Modeling In-patients with Chronic Non-Communicable Diseases Using Parametric Shared Frailty Models

Belachew Melese Hunde (belachewmelese3@gmail.com)

Research article

Keywords: Log-logistic, Inverse Gaussian, Parametric Frailty, Non-communicable Diseases

Posted Date: October 31st, 2019

DOI: https://doi.org/10.21203/rs.2.16662/v1

Abstract

Background Non-communicable diseases, known as chronic diseases, are not contagious in their nature. They progress slowly in affecting the health of a person. They are the leading causes of death in all continents except Africa, but current projections indicated that by 2025 the largest increases in the non-communicable diseases deaths will occur in Africa. In Ethiopian about 34% of patients suffered from chronic non communicable diseases and there is a gap of estimating the time-to-death of these patients to manage the diseases progression in earlier. Thus, this study was aimed in estimating the survival outcome (death times) of the retrospective follow-up studies of registered inpatients for three years in three hospitals of Oromiya National Regional State, Ethiopia.

Methods To describe the prediction and diseases progression of non-communicable diseases, different types of parametric frailty models were compared. Hospitals of the patients were considered as the unobserved variable in the models. The Exponential, Weibull and log-logistic as baseline hazard functions and the gamma and inverse Gaussian for the frailty distributions were checked for their performance using both AIC criteria and Likelihood ratio test.

Results On average death times of chronic diseases was 12 days with the maximum of 40 days. Of 646 chronic non-communicable hospitalized patients about 41.5% were died. The log-logistic model with inverse Gaussian frailty has the minimum AIC and LRT value among the models compared. The hospital of the patient has a significant effect in modeling time-to-death of chronic diseases datasets.

Conclusion The log-logistic with inverse Gaussian frailty model fitted better than other distributions for the chronic diseases data sets. Therefore, considering the hospital as random effects has a significant impact on time-to-death for NCDs patients and therefore, it is recommendable to include frailty to act as covariates for capturing any dependency under clustered time-to-event methods.

Background

Non communicable diseases (NCDs), also known as chronic diseases, tend to be of long duration and are the result of a combination of genetic, physiological, environmental and behaviours factors. NCDs already disproportionately affect low- and middle-income countries when more than three quarters of global NCD death (32 million) occur annually. In African nations, deaths from, NCDs are projected to exceed the combined deaths of communicable and nutritional diseases and maternal and prenatal deaths as the most common causes of death by 2030 (1).

In Ethiopia World Health Organization (WHO) estimated that in 2011 alone 34% of population is dying from non-communicable diseases, with a national cardiovascular disease prevalence of 15%, cancer and chronic obstructive pulmonary disease prevalence of 4% each, and diabetes mellitus prevalence of 2% (2). This WHO estimation is comparable with East African countries, such as Kenya, Uganda, and Eritrea in 2011. Previous studies on NCDs are mainly descriptive in nature and limited to the study of association between related variables. A Few studies have been done on risk factors of NCDs using chi-square and

logistic regression and most of these studies are based on small scale survey data concentrated on certain area. It is important to study factors that may affect the survival time of chronic diseases patients using advanced survival methods. Standard survival data (also called time -to-event data) arise in studies where the time from some origin to an end-point is measured. The end-point is defined by occurrence of a certain event of interest. In this research we assessed various techniques in survival analysis and we applied to chronic NCDs data. The classical proportional hazards model popularized by Cox (3) is assumed identically and independently distributed samples. But, in chronic disease the independence assumption is not reasonable across the hospitals. In that case the frailty models are appropriate for data analysis. The frailty is an unobserved random factor that modifies multiplicatively the hazard function of an individual or a group or cluster of individuals and developed by Vaupel *et al.* (4) for the first time. Clayton (5) relaxed the frailty model to multivariate situation using chronic disease incidence in families' data sets. The random effect, called frailty describes the common risk or the individual heterogeneity, acting as a factor on the hazard function (6–12). In this study I assumed that the patients considered as participants may have different style of life because they came from different community and different geographical locations. Thus, survival patients within the same hospital may have dependence at hospital level due to treatment effects. Hence, shared frailty models were appropriate under such conditions by assuming that patients within the same hospital shares similar risk factors by accounting hospital as hidden heterogeneity. This model assumes a conditional independence where a random effect takes into account the effects of unobserved or unobservable heterogeneity, caused by different sources to all individuals in a hospital (7, 8). Disregarding the correlation among patients' due to hospital level may lead to underestimate standard errors and parameter estimates may be biased and inconsistent (11).

Therefore, this study was explored the effects of prognostic factors and average survival times among NCDs inpatients in different hospitals for random right censoring.

