1

<u>Original Research Article</u>

3

Application of Multiplicative and Additive Hazards Models to Injury Prevention among Healthcare Workers

6
7
8
9

10

ABSTRACT

12
13

The Cox multiplicative model is used widely in survival analysis, where the covariates act multiplicatively on unknown baseline hazards. However, the Cox model requires the proportionality assumption, which limits its applications. The additive hazards model has been used as an alternative to the Cox model, where the covariates act additively on unknown baseline hazards. In this study, performance of the Cox multiplicative hazards model and the additive hazards model has been demonstrated, using an injury prevention study. Both the multiplicative and additive hazards models showed similar results in selecting significant covariates in the final model in our study. The coefficient of the covariates in the additive hazards model is easy to interpret in an additive manner and should be considered when the proportionality assumption of the Cox model is doubtful. The multiplicative and additive hazards models describe different features of the association between the risk factors and the study outcomes. They may be used each other as supplementary approach for further understanding of the data.

27
28
29
30
Keywords: survival analysis, Cox model, multiplicative model, additive model, injury prevention, healthcare worker

33

34

35

36

37

38

39

40

41
42   **1. INTRODUCTION**

43   In survival analysis, the Cox hazards model [1] is the most widely used in survival

44   analysis. In this model, the effect of the covariates acts multiplicatively on some

45   unknown baseline hazard. However, when the proportionality assumption is not satisfied,

46   the Cox model can lead to potentially biased estimates and conclusions [2]. Alternatively,

47   additive hazards model has been proposed. The additive hazards model assumes that the

48   covariates act in an additive manner on an unknown baseline hazard. Aalen's additive

49   model as a non-parametric approach specifies how the hazard rate depends on covariates

50   in a linear way and allows one to assess possible changes in the influence of the

51   covariates over time [3]. The estimation procedure for Aalen's model was determined by

52   the cumulative regression functions. By several authors, applications of Aalen's model

53   have been described and further development has been recommended [4-9]. Aalen's

54   approach leads to weighted comparisons of the crude estimate of the hazards rate of each

55   group as compared to a baseline group [10]. This weighting leads to inconsistent tests in

56   the sense that the test statistic depends on which group someone picks as the baseline

57   group.

58

59   The Lin and Ying observed that this lack of progress is attributed to the fact that the

60   partial likelihood approach cannot be used directly to eliminate the baseline hazard in

61   estimating the intercept [11, 12]. They have developed procedures with high efficiencies

62   for making inferences about the regression parameters under the additive hazards model

63   with an unspecified baseline hazards function. In their study, a simple semi-parametric

64   estimating function for the intercept was constructed, which imitated the martingale

65   feature of the partial likelihood score function for baseline hazards. In the subsequent

66   paper, they suggested the semi-parametric analysis of general additive-multiplicative

67   hazard models for the counting process and the additive hazards regression models for

68   survival data but compared these with the frailty model [13]. Others applied the additive

69   hazards model to competing risks setting [14, 15]. Yin and Cai [16] proposed an additive

70   hazard model for multivariate failure time data. Lim and Zhang [17, 18] compared both

71   the additive and multiplicative hazards models in recurrent event data. As an extension, a

72   flexible additive-multiplicative hazard model based on Aalen's and Cox's models have

73   been proposed [19-24]. For additive-multiplicative hazard model, some covariate effects

74   are believed to result in multiplicative effects whereas other effects are best described as

75   additive. However, in practice, it is not easy to decide which covariates to be included

76   additively and which ones to be included multiplicatively. For the additive model, plots

77   of the cumulative regression function provided an appealing explanation for how the

78   hazards profiles were distributed.

79

80   **2. STUDY DESCRIPTION**

81   Patient handling injuries are common among health care workers and the risk of injury

82   increases with the number of patient handling tasks performed. Studies showed that a

83   transfer, lifting and repositioning (TLR) program may prevent injuries while performing

84   one type of manoeuvre and not another depending on the emphasis of the intervention.

