hat{\beta}' Z_j)\{\widehat{H_o}(t_h) - \widehat{H_o}(t_{h-1})\}$$

for j=1,..,n and k=1,…,p. The first term $\delta_j\{Z_{jk} - \overline{Z_K}(T_J)\}$ is Schoenfeld's partial residual. Therefore, we used plots of approximate difference in estimated coefficients scaled by standard error for the coefficients against the subject number to gage the influence of the jth observation on each covariate respectively.

To assess the overall fit of the Cox regression models, Cox-Snell residuals plots method was used. [49] The Cox-Snell residual is defined as

$$r_j = \widehat{H_0}(T_j)\exp(\hat{\beta}' Z_j)$$

where $\widehat{H_0}$ is Breslow's estimator of the baseline hazard rate; Z_j is the covariate vector for subject j; $\hat{\beta}$ is the coefficient vector. If the model is correct and the $\hat{\beta}$'s are close to the true values, then the r_j's should look like a censored sample from a unit exponential distribution. To check whether the r_j's behave as a sample from a unit exponential, we could compute the Nelson–Aalen estimator of the cumulative hazard rate of the r_j 's. If the unit exponential distribution fits the data, this estimator then should be approximately equal to the cumulative hazard rate of the unit exponential ($H_E(t) = t$). Thus, a plot of the estimated cumulative hazard rate of the r_j, which is $\widehat{H_r}(r_j)$, versus r_j should be a 45º straight line through the origin if the model is good of fit.

The nominal significant level for all the analyses was 0.05. Data analyses of Cox regression models were conducted using SAS version 9.4 (SAS Institute, www.sas.com); the analyses including Pearson Chi-square Test, ANOVA test, Kaplan-Meier plot, log-rank test, and model diagnosis were conducted using R 3.0.2 (The R Foundation for Statistical Computing, www.r-project.org). All the graphs were generated using R3.0.2. R package "survival" (Therneau & Lumley, www.r-project.org) was used for Kaplan-Meier plot, log-rank test and model diagnosis.

III Results

(All of the tables/figures are in Appendix A)

3.1 Participants Characteristics

At baseline the study population (Table 1) has an average age of 64.77 ± 20.55 years (3 patients with missing values of age), 34% of the patients are male, 37% have diabetes records. Patients' distribution by group of CKD status are 55% without CKD and 44% with CKD. The median value of eGFR is 68.77 mL/min/1.73 m^2 (interquartile range: 30.64 mL/min/1.73 m^2). 866(75%) of patients with 25(OH)D <30 ng/mL were categorized into vitamin D deficiency group. Among patients who do not have CKD, 70% are with vitamin D deficiency, while among those who have CKD, 81% are with vitamin D deficiency. By examining the factors correlated with vitamin D

deficiency level, there is no significant difference in age or gender or presence of diabetes among patients with vitamin D deficiency and normal vitamin D level (25(OH)D >=30 ng/mL). Conversely, eGFR is significantly lower among vitamin D deficiency group compared to the normal vitamin D group ($P < 0.001$).

3.2 Marginal Association between Mortality and Potential Risk Factor

Of the 1159 subjects in the study, 380 died during an average follow-up of 427.3 days (median 329 days, range 2 - 1486 days). In crude analysis without adjustments, there is no significant difference in mortality risk for patients with low vitamin D level and patients with normal vitamin D level (P=0.677). By examining the marginal association of mortality with each covariate separately, neither gender nor presence of diabetes is significantly associated with mortality risk (P=0.07; P=0.377 respectively). Age is positively associated with mortality risk (P<0.001). A higher eGFR is significantly associated with a lower risk of mortality (P<0.001). Figure 1 summarizes the marginal associations between mortality and each risk factor.

3.3 Multivariate Survival Analyses

Based on the test results of the marginal association of mortality with risk factors, we adjust age, eGFR for the multivariate Cox regression model of all-cause mortality with vitamin D level. Considering the eGFR is significantly lower in patients with vitamin D deficiency compared to that in patients with normal vitamin D, the 2-way interaction between vitamin D deficiency and eGFR is also added into the Cox regression model. A fitted model was shown as below:

$$h(t) = h_0(t)\exp\left(\beta_1 I_{\text{vitamin D deficiency}} + \beta_2 \text{eGFR} + \beta_3 \text{Age} + \beta_4 I_{\text{vitamin D deficiency}} \times \text{eGFR}\right) \quad (1)$$

The hazard ratio (HR) for mortality among patients with vitamin D deficiency compared to patients with normal vitamin D level is 1.049 (95% CI, 0.822 – 1.337; Table 2), indicating a non-significant impact of vitamin D deficiency on all-cause mortality. However, the local Likelihood Ratio test shows the interaction between vitamin D deficiency and eGFR is significant in the

model (1) (P=0.0265, Table 2). This indicates that the impact of eGFR on all-cause mortality among patients with normal vitamin D might be different from those with vitamin D deficiency.