Methods

Study sample and setting

The cross-sectional retrospective follow-up study data was obtained from Adama, Asella and Bishoftu hospitals that are located in the Oromiya National Regional State (ONRS), Ethiopia. Patients were included for this study if they were hospitalized in one of these three hospitals from January 1, 2012 to January 30, 2014 and if they aged 15 or more years. A total of 21,529 in-patients with chronic NCDs were admitted in three hospitals. Of these 646 patients were selected randomly for the study. The outcome variable was the survival time of NCDs patients measured (in months) from date of hospitalized patients enter/start date until the date of death. The event of interest was death and if the patients are not died, I considered as censored outcome either the patients lost to follow-up or still alive or the patients may be died due to other causes which are not related to the study of event of interest. Under this paper *Age, sex, marital status, Place of residence, Educational status, Tobacco use, Alcohol consumption, Family history of chronic disease patients and Diastolic Blood Pressure (DBP)* were considered and assessed as there

are the risk factors to decelerated time-to-death of chronic NCDs patients or not. For educational status we used four categories; 0 not educated 1 primary, 2 secondary and 3 above secondary educated patients. Marital status was classified as 0 for single, 1 married and 2 others. Sex of patients was categorized as 0 female and 1 male. Place of residence was categorized as 0 for urban and 1 for rural. *Tobacco use* was categorized based on whether the patients smoking or not. Alcohol consumers also divided into alcohol consumer or not. Also family histories categorize into no and yes. DBP categorized into three 0 low, 1 normal and 2 abnormal patients. The statistical methods tools such as exponential, Weibull, and log logistic hazard functions were applied and compared for their efficiencies and for the frailty distributions, Gamma and inverse Gaussian distributions were assumed. For comparison of different distributions, the Akaikie Information Criteria (AIC) criteria was used, but for comparing nested models, Likelihood Ratio Test (LRT) was used. Data were analyzed using R statistical software version 3.2.1.

Shared frailty model

The general formula for shared frailty model is written as follow. Given the random effects, denoted by wi, the survival times in hospital i ($1 \leq i \leq n$) are assumed to be independent and the proportional hazard frailty model is formulated as:

$$h_{ij}(t / X_{ij}, w_i) = \exp(\beta' X_{ij} + w_i) h_o(t)$$

When the proportional hazards assumption does not reasonable, an alternative model is the parametric frailty model which is given as:

$$h_{ij}(t / X_{ij}, w_i) = h_o(\exp(\beta' X_{ij} + w_i)t) \exp(\beta' X_{ij} + w_i)$$

Where i indicates the ith hospital and j indicates the jth individual for the ith hospital, ho(.) is the baseline hazard, wi the random effects of all the subjects in hospital i, Xij the vector of covariates for subject j in hospital i, and β the vector of regression coefficients.

We assumed that *Z* (where Z = exp (wi)) has the gamma or inverse Gaussian distribution so that the hazard function depends upon this frailty that acts multiplicatively on it. The main assumption of a shared frailty model is that all individuals in hospital i share the same value of frailty Zi i = 1... n. The survival time was assumed to be conditionally independent with respect to the shared (common survival times) frailty. The shared frailty was the cause of dependence between survival times within the hospitals.

In this paper, firstly the univariable analysis was done independently to select the significant variables that were involved in the multivariable analysis.

The multivariable survival analysis was employed under the distributions such as Exponential, Weibull and Log-logistic for the baseline hazard function and the gamma and inverse Gaussian were considered for frailty distributions. Univariable analysis had done using the following covariates: *age, education status, alcohol consumption, tobacco use and place of residence of patients.* However, *sex, marital status, DBP* and *family history* were omitted because there were insignificant in univariable analysis.

Results

In this paper, samples of 646 chronic non-communicable hospitalized patients were considered. The medical cards of those patients were reviewed, out of which 46.6% were females and 53.4% were males. Among those patients, 41.5% died while 58.5% censored in the hospital as shown in Table 1.

Under multivariable frailty models: education level, alcohol consumption and tobacco use were significant covariates. These covariates were the prognostic factors for time-to-death of NCDs patients in the three hospitals. However, the patient age and place of residence had no significant effect on the death of inpatients of chronic NCDs in three hospitals. Hospitality had significant effect for both weibull and log-logistic frailty models at 5% level of significance. The smallest AIC value 2014.493 was appeared for the Log-logistic Inverse Gaussian Frailty (LIGF) model when compared to others models considered in this paper. This value showed that the LIGF model was best model to describe the Chronic NCDs dataset using different types of parametric frailty models (See Table 2).

Based on LIGF model, secondary school educated, above secondary educated patients, alcohol consumption and tobacco use were significant. The confidence interval of acceleration factor () of all secondary educated, above secondary educated patients, alcohol consumption and tobacco user excluded 1, at 5% level of significance (Table3).

The value of the shape parameter $\rho = 1.955$ in the best model is greater than one indicating that the shape of hazard function was unimodal that means it increased up to some time and then decreased. The predicted heterogeneity in the population of patients among the hospitals was 0.091 , and the correlation within hospital τ was about 22.3%.

The predicted hospitality effect values increased with range of 0.933 to 1.081 with increasing the median time of the hospitals for chronic NCDs dataset (see Figure1). This means, the estimated hospital effect values were lower for lower values of event times and higher for higher value of event times among the hospitals.

The relative risk functions obtained by the 25^{th}, 50^{th} and 75^{th} quantiles of frailty distribution ($z = 0.933$, $z = 0.972$ and $z = 1.081$ respectively) were plotted for the LIGF model (Figure2). Using the maximizing the likelihood approach, the LIGF model estimated parameters were $\tau = 0.223$, $\rho = 1.955$ and $\theta = 0.091$.

From the result of this study the frailty value(z>1) showed larger heterogeneous between hospital effects in which patient may die earlier and when z<1 indicated the less heterogeneous hospital effect in which patient has prolong their length of life time for the conditional hazard functions given by the 75^{th}, 50^{th} and 25^{th} quantile frailty values. Entirely, the conditional hazard functions were equal at the beginning time (t = 0). but, the gap expanded through time, specifically at middle time. The hazard function patterns followed unimodal (increased up to some point and then decreased) because the shape parameter for the baseline hazard function was greater than one ρ = 1.955.