85   To evaluate patient handling injuries following a multi-factor ergonomic intervention

86   program among health care workers, a quasi-experimental study which had a TLR

87   intervention group and a non-randomized control group was conducted. Descriptions of

88   the overall study design and profile have been published elsewhere [25, 26]. Briefly, this

89   study was conducted in two Health Regions (3 hospitals for the intervention and 3 for the

90   control) in Saskatoon, Canada, from September 2002 to December 2006. The hospitals

91   were matched on hospital type and size. The TLR intervention program component

92   consisted of staff education on anatomy, injuries, body mechanics, personal health, lifting

93   and patient handling procedures, standardized patient handling needs assessment and

94   patient handling algorithms. All direct health care workers, who were employed as such

95   in the study time periods, were eligible for inclusion into the study. Injuries occurred in

96   lower and upper back, shoulder, neck, extremity, and other body parts were included. The

97   control hospitals had not received any form of injury prevention program during the

98   study period other than standard occupational health and safety practice. Each

99   intervention and control hospital was followed for two year after completion of the

100   intervention program. Gender, age, occupation type, work department, and hospital size

101   were also obtained from the database. The primary outcome was the times to the event of

102   TLR related injury occurring in subjects during the study time. A total of 1,467 subjects

103   were eligible for the study.

104   With this example, we use three models (Cox multiplicative hazards model, Aalen's

105   additive hazards model, and Lin & Ying's additive hazards model) (i) to determine which

106   combination of potential explanatory variables affects the form of the hazard function and

107   (ii) to obtain an estimate of the hazard function itself for an individual. We will also

108   examine the goodness-of-fit analysis of the models.

109

110   **3. MODELS AND METHODS**

111   Within the framework of the multiplicative or additive hazards regression models, a

112   variety of models have been proposed and utilized in real applications. The Cox

113   multiplicative and Lin & Ying's additive hazards models received the greatest attention

114   due to relatively easy interpretation of the covariate effects. These two models assume

115   unspecified baseline hazards and constant covariate effects. In our study, we will assume

116   that all censoring is non-informative and independent, i.e., knowledge of a censoring time

117   for a subject provides no further information about the subject's likelihood of survival at

118   a future time.

119

120   *3.1. Basic Notations*

121   Suppose that there are n subjects in a study. Let T_i be the time when the event of interest

122   occurs for the ith subject and C_i be the corresponding censoring time. T_i is measured from

123   the subject's study enrollment and the censoring C_i occurs after the subject has been

124   entered into a study to the right of the last known failure time; thus, it is right censoring.

125   When T_i is subject to right censoring, the failure time X_i is a minimum of (T_i, C_i), i.e., X_i

126   is equal to T_i if the event was observed and is equal to C_i if it is censored. Let $\delta_i =$

127    $I(T_i \leq C_i)$, where $I(.)$ is an indicator function and takes the value 1 when $T_i \leq C_i$ and is 0

128   otherwise. Let Z_i be a covariate vector of p-dimensions for the ith subject. The hazard

129   function for the ith subject, $\lambda_i(t)$, is assumed to take either multiplicative or additive

130   forms.

131

132   *3.2. Multiplicative Hazards Model*

133   The Cox model is one of the most commonly used multiplicative hazards models.

134   The effect of the covariates in the Cox model was to act multiplicatively on some

135   unknown baseline hazards. The model is very useful in practice because either the

136   estimated coefficients themselves or simple functions of them can be used to provide

137   estimates of hazard ratios. In addition, statistical software is readily available, and it is

138   easy to fit models, check model assumptions, and assess model fit.

139

140   For Cox proportional hazards model, the hazard function is

141    $$\lambda(t) = \lambda_0(t)\, e^{\vec{\beta}' \vec{Z}(t)} \tag{1}$$

142   where t is the time since a subject's study enrollment. Note that $\lambda_0(t)$ are unspecified

143   baseline hazard functions. The corresponding partial likelihood function [2] is

144    $$L(\vec{\beta}) = \prod_{i=1}^{n} \left\{ \frac{e^{\vec{\beta}' \vec{Z}_i(X_i)}}{\sum_{j=1}^{n} Y_j(t)\, e^{\vec{\beta}' \vec{Z}_j(X_j)}} \right\}^{\delta}, \tag{2}$$

145   where $Y_j(t) = I(X \geq t)$ is a risk set indicator. $\vec{\beta}$ is a p-vector of regression coefficients of

146   Z_i. In order to draw a semi-parametric inference on $\vec{\beta}$ for the model (1), the score

147   functions $U(\vec{\beta})$ are obtained by differentiating the logarithm of $L(\vec{\beta})$ with respect to $\vec{\beta}$.