To examine this hypothesis, we then stratified the study subjects by vitamin D levels (normal and deficiency) and conducted survival analyses by strata, adjusting for age, eGFR. The fitted stratified models are as below, where $h_1(t)$ is the model for patients with normal vitamin D and $h_2(t)$ is the model for patients with normal vitamin D.

$$h_1(t) = h_{01}(t)\exp(\beta_{11}\text{eGFR} + \beta_{21}\text{Age}) \qquad (2)$$

$$h_2(t) = h_{02}(t)\exp(\beta_{12}\text{eGFR} + \beta_{22}\text{Age}) \qquad (3)$$

Table 3 gives a summary of the estimates for model (2) and (3). After adjustment, for the patients with normal vitamin D, the HR for all-cause mortality with 1 mL/min/1.73 m^2 increase in eGFR is 1.004 (95% CI, 0.988-1.009; Table 3). For the patients with vitamin D deficiency, the HR for all-cause mortality with 1 mL/min/1.73 m^2 increase in eGFR is 0.997 (95% CI, 0.994-1.000; Table 3). The coefficients of eGFR in model (2) and model (3) are significantly different (P=0.014). Therefore the negative impact of a higher eGFR on mortality risk is indicated to be evident among patients with vitamin D deficiency, but is no longer significant for patients with normal vitamin D. An older age is significantly associated with a higher risk of mortality in model (1), model (2) and model (3) (Table 1-3).

We then add $eGFR^2$ into the stratified models to see if there is a non-linear association between eGFR and mortality.

$$h_3(t) = h_{03}(t)\exp(\beta_{13}\text{eGFR} + \beta_{23}\text{Age} + \beta_{33}\text{eGFR}^2) \qquad (4)$$

$$h_4(t) = h_{04}(t)\exp(\beta_{14}\text{eGFR} + \beta_{24}\text{Age} + \beta_{34}\text{eGFR}^2) \qquad (5)$$

Model (4) and model (5) are the models with $eGFR^2$ for normal vitamin D group and vitamin D deficiency group respectively. In the summary estimates of these two models (Table 4), the local Likelihood Ratio Test results show that neither eGFR nor $eGFR^2$ is significant in model (4) and

the HR for all-cause mortality with one unit increase in eGFR at 60 mL/min/1.73 m^2 is 1.000 (95% CI, 0.993-1.007). Inversely, both the eGFR and $eGFR^2$ are found to be significant in model (5), with a HR of 0.994 (95% CI, 0.991-0.997) on mortality with one unit increase in eGFR at 60 mL/min/1.73 m^2. This still indicates eGFR has a stronger impact on the mortality among patients with vitamin D deficiency.

3.4 Model Diagnosis

The results from testing the proportional hazards assumption for model (1) suggest that the proportional hazards assumption may be adequate for each covariate in the model (P>0.5 for all covariates). The plots based on Schoenfeld residuals for each covariate are close to zero and thus give a result consistent with the hypothesis test results (Figure2-4).

Looking at the Score residuals method plots for each covariate in model (1) (Figure 5), we find subject ID 895 and 997 have obvious impact on the estimation of the coefficients for vitamin D deficiency and eGFR; subject ID 74 have large influence on estimated coefficient for eGFR; several more subjects have higher impact on the coefficient estimates of age (ID 170, 513 and 718), but with a relative smaller scale. For example, subject ID 897 who was 25.74 years old and had a normal vitamin D level, with a 294.3016 mL/min/1.73 m^2 eGFR, but only had a survival time of 26 days, while based on the full sample analysis, he/she should live much longer.

The Cox-Snell residuals plot for model (1) suggests that the model does not fit too badly, except for a tail where the variability in the estimate of the cumulative hazard rate is large (Figure 6). For model (2)-(5), all the Cox-Snell residual plots show a larger variability of the estimate of the cumulative hazard rate in the tails, but still suggest a not bad fitting result. Without considering the tail part, the estimates of model (3) and model (5) are closer to the 45^o line than that of model (2) and model (4), indicating the models for vitamin D deficiency group fit better than the models for normal vitamin D. Comparing the Cox-Snell residuals plots for model (1)-(5), we can see the estimates for stratified models (model (2)-(5)) lie more above the 45^o line. This could support the

opinion that stratified models fit better than the unstratified model (model (1)), which means the impact of CKD status might do have a different impact among patients with different vitamin D levels.

IV Discussion

This study shows eGFR have a stronger and significant negative impact on the risk of all-cause mortality among patients with vitamin D deficiency (25(OH)D<30 ng/mL), compared to an observed insignificant impact on mortality risk among patients with normal vitamin D status. One unit increase in eGFR at a baseline of 60 mL/min/1.73 m^2 would have a 0.994 lower risk for all-cause mortality. This finding is consistent with the results from previous study that kidney disease and vitamin D deficiency have an interacted effect on all-cause mortality. However, no significant association between vitamin D deficiency and all-cause mortality is found among our whole study population. The similar conclusion is obtained from the two subgroups, patients with CKD and patients without CKD. We also try different cutoffs of 25(OH)D level to define vitamin D deficiency. As an alternative, 25(OH)D < 15 ng/mL is considered as a more strict definition of vitamin D deficiency. The results based on the cutoff of 15 ng/mL as vitamin D deficiency are consistent with our findings based on the cutoff of 30 ng/mL.