To check how well the baseline hazard was fitted the cumulative hazard function vs. the time for the exponential; the logarithm cumulative hazard function vs. the logarithm of time for Weibull and the logarithm of the failure odds vs. the logarithm of time for the log-logistic models was drawn (Figure3). The plot of Weibull was closest to reference line and approximately linear than the other plots, although only few observations were scattered at the beginning time. The patterns suggested that the weibull hazard function was appropriate in the model.

The Cox-Snell residuals together with their cumulative hazard function were obtained by fitting the exponential, Weibull and log-logistic models to chronic NCDs dataset, through maximum likelihood estimation (Figure 4). The plots showed that the Cox-Snell residuals fitted to assess the log-logistic model for the dataset were approximately nearest to the line through the origin as compared to another models, again indicating that the log-logistic model fitted the NCDs dataset well.

A quantile-quantile (q-q) was plotted to check if the AFT provided how well fit to the data using by two different survival groups. Graphically the adequacy of the AFT model was checked by comparing the significantly different variables: education level (not educated and secondary educated patients; not educated and above secondary educated patients), no alcohol consumption and alcohol consumption patients; similarly the tobacco user or non-user patients (Figure 5). Based on the figure, nearly the plots seems as to be linear for three covariates such as education level, alcohol consumption and tobacco use with slopes of about the acceleration factors 0.52, 0.405, 1.631 and 1.659 respectively. Therefore, the log-logistic was considered as baseline than both exponential and Weibull distribution for parametric frailty model to fit Chronic NCDs dataset.

Discussion

The main objective of this study was modeling average death time of CNCDs dataset under different types of parametric frailty models. AIC criteria and LRT methods were employed to compare the distributions of the models and a model with minimum AIC and LRT was selected to be the best (12, 13). Among the considered models, the LIGF model has minimum AIC value of 2014.493 which described the NCDs dataset correctly when compared with other models for this paper. This study was agreed with Banbeta *et al.* (13) study an application of frailty model on severe acute malnutrition data set.

The finding of this study indicated that the expected survival time of NCDs inpatients is lower for no educated patients, or if the patient has previously experienced NCDs. The expected survival time also decreased with alcohol consumption and for tobacco users for hospitalized patients. This result is consistent with studies in the literature, but researchers tend to examine the effects of covariates on patients using logistic regression model (14). The median survival time of NCDs hospitalized patients were 12 months. The leading causes of death in this study were Chronic Heart Failure (CHF) diseases, followed by, ischemic heart failure (IHF) diseases and diabetes mellitus (DM).

This study showed that types of hospital may influence on the survival time of chronic NCDs patients which might be due to the treatment effects, the infrastructures problem, due to insufficient skill of physicians in the hospital and others factors. Thus, taking into account the hospital effect was reasonable in modeling the relative risk for clustered survival model for prediction purpose. Patients with smallest median time were expected to have smaller heterogeneity and a bigger risk of hazards (8). This indicated that the patients may lead to death in short time in the respective hospital. Hospital that has greater variability has more likely larger death than the less variable hospital.

The random effect distribution emerged by Hougaard (16) as an alternative to gamma distribution was better in my dataset when compared to the gamma frailty distribution. The variability in the hospitals was estimated to be $\theta = 0.091$, and the correlation within hospitals was $\tau = 22.3\%$.

This study showed that NCDs data set was best fitted by the log-logistic baseline as compared to the exponential and weibull hazard functions. According to the diagnostic plots, the *log failure odds* of log-logistic baseline with *log time* was more linear as compared to the plots of exponential (*cumulative hazard* versus *time)* and weibull (*log cumulative hazard* versus *log time)*,, showing the NCDs dataset was best modeled by the log-logistic as baseline. This result was also checked by the cumulative hazard plots for the Cox-Snell residuals of the exponential, weibull and the log-logistic distributions. The plot was more closed to the reference line in the log-logistic distribution which indicated the log-logistic baseline was best fit my dataset. Also a q-q plot was fitted to check if the accelerated failure time provided an adequate fit to NCDs dataset and the log-logistic as baseline was good fit the NCDs data than others.

The significant prognostic covariates under univariable analysis were *age, education status, alcohol consumption, tobacco use and place of residence of patients.* The result of LIGF model showed that the *education status, alcohol consumption, tobacco use* were the significant prognostic factors for the time-to-death of chronic NCDs. This is in line with studies conducted in Cuba on frailty prevalence and its effect on dependency incidence and mortality in older adults, as well as its associated risk factors (17) and the study resulted that age, sex and lower education level was not related to frailty of mortality.

Conclusion

Among several parametric frailty models, LIGF model fitted well the time-to-death of chronic NCDs dataset in the three hospitals of ONRS. The study showed that hospital level has a significant effect on the survival time of the patients of chronic NCDs. The median death time of the patients was about 12

months with maximum time of 40 months of which 41.5% were died. Under this study, although chronic NCDs patient has long span of progression, the median of time-to-death was very short because of the patients go to a hospital after the diseases become in serious conditions period.

Based on the result of this study education level, alcohol consumption and tobacco use were identified as the most prognostic factors that decelerated the survival time of non-communicable diseases patients in the three public hospitals. We recommend that interventions such as mass media campaigns peer /outreach education and life skill programs be introduced to educate patients' for early diagnosis and to follow-up their health problem of NCDs carefully.