148   The maximum partial likelihood estimator $\hat{\vec{\beta}}$ is obtained by solving the corresponding

149   score equation, $\dfrac{\partial \ln L(\vec{\beta})}{\partial \vec{\beta}} = 0$. The variance-covariance matrix is estimated from the

150   inverse of the information matrix, $I^{-1}(\hat{\vec{\beta}})$.

151

152   *3.3. Additive Hazards Model*

153   The simple additive hazards model given by Cox and Oakes [27] is $h(t|Z) = h_0(t) + \varphi(Z)$

154   where $\varphi(0)=0$ and $\varphi(Z)$ is constrained so that the right-hand side is non-negative. $h_0(t)$ is

155   the baseline hazard and the covariates act in an additive manner on an unknown baseline

156   hazards rate. Aalen's additive model 3, 28] and Lin and Ying's additive models (L-Y

157   model) [11] have received great attention in the literature. In Aalen's model, the

158   unknown risk coefficients are allowed to be functions of time so that the effect of a

159   covariate may vary over time. The least-squares approach is used to estimate the

160   cumulative regression functions and the standard errors of these functions [29]. In the L-

161   Y model, the time-varying regression coefficients in Aalen's model are replaced by

162   constants and the estimating equation is obtained from the score function to estimate the

163   model. In the next section, these additive hazards models will be reviewed.

164

165   *3.3.1. Aalen's additive hazards model*

166   In the Aalen's additive hazards model, the covariates are assumed to impact additively

167   upon an unknown baseline hazard, but the effects are not constrained to be constant [28].

168   Thus, the hazard function under the Aalen's model for the *i*th subject with a *p*-vector of

169   the covariates $Z_i = (z_{i1}, \ldots, z_{ip})$ is defined as:

170          $$\lambda_i(t) = \lambda_0(t) + \gamma_1(t)\, z_{i1}(t) + \ldots + \gamma_p(t)\, z_{ip}(t). \tag{1}$$

171   where $\lambda_0(t)$ is an unspecified baseline hazard function, and coefficient $\gamma_k(t)$ is allowed to

172   vary freely over time, where $k = 1, 2, \ldots, p$. Aalen shows that if a covariate is

173   independent of all the other covariates in the model, then the regression model with this

174   covariate eliminated is the same as the regression model with this covariate included [28].

175   Note that this fact is not true for the Cox proportional hazards model. The additive effect

176    $\gamma_k(t)$ may change in magnitude and even sign with time. As it is not straightforward to

6

177   estimate $\lambda_0(t)$ non-parametrically, direct estimation of the coefficient $\gamma_k(t)$ is difficult.

178   Aalen and others [8, 28] have developed least square estimation of integrated coefficients

179    $$\Gamma_k(t) = \int_0^t \gamma_k(u)\ du\ .$$

180

181   The usual method of representing the effect $\gamma_k(t)$ is to graph them against time. To define

182   how the effects of covariates changes over the time, cumulative regression function plots

183   estimated by the Aalen's model can be examined. The values of $\gamma_k(t)$, the absolute

184   increase in hazard at time t, are not actually observed, but their relative size may be

185   inferred from the slope of the line. The Aalen's plots are obtained by estimating the

186   instantaneous contributions of covariates to the hazard at each distinct failure time and

187   summing up the resulting estimates. The slop of such plots indicates whether a specific

188   covariate has a constant or a time-dependent effect [6]. Slope of an estimated cumulative

189   regression function is positive when covariate increase corresponds to hazard increases,

190   and negative when covariate increases correspond to hazard decrease. Cumulative-sums

191   slop approaches zero when a covariate has no effect on the hazard.

192

193   *3.3.2. Lin & Ying's (L-Y) additive hazards model*

194   We know from Aalen's additive hazards model the conditional hazards rate of a subject,

195   given a set of covariates, and that the regression coefficients are the function of time. Lin

196   and Ying proposed an alternative additive hazards regression model, which is the most

197   closely connected and analogue to the Cox model [11-13]. The L-Y additive hazards

198   model for the ith subject with covariate vector $Z_i = (z_{i1}, \ldots, z_{ip})$ is $\lambda_i(t),$

199    $$\lambda_i(t) = \lambda_0(t) + \gamma_1\ z_{i1}(t) + \ldots + \gamma_p\ z_{ip}(t).$$

200   The covariates are assumed to act additively on a baseline hazard $\lambda_0(t)$ and coefficient

201   γ_k is constant additive effects, where $k = 1, 2, \ldots, p$. Lin and Ying [11] propose a

202   heuristic estimation method based on a estimating equation due to the Cox's partial

203   likelihood. Their method successfully treats the baseline hazard as nuisance and removed

204   them from estimating the regression coefficients. Using the counting process and

205   martingale approach, they obtained closed-form estimators for the regression parameters

206   and the cumulative baseline hazard function.