One major limitation of our study is that we may not have sufficient potential confounders in the dataset to be well sampled and controlled, so as to allow for an independent assessment of the impact of vitamin D deficiency on all-cause mortality. From the findings on previous studies, some factors like the season and race, are the risk factors associated with vitamin D level in the body. Therefore, ignoring the seasonal variation of 25(OH)D test results and the race information of our study subjects could lead to attenuation of the effect of vitamin D on mortality. Besides, based on the model diagnosis conducted in our previous analysis, several individual observations who showed a heavy influence on the estimations also need a further investigation. The

association between vitamin D deficiency and cause-specific mortality is another analysis which can be done in the future.

In our study, we follow a traditional purposeful selection method to build the multivariate model, i.e., fitting a multivariable model containing all variables that are significant in a univariate analysis and then checking nonlinearity and interaction. This method can be easily conducted since we have only a few covariates here. If there will be more covariates included in the future, a more appropriate and accurate variable selection method should be used. For example, the Lasso method can be used to reduce the model dimension to provide parsimonious interpretation while ensuring good prediction accuracy.[50]

References

1. Holick, M.F. and T.C. Chen, *Vitamin D deficiency: a worldwide problem with health consequences.* Am J Clin Nutr, 2008. **87**(4): p. 1080S-6S.
2. Heaney, R.P., *Vitamin D in health and disease.* Clin J Am Soc Nephrol, 2008. **3**(5): p. 1535-41.
3. van Schoor, N.M. and P. Lips, *Worldwide vitamin D status.* Best practice & research Clinical endocrinology & metabolism, 2011. **25**(4): p. 671-680.
4. Clemens, T.L., et al., *Increased skin pigment reduces the capacity of skin to synthesise vitamin D3.* Lancet, 1982. **1**(8263): p. 74-6.
5. Matsuoka, L.Y., et al., *Sunscreens suppress cutaneous vitamin D3 synthesis.* J Clin Endocrinol Metab, 1987. **64**(6): p. 1165-8.
6. Holick, M., L. Matsuoka, and J. Wortsman, *Age, vitamin D, and solar ultraviolet.* The Lancet, 1989. **334**(8671): p. 1104-1105.
7. Wortsman, J., et al., *Decreased bioavailability of vitamin D in obesity.* Am J Clin Nutr, 2000. **72**(3): p. 690-3.
8. Holick, M.F., *Vitamin D deficiency.* N Engl J Med, 2007. **357**(3): p. 266-81.
9. Dawson-Hughes, B., et al., *Estimates of optimal vitamin D status.* Osteoporos Int, 2005. **16**(7): p. 713-6.
10. Pearce, S.H. and T.D. Cheetham, *Diagnosis and management of vitamin D deficiency.* BMJ, 2010. **340**: p. b5664.
11. Ford, L., et al., *Vitamin D concentrations in an UK inner-city multicultural outpatient population.* Ann Clin Biochem, 2006. **43**(Pt 6): p. 468-73.
12. Forrest, K.Y. and W.L. Stuhldreher, *Prevalence and correlates of vitamin D deficiency in US adults.* Nutr Res, 2011. **31**(1): p. 48-54.
13. Ginde, A.A., M.C. Liu, and C.A. Camargo, Jr., *Demographic differences and trends of vitamin D insufficiency in the US population, 1988-2004.* Arch Intern Med, 2009. **169**(6): p. 626-32.
14. Scragg, R. and C.A. Camargo, Jr., *Frequency of leisure-time physical activity and serum 25-hydroxyvitamin D levels in the US population: results from the Third National Health and Nutrition Examination Survey.* Am J Epidemiol, 2008. **168**(6): p. 577-86; discussion 587-91.
15. Ogden, C.L., et al., *Prevalence of overweight and obesity in the United States, 1999-2004.* JAMA, 2006. **295**(13): p. 1549-55.