Abbreviations

AFT

Accelerated Failure Time

AIC

Akaikie Information Criterion

CHF

Chronic Heart Failure

DBP

Diastolic Blood Pressure

DM

Diabetes Mellitus

IHF

ischemic heart failure

LIGF

Log-logistic Inverse Gaussian Frailty

LRT

Likelihood Ratio Test

NCD

Non-Communicable Disease

ONRS

Oromiya National Regional State

WHO

World Health Organization

Declarations

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Consent to publish

Not Applicable, because there was no picture or/and video taken for this research article.

Ethics approval and consent to participate

The study was approved by the graduate council of department of statistics and Haramaya University. Letter of support was obtained from school of graduate study, Haramaya University. Authorities of the respective hospitals were informed about the study and they have granted us full access to documents. The outcome of the study was communicated to the Hospitals authorities to use as guidance in their future health care delivery.

Competing interests

I declare that there is no competing interest.

Funding

I thank College of Computing and Informatics, Haramaya University for providing the funding for this study.

Authors' Contributions

BM designed the idea, analyzed and interpreted the data, and drafted and edited the manuscript. An author read and approved the final manuscript.

Acknowledgements

An author gratefully acknowledged the three public hospitals in Oromiya National State and College of health and medical for helping me during data collection.

Authors' Information

[1]Department of Statistics, College of Natural and Computational Science, Addis Ababa University, Addis Ababa, Ethiopia

References

1. WHO, 2018: Report on Non communicable diseases. (Accessed on 11 May 2019)
2. World Health Organization: Global status report on alcohol and health 2011. Geneva. http://www.who.int/substance_abuse/publications/global_alcohol report/en/ (accessed on: June 14, 2019).
3. Cox D: Regression Models and Life Tables. Series B (Methodological). J Roy Stat Soc 1972, 34: 187–220.
4. Vaupel J, Manton K, Stallard E: The Impact of Heterogeneity in Individual Frailty on the Dynamics of Mortality. Demo 1979, 16: 439–454.
5. Clayton D. G: A Model for Association in Bivariate Life Tables and Its Application in Epidemiological Studies of Familial Tendency in Chronic Disease Incidence. Biometrika 1978, 65:141–151.
6. C. M. A. Wintrebert, A. H. Zwinderman, E. Cam, R. Pradel and J. C. van Houwelingen (2005). Joint Modelling of Breeding and Survival in the Kittiwake Using Frailty Models. Ecological Modelling 181: 203 –213.
7. Alan J: Parametric methods in the analysis of survival data. Mic elec Rel 1980, 20:477–481.
8. Hougaard P: Analysis of Multivariate Survival Data. Springer-Verlag, Newyork; 2000.
9. Collett D: Modelling survival data in medical research, Second Edition. Chapman and Hall/CRC, London; 2003.
10. Wienke A: Frailty Models in Survival Analysis. Chapman and Hall/CRCbiostatistics series; 2011.
11. Duchateau L. and Janssen, P: The Frailty Model. Springer-Verlag, New York; 2007.
12. Munda M: Parametric frailty models in R. Am Stat Assoc 2012, 55:1–21.
13. Banbeta et al.: Modeling time-to-cure from severe acute malnutrition: application of various parametric frailty models. Archives of Public Health 2015 73:6. Doi: 10.1186/2049–3258–73.

14. Ayalew Tegegn, Muluneh, Abraham Haileamlak, FasilTessema, Fessahaye Alemseged, Kifle Woldemichael, Makonnen Asefa et al: Population based survey of chronic non-communicable diseases at Gilgel Gibe field research center, southwest Ethiopia. Ethiopian J Health Sc 2012, 22:7–18.
15. Hougaard P: Frailty Models for Survival Data. Lifetime Data Analysis 1995, 1: 255–273.
16. Hougaard P: Life table methods for heterogeneous populations: Distributions describing heterogeneity. Biometrika 1984, 71: 75–83.
17. Jesús L, López A, Valhuerdi A, Guerra M, Llibre-Guerra J, Sánchez S, Bosch R, Zayas T and Moreno C: Frailty, Dependency and Mortality Predictors in a Cohort of Cuban Older Adults, 2003–2011. MEDICC Review 2014, 1:16.

Tables

Table 1. Descriptive results of chronic disease death events vs. different demographic, health and risk behavior variables

Covariates	Censored	Death (%)	Total	Median (in months)
Sex				
Female	189	112(37.2)	301	13
Male	189	156(45.2)	345	11
Marital status				
Single	74	32 (30.2)	106	13
Married	252	173 (40.7)	425	12
Others	52	63 (54.8)	115	11
Education level				
No educate	183	194 (51.5)	377	11
Primary	80	50 (38.5)	130	13
Secondary	62	14 (18.4)	76	27
Above	51	10 (16.4)	61	NA
Place of residence				
Urban	160	80 (33.3)	240	13
Rural	218	188 (46.3)	406	11
Alcohol consumption				
No	201	71 (26.1)	272	21
Yes	177	197 (52.4)	376	11
Tobacco use				
No	247	152 (38.1)	399	14
Yes	131	116 (47)	247	11
Type of disease				
Diabetes	110	68(38.2)	178	16
CHF	84	86 (50.6)	170	11
Stroke	73	73 (50)	146	11
Others	111	41 (27)	152	14

Covariates	Censored	Death (%)	Total	Median (in months)
Diastolic blood pressure				
Low (<60)	72	140 (66)	212	10
Normal (60-90)	258	21 (7.5)	279	NA
High (>90)	48	107 (69)	155	10
Family history				
No	263	183 (41)	446	12
Yes	115	85 (42.5)	200	12
Over all	378	268 (41.5)	646	12

Table 2. Comparison of frailty models for CNCDs data

Baseline function	Frailty distribution	LRT	AIC
Exponential	Gamma	-1029.726	2077.451
	Inverse-Gaussian	-1029.85	2077.700
Weibull	Gamma	-1000.294	2020.588
	Inverse-Gaussian	-1000.293	2020.587
Log-logistic	Gamma	-997.263	2014.526
	Inverse-Gaussian	-997.247	2014.493

Table 3. Log logistic – inverse- Gaussian multivariable frailty model

Variables		se()		p-value	95% CI LCL	95% CI UCL
Age	0.009	0.005	1.009	0.058	1.000	1.019
Education level (no education R)						
Primary	-0.142	0.199	0.868	0.477	0.587	1.282
Secondary	-0.654	0.301	0.520	0.030 *	0.288	0.938
Above	-0.903	0.342	0.405	0.008 **	0.207	0.793
Place (urban R)	0.257	0.147	1.293	0.081	0.969	1.725
Alcohol(No R)	0.489	0.151	1.631	0.001 ***	1.212	2.194
Tobacco use (No R)	0.506	0.130	1.659	0.000 ***	1.284	2.140
θ = 0.091			AIC = 2014.493			

Figures

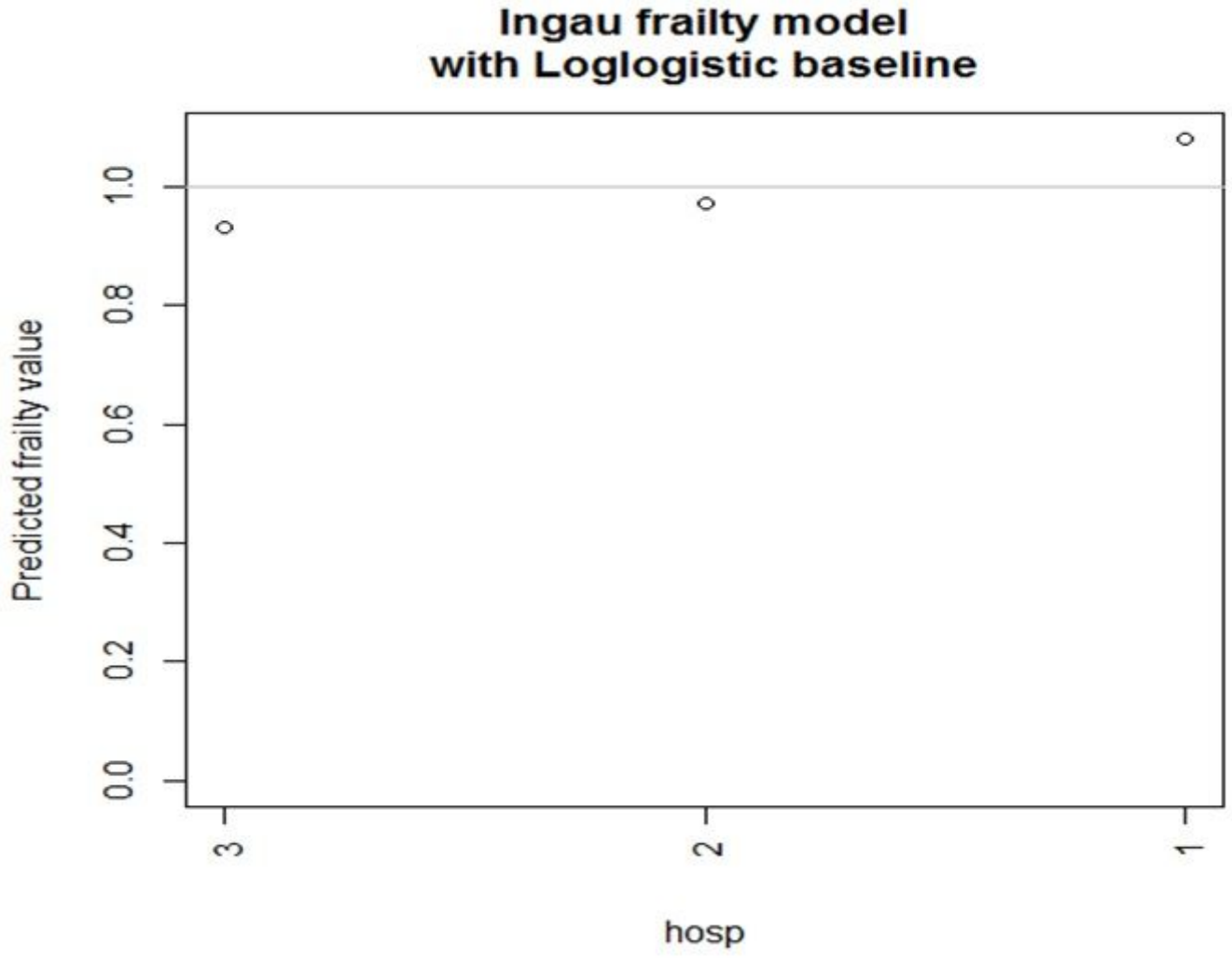

Figure 1

Prediction of frailties for the CNCDs dataset as given by the parametric log logistic-inverse Gaussian frailty model