16. Bischoff-Ferrari, H.A., et al., *Fracture prevention with vitamin D supplementation: a meta-analysis of randomized controlled trials.* JAMA, 2005. **293**(18): p. 2257-64.
17. Murad, M.H., et al., *Clinical review: The effect of vitamin D on falls: a systematic review and meta-analysis.* J Clin Endocrinol Metab, 2011. **96**(10): p. 2997-3006.
18. Witham, M.D., M.A. Nadir, and A.D. Struthers, *Effect of vitamin D on blood pressure: a systematic review and meta-analysis.* J Hypertens, 2009. **27**(10): p. 1948-54.
19. Grandi, N.C., L.P. Breitling, and H. Brenner, *Vitamin D and cardiovascular disease: systematic review and meta-analysis of prospective studies.* Prev Med, 2010. **51**(3-4): p. 228-33.
20. Zhao, X.Y. and D. Feldman, *The role of vitamin D in prostate cancer.* Steroids, 2001. **66**(3-5): p. 293-300.
21. Grant, W.B., *How strong is the evidence that solar ultraviolet B and vitamin D reduce the risk of cancer? An examination using Hill's criteria for causality: An examination using Hill's criteria for causality.* Dermato-endocrinology, 2009. **1**(1): p. 17-24.
22. Kilkkinen, A., et al., *Vitamin D status and the risk of lung cancer: a cohort study in Finland.* Cancer Epidemiol Biomarkers Prev, 2008. **17**(11): p. 3274-8.
23. Adorini, L. and G. Penna, *Control of autoimmune diseases by the vitamin D endocrine system.* Nat Clin Pract Rheumatol, 2008. **4**(8): p. 404-12.
24. Hutchinson, M.S., et al., *Low serum 25-hydroxyvitamin D levels are associated with increased all-cause mortality risk in a general population: the Tromso study.* Eur J Endocrinol, 2010. **162**(5): p. 935-42.
25. Schottker, B., et al., *Strong associations of 25-hydroxyvitamin D concentrations with all-cause, cardiovascular, cancer, and respiratory disease mortality in a large cohort study.* Am J Clin Nutr, 2013. **97**(4): p. 782-93.
26. Skaaby, T., et al., *Vitamin D status and cause-specific mortality: a general population study.* PLoS One, 2012. **7**(12): p. e52423.
27. Thomas, M.K., et al., *Hypovitaminosis D in medical inpatients.* N Engl J Med, 1998. **338**(12): p. 777-83.
28. Mehrotra, R., et al., *Hypovitaminosis D in chronic kidney disease.* Clin J Am Soc Nephrol, 2008. **3**(4): p. 1144-51.
29. Doorenbos, C.R., et al., *Possible renoprotection by vitamin D in chronic renal disease: beyond mineral metabolism.* Nat Rev Nephrol, 2009. **5**(12): p. 691-700.
30. Hilpert, J., et al., *Expression profiling confirms the role of endocytic receptor megalin in renal vitamin D3 metabolism.* Kidney Int, 2002. **62**(5): p. 1672-81.

31. Willnow, T.E. and A. Nykjaer, *Pathways for kidney-specific uptake of the steroid hormone 25-hydroxyvitamin D3.* Curr Opin Lipidol, 2002. **13**(3): p. 255-60.
32. Zhang, Y., et al., *Vitamin D receptor attenuates renal fibrosis by suppressing the renin-angiotensin system.* Journal of the American Society of Nephrology, 2010. **21**(6): p. 966-973.
33. National Kidney, F., *K/DOQI clinical practice guidelines for chronic kidney disease: evaluation, classification, and stratification.* Am J Kidney Dis, 2002. **39**(2 Suppl 1): p. S1-266.
34. Autier, P. and S. Gandini, *Vitamin D supplementation and total mortality: a meta-analysis of randomized controlled trials.* Arch Intern Med, 2007. **167**(16): p. 1730-7.
35. Wang, A.Y., et al., *Serum 25-hydroxyvitamin D status and cardiovascular outcomes in chronic peritoneal dialysis patients: a 3-y prospective cohort study.* Am J Clin Nutr, 2008. **87**(6): p. 1631-8.
36. Barreto, D.V., et al., *Vitamin D affects survival independently of vascular calcification in chronic kidney disease.* Clin J Am Soc Nephrol, 2009. **4**(6): p. 1128-35.
37. Pecovnik-Balon, B., et al., *Vitamin D as a novel nontraditional risk factor for mortality in hemodialysis patients.* Ther Apher Dial, 2009. **13**(4): p. 268-72.
38. Gracia-Iguacel, C., et al., *Vitamin D deficiency in dialysis patients: effect of dialysis modality and implications on outcome.* J Ren Nutr, 2010. **20**(6): p. 359-67.
39. Drechsler, C., et al., *Vitamin D deficiency is associated with sudden cardiac death, combined cardiovascular events, and mortality in haemodialysis patients.* Eur Heart J, 2010. **31**(18): p. 2253-61.
40. Wolf, M., et al., *Vitamin D levels and early mortality among incident hemodialysis patients.* Kidney Int, 2007. **72**(8): p. 1004-13.
41. Ravani, P., et al., *Vitamin D levels and patient outcome in chronic kidney disease.* Kidney Int, 2009. **75**(1): p. 88-95.
42. Drechsler, C., et al., *Vitamin D status and clinical outcomes in incident dialysis patients: results from the NECOSAD study.* Nephrol Dial Transplant, 2011. **26**(3): p. 1024-32.
43. Jassal, S.K., et al., *Vitamin d, parathyroid hormone, and cardiovascular mortality in older adults: the Rancho Bernardo study.* Am J Med, 2010. **123**(12): p. 1114-20.
44. Pilz, S., et al., *Vitamin D status and mortality risk in CKD: a meta-analysis of prospective studies.* Am J Kidney Dis, 2011. **58**(3): p. 374-82.