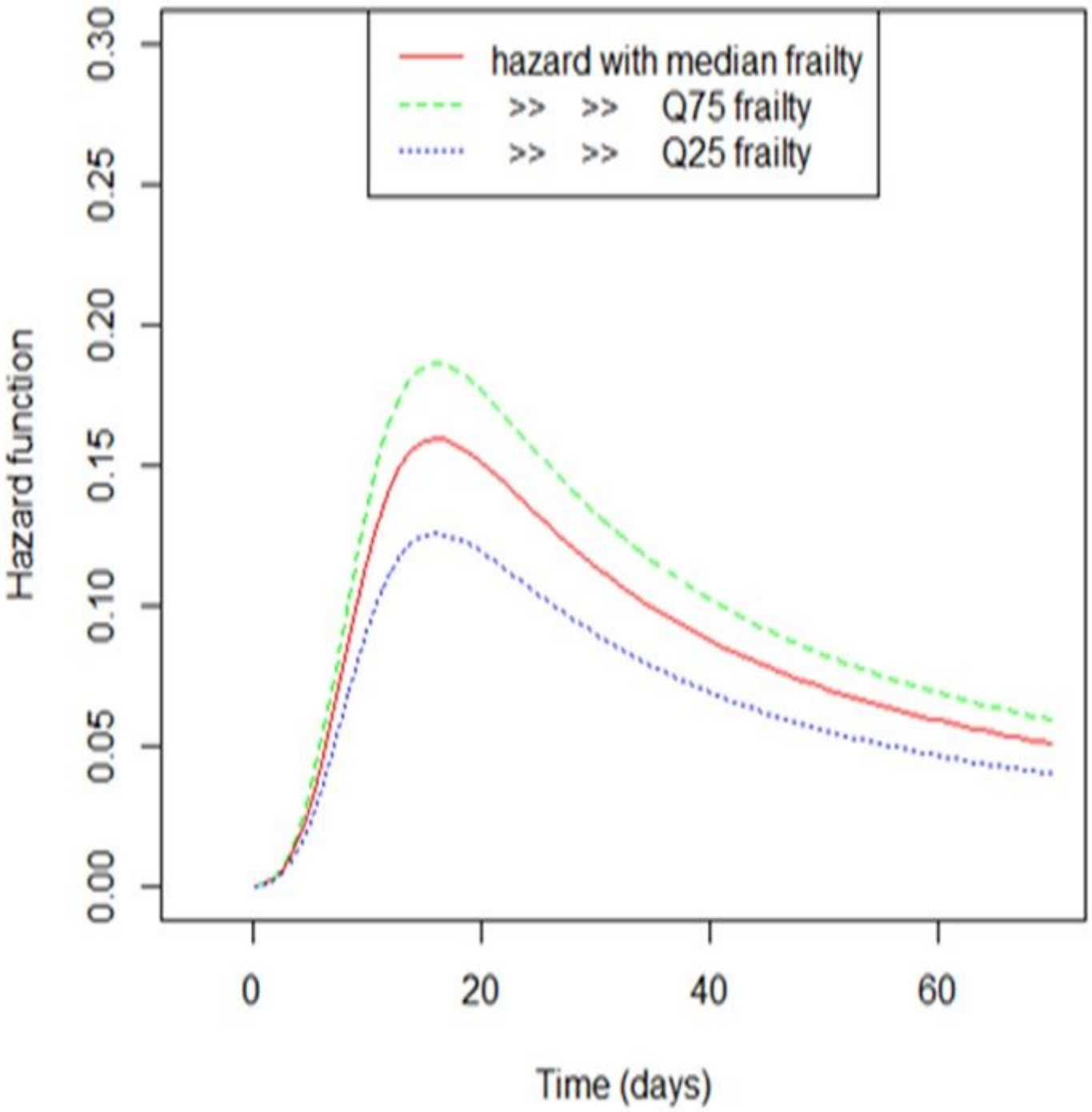

Figure 2

Conditional hazard rates of the log-logistic- inverse Gaussian frailty model for the CNCDs dataset

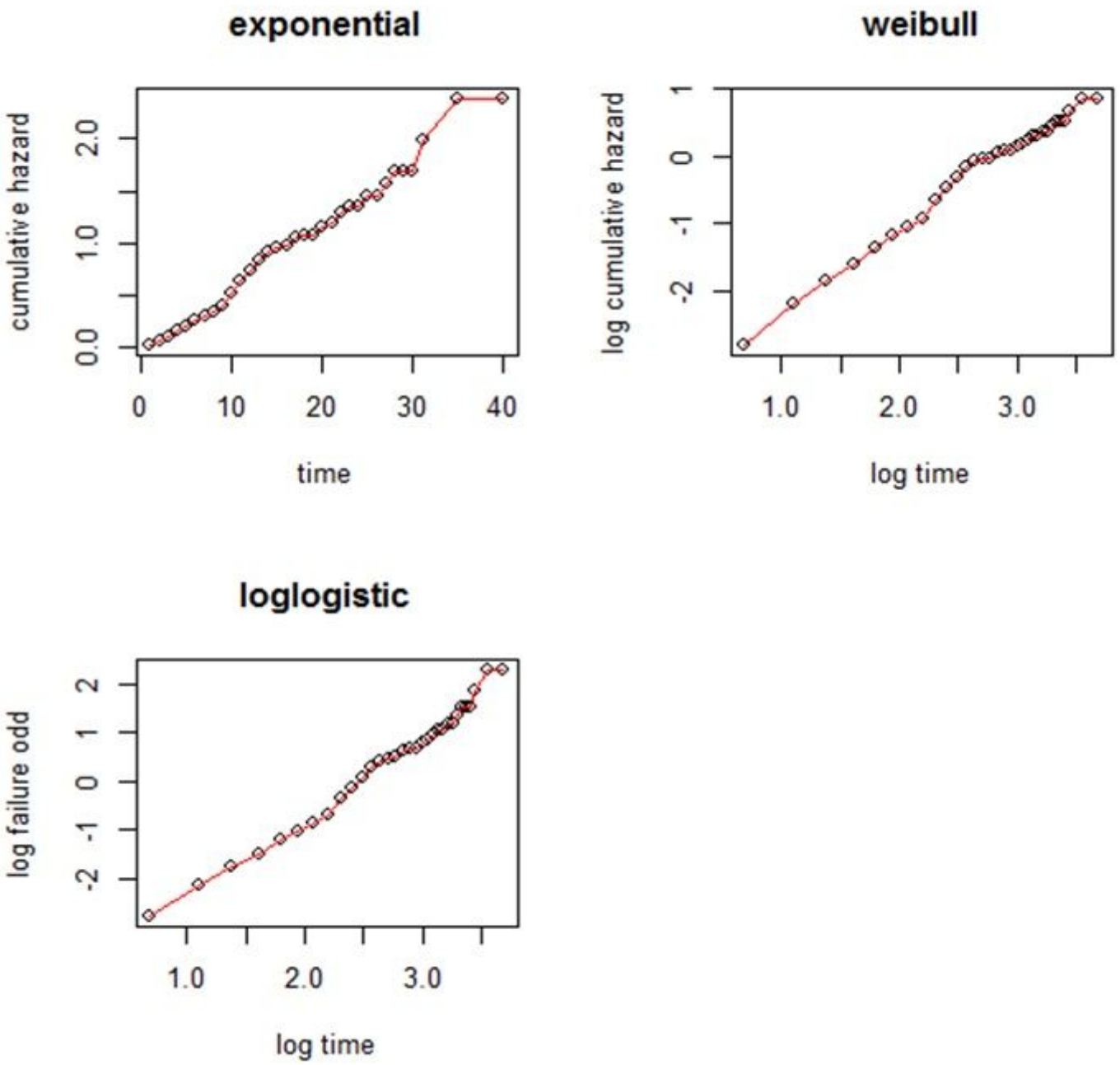

Figure 3

Graphical evaluation of the exponential, weibull and log-logistic assumptions

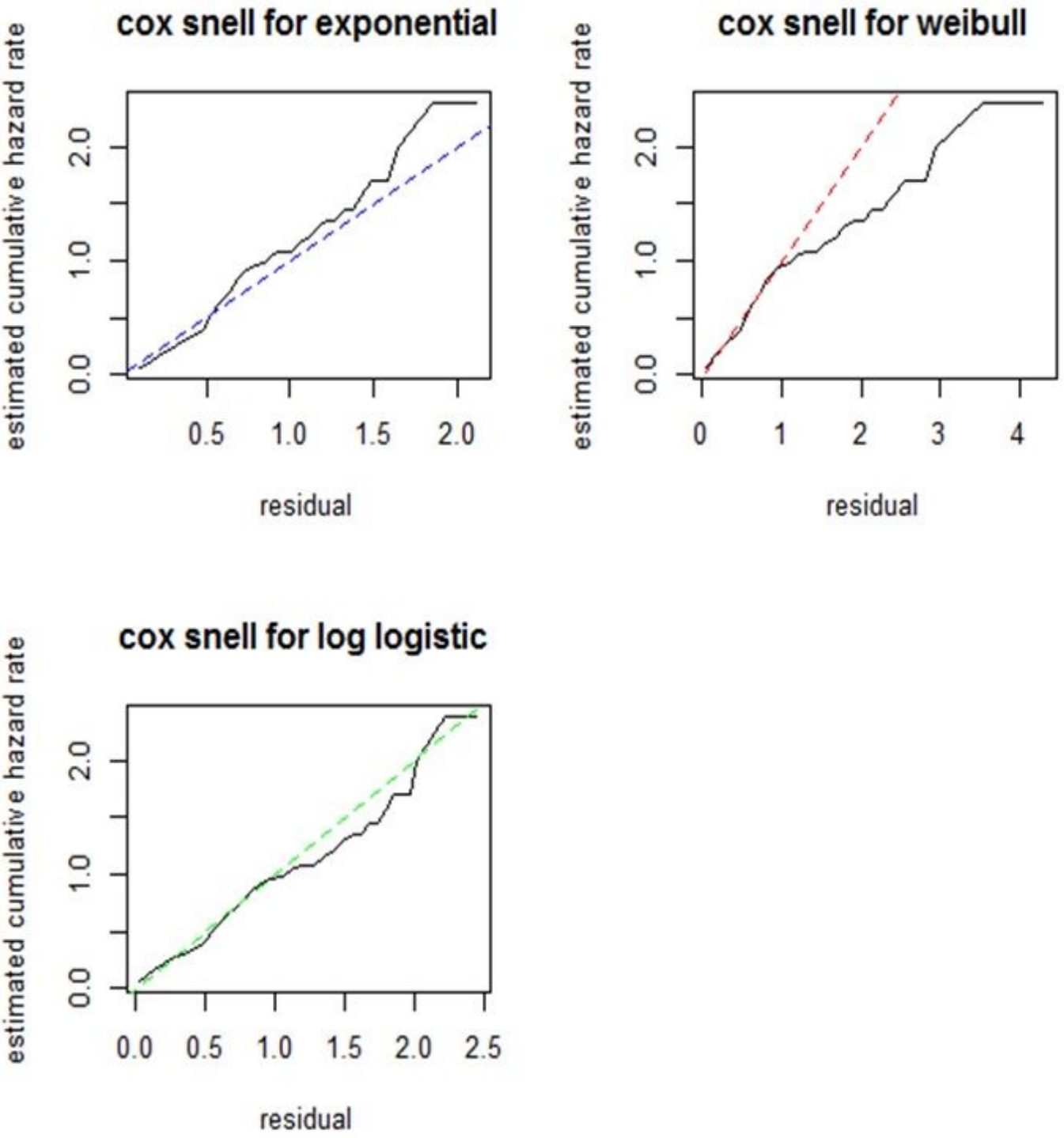

Figure 4