45. Levey, A.S., et al., *Using standardized serum creatinine values in the modification of diet in renal disease study equation for estimating glomerular filtration rate.* Annals of internal medicine, 2006. **145**(4): p. 247-254.
46. Kaplan, E.L. and P. Meier, *Nonparametric estimation from incomplete observations.* Journal of the American statistical association, 1958. **53**(282): p. 457-481.
47. Cox, D.R., *Regression models and life-tables*, in *Breakthroughs in Statistics*. 1992, Springer. p. 527-541.
48. Grambsch, P.M. and T.M. Therneau, *Proportional hazards tests and diagnostics based on weighted residuals.* Biometrika, 1994. **81**(3): p. 515-526.
49. Cox, D.R. and E.J. Snell, *A general definition of residuals.* Journal of the Royal Statistical Society. Series B (Methodological), 1968: p. 248-275.
50. Tibshirani, R., *The lasso method for variable selection in the Cox model.* Stat Med, 1997. **16**(4): p. 385-95.

Appendix A

Table1. Characteristics of study participants at baseline by vitamin D status

	All N=1159	Low Vitamin D (25D <30 ng/mL) N=866	Normal Vitamin D (25D >=30 ng/mL) N=293	P-Value
Age[a] (N=1156)	64.77 ± 20.55	64.60 ± 19.36	65.52 ± 23.65	0.505[d]
Gender[a]				
Male	398 (34.34)	288 (33.26)	110 (37.54)	0.1817[c]
Female	761 (65.66)	578 (66.74)	183 (62.46)	
eGFR[b] (N=1155)	66.20 (59.58)	62.80 (64.77)	76.50 (47.47)	<0.001[d]
CKD stage group[a]				
CKD (eGFR>=60)	643 (55.48)	451 (52.14)	192 (66.21)	<0.001[c]
No CKD (eGFR<60)	512 (44.18)	414 (47.86)	98 (33.79)	
Missing	4 (0.34)			
Diabetes[a]				
Yes	431 (37.12)	339 (67.94)	92 (73.02)	0.3204[c]
No	195 (16.80)	160 (32.06)	34 (26.98)	
Missing	535 (46.08)			

25D, 25-hydroxyvitaminD; CKD, chronic kidney disease;
eGFR, estimated glomerular filtration rate

[a] Categorical variable statistics showed as N(%);Continuous variable statistics showed as Mean ± SD

[b] Continuous variable highly skewed; statistics showed as Median(interquartile range)

[c] P-value for Pearson Chi-Square Test (two-sided)

[d] P-value from ANOVA test

[e] P-value from ANOVA test based on transformation log(eGFR)

Table2. Summary of Estimates for Model (1)

Variable	Coefficient (β)	Standard Error	P-value	Hazard Ratio	95% CI of HR	
Vitamin D deficiency	0.5376	0.2528	0.0335	1.049[a]	0.822	1.337
eGFR	0.0037	0.0025	0.1435	1.004[b]	0.999	1.009
				0.997[c]	0.994	1.000
Age	0.0245	0.0031	<0.0001	1.025[d]	1.019	1.031
Vitamin D deficiency* eGFR	-0.0068	0.0029	0.0208	-	-	-

[a] Vitamin D deficiency vs. normal vitamin D at eGFR=72.5675
[b] 1 unit increase in eGFR for normal vitamin D level
[c] 1 unit increase in eGFR for vitamin D deficiency level
[d] 1 unit increase in age

Table3. Summary of Estimates for Model (2) vs. Model (3)

Model	Variable	Coefficient (β)	Standard Error	P-value	Hazard Ratio	95% CI	
(2)[a]	eGFR	0.0039	0.0028	0.1622	1.004[c]	0.998	1.009
	Age	0.0247	0.0060	<.0001	1.025[d]	1.013	1.037
(3)[b]	eGFR	-0.0031	0.0015	0.0437	0.997[c]	0.994	1.000
	Age	0.0242	0.0036	<.0001	1.025[d]	1.017	1.032

[a] model for patients with normal vitamin D, adjust for eGFR and Age
[b] model for patients with vitamin D deficiency, adjust for eGFR and Age
[c] 1 unit increase in eGFR
[d] 1 unit increase in age

Table4. Summary of Estimates for Model (4) vs. Model (5)

Model	Variable	Coefficient (β)	Standard Error	P-value	Hazard Ratio	95% CI	
(4)[a]	eGFR	-0.0032	0.0053	0.5522	1.000[c]	0.993	1.007
	Age	0.0247	0.0061	<0.0001	1.025[d]	1.013	1.037
	$eGFR^2$	0.0000254	0.0000156	0.1019			
(5)[b]	eGFR	-0.0122	0.0029	<0.0001	0.994[c]	0.991	0.997
	Age	0.0267	0.0037	<0.0001	1.025[d]	1.017	1.032
	$eGFR^2$	0.0000494	0.0001684	<0.0001			

[a] model for patients with normal vitamin D, adjust for eGFR, Age and $eGFR^2$
[b] model for patients with vitamin D deficiency, adjust for eGFR, Age and $eGFR^2$
[c] 1 unit increase in eGFR at 60
[d] 1 unit increase in age