Cox-Snell residuals for exponential, weibull and log logistic models to the CNCDs data

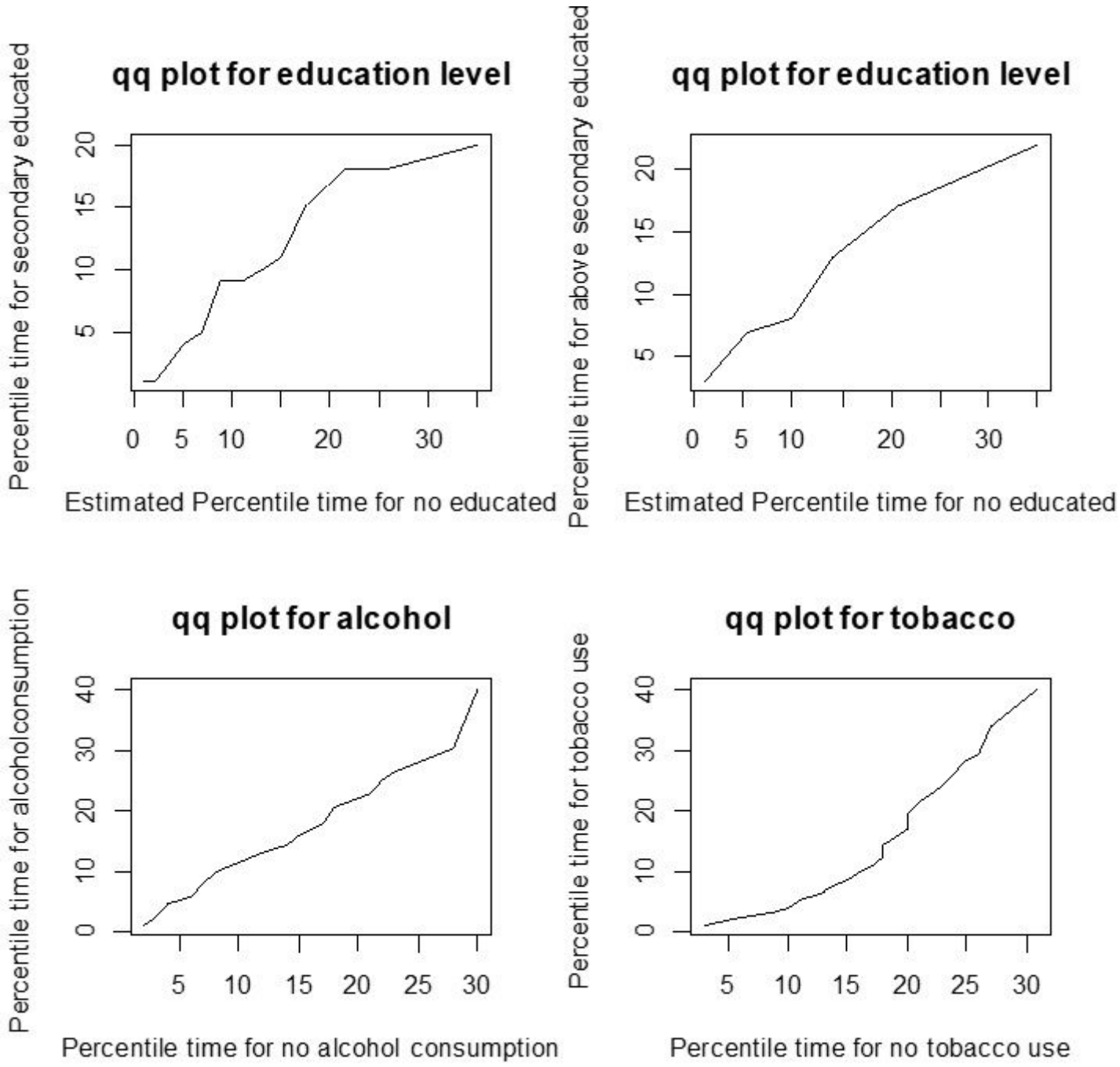

Figure 5

Q-Q plot to check the adequacy of the accelerated failure time model