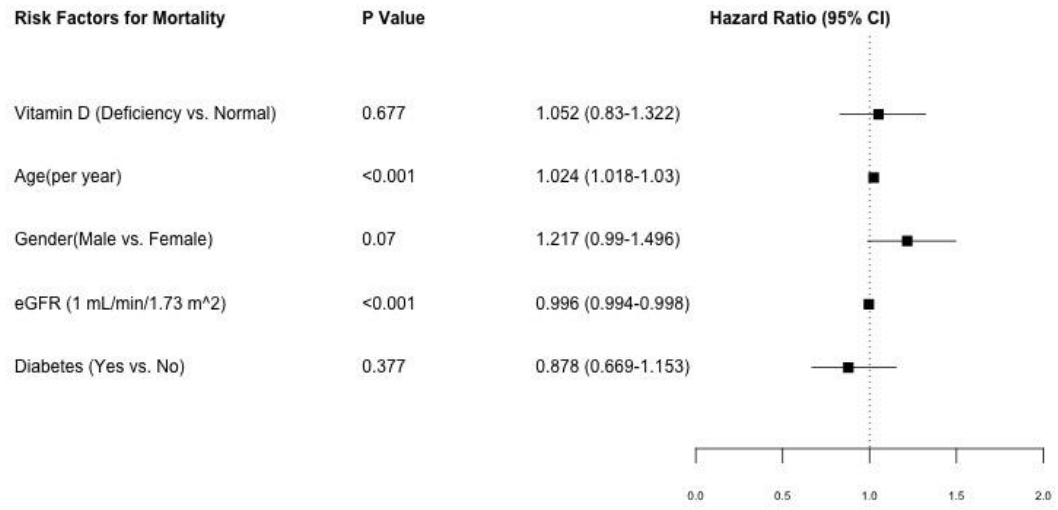

Figure 1. Hazard ratios, 95% confidence intervals (95% CI) and P values for all-cause mortality marginally associated with each risk factor

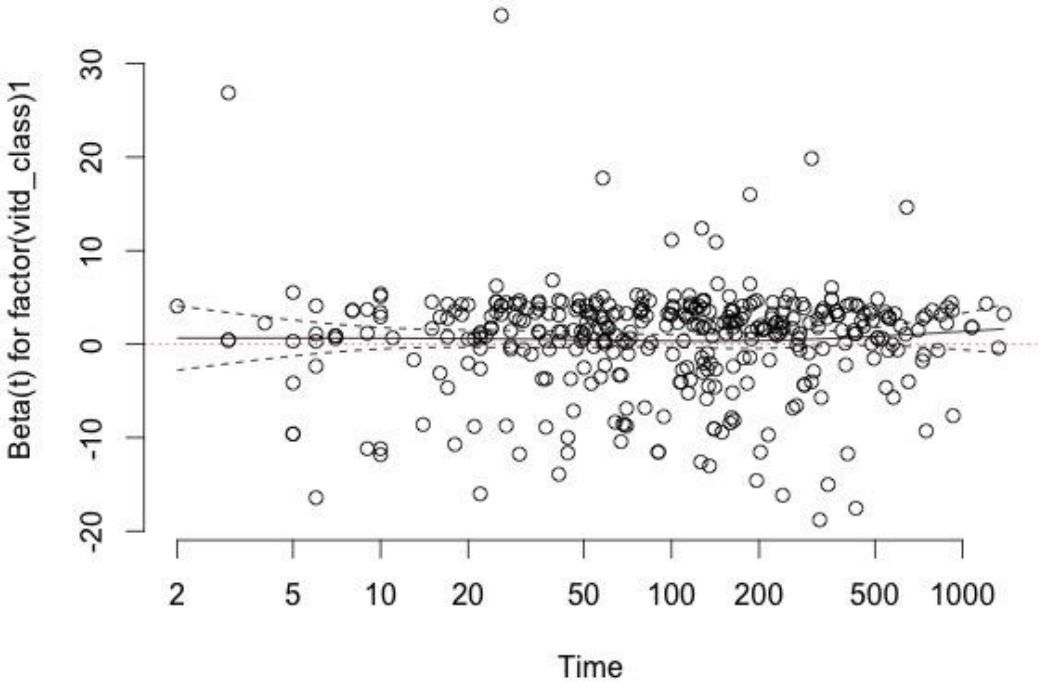

Figure 2. Time-dependent coefficient estimates for vitamin D group based on scaled Schoenfeld-residuals in model (1)

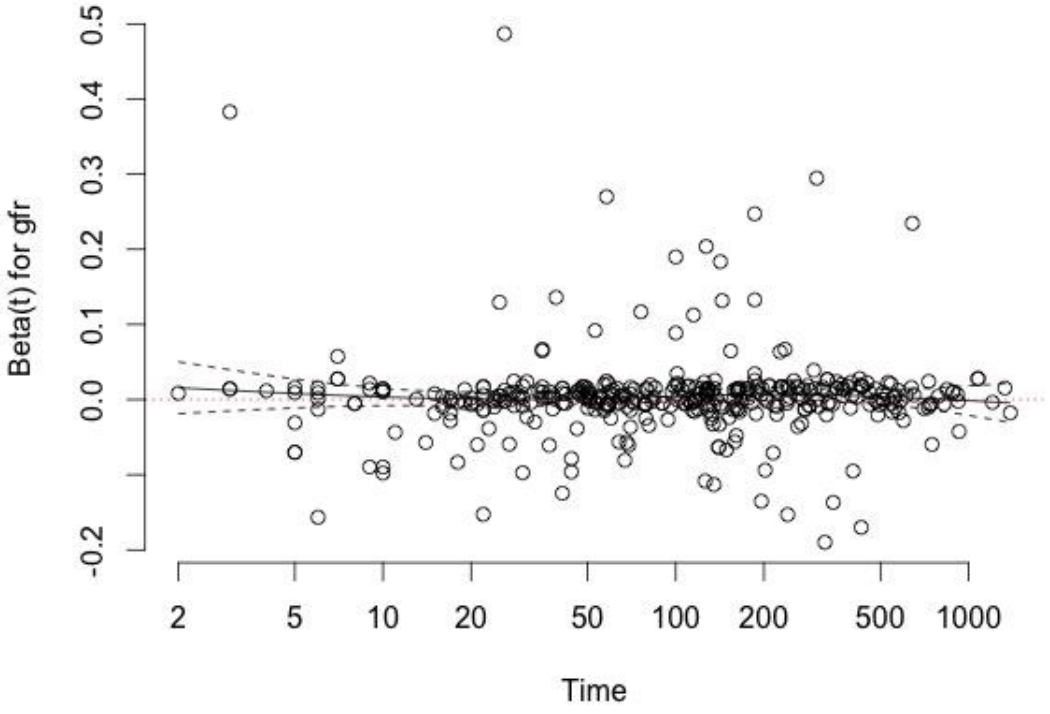

Figure 3. Time-dependent coefficient estimates for eGFR based on scaled Schoenfeld-residuals in model (1)

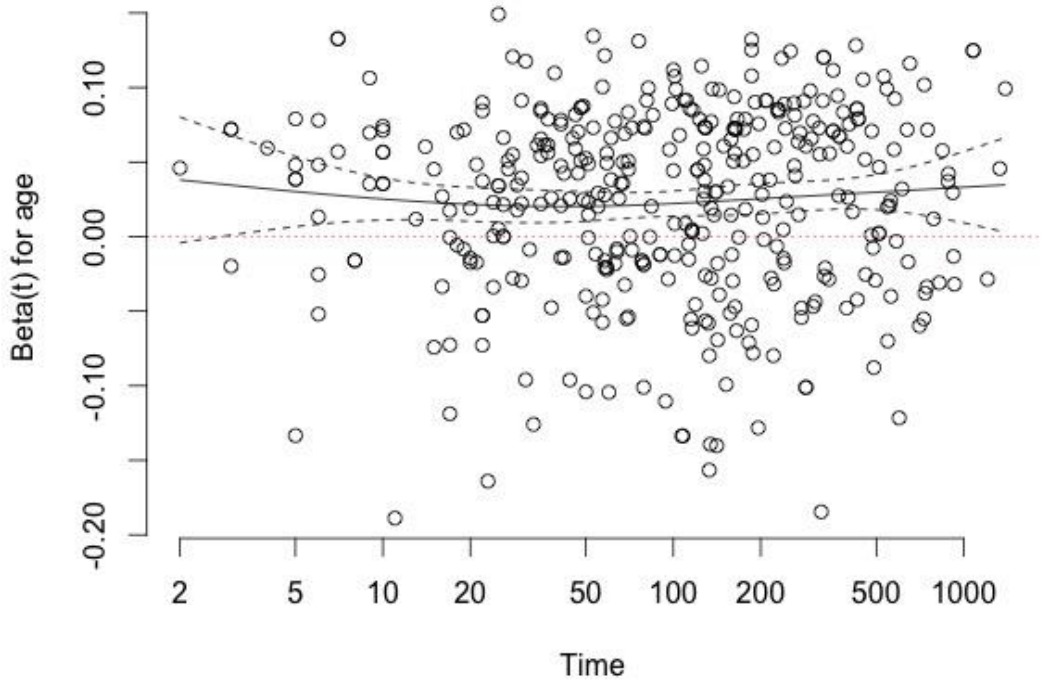

Figure 4. Time-dependent coefficient estimates for age based on scaled Schoenfeld-residuals in model (1)

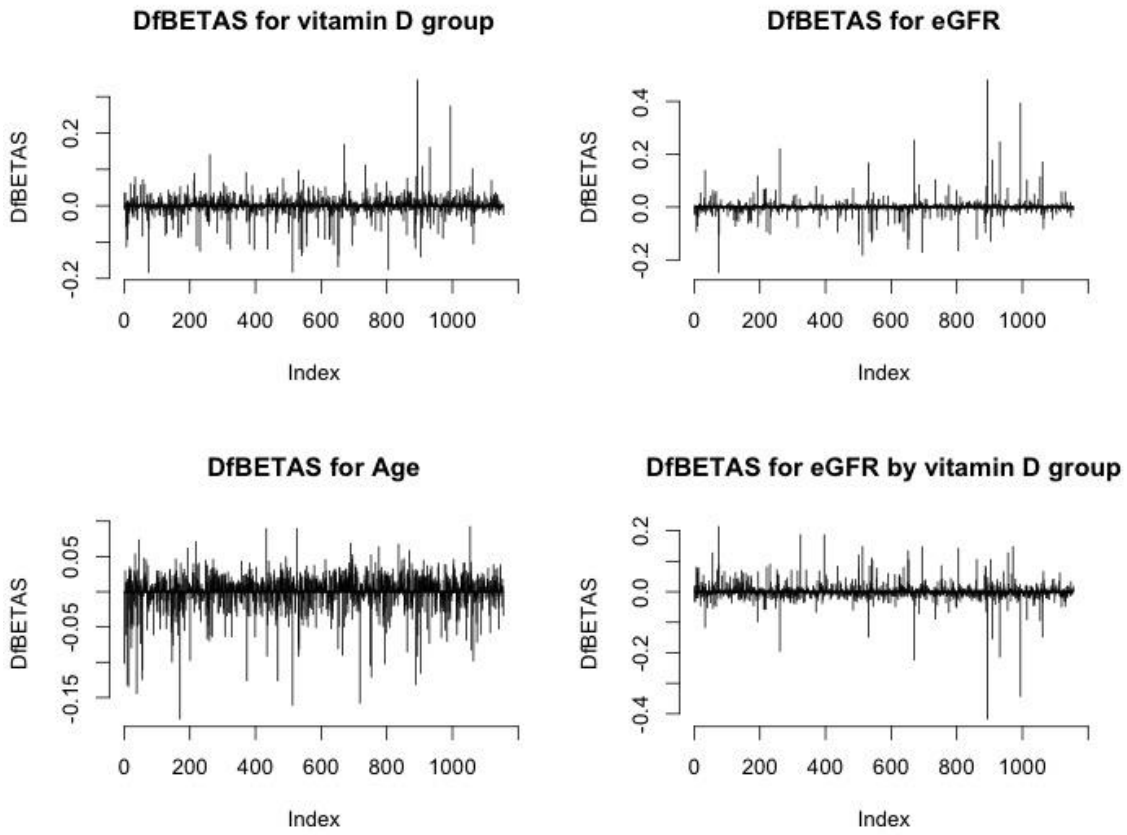

Figure 5. Influence observations for vitamin D group, eGFR, age and interaction of vitamin D group with eGFR in model (1)

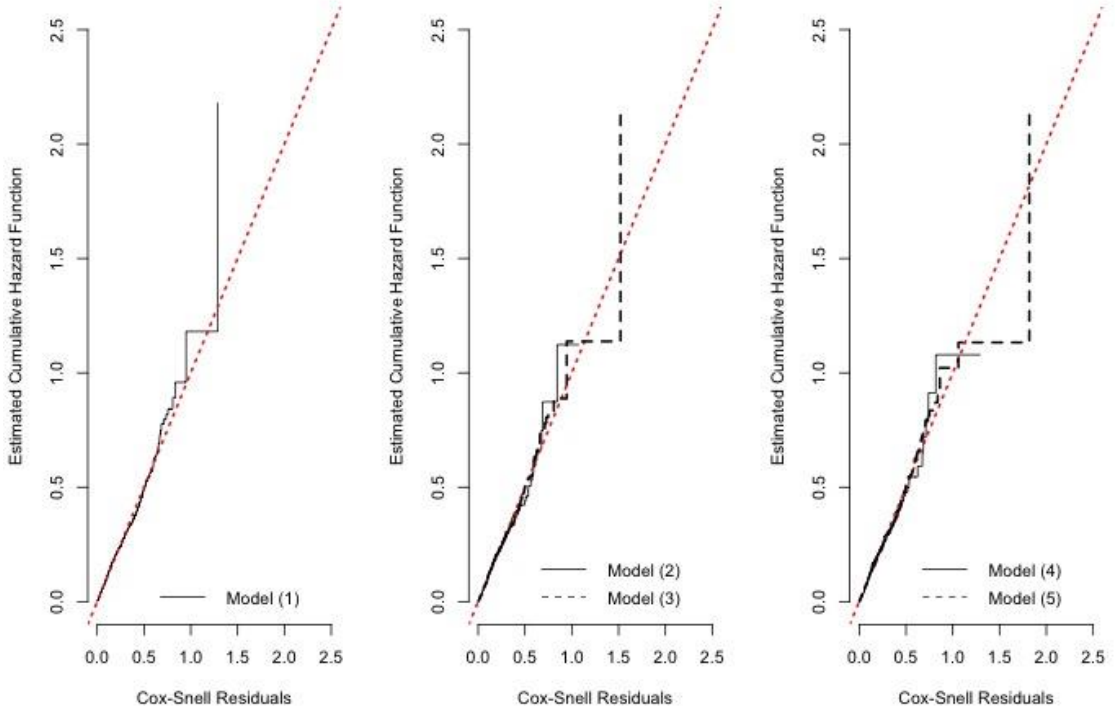

Figure 6. Cox-Snell residual plots for model (1)-(5)