207

208   In order to draw semi-parametric inference on the coefficient $\vec{\gamma}$ for model, the key

209   quantities are given by:

210
$$U = \sum_{i=1}^{n} \int_{0}^{\tau} [Z_i(t) - \bar{Z}(t)] \, dN_i(t)$$

211
$$A = \sum_{i=1}^{n} \int_{0}^{\tau} [Z_i(t) - \bar{Z}(t)]^{\otimes 2} \, Y_i(t) \, dt$$

212
$$B = \sum_{i=1}^{n} \int_{0}^{\tau} [Z_i(t) - \bar{Z}(t)]^{\otimes 2} \, dN_i(t)$$

213

214   where, for any vector a, $a^{\otimes 2} = aa^T$; τ is a pre-specified time point usually set to

215    $\max\{ X_1, X_2, \ldots, X_n \}$ such that all observed failures are included in the analysis, and

216
$$\bar{Z}(t) = \sum_{i=1}^{n} Y_i(t) Z_i(t) \sum_{i=1}^{n} Y_i(t)$$

217   is the at-risk weight covariate mean at time t. Lin and Ying [11]proposed to estimate $\vec{\gamma}$

218   by

219
$$\hat{\vec{\gamma}} = A^{-1} U,$$

220   while the estimated variance of $\vec{\gamma}$ was derived to be:

221
$$\hat{V}(\hat{\gamma}) = A^{-1} B A^{-1}.$$

222

223   Here, neither A nor B involves the regression parameter. They showed that $\hat{\vec{\gamma}}$ is

224   asymptotically normal with mean $\vec{\gamma}$ and with a variance-covariance matrix consistently

225   estimated by $\hat{V}(\hat{\vec{\gamma}})$. More precisely, $(A^{-1} B A^{-1})^{-1/2} (\hat{\vec{\gamma}} - \vec{\gamma})$ converges in distribution to

226    $N(0, 1)$.   The L-Y model has a limitation that the linear predictor $\vec{\gamma} Z_i(t)$ needs to

227   constrained to ensure positivity [13]. One may avoid this constraint by replacing $\vec{\gamma} Z_i(t)$

228   by $e^{\bar{\gamma}Z(t)}$, in which case $\lambda_0(t)$ pertains to the hazard function under $\bar{\gamma}\,Z_i(t) = -\infty$ rather

229   than under $Z_i(t) = 0$.

230

231   *3.3.3. Model Goodness of Fit*

232   The use of diagnostic procedures for model checking is an essential part of the modeling

233   process. While there are several residuals plots for testing the goodness of fit for the Cox

234   model [2], the residuals plot for the additive models is limited. Arjas' plot was used to

235   assess the adequacy of the fit of the additive model [17, 31-33]. The concept behind

236   Arja's plot is to plot expected number of failures against actual number of the injury event

237   with different covariate values. Arjas' plot is not a true residual plot, but deviations from

238   the 45° slope will give essentially the same information, which is a clearer indication of

239   lack of model fit.

240

241   SAS version 9.2 and R were used for the analysis in this study. The additive hazards

242   models are not available in commonly used computer packages, while for the Cox model

243   most statistical software are readily available and easy to use to fit models, check model

244   assumptions and assess model fit. Both the Aalen and L-Y additive hazards models can

245   perform by either a SAS macro available at

246   http://www.mcw.edu/FileLibrary/Groups/Biostatistics/Software/addmacro.txt [34] or a

247   combination of PROC PHREG and PROC REG [33].

248

249

250   **4. APPLICATION TO INJURY PREVENTION STUDY**

251   A total of 1,467 subjects (789 from the intervention group and 678 from the control

252   group) were eligible for the present study. Of these subjects, 263 subjects had the event

253   of the TLR related injury with 114 (14.4%) from the intervention group and 149 (22%)

254   from the control group. Our study observation was completed at December 1, 2006. The

255   Kaplan-Meier analysis was performed to assess the overall difference among the intervention

256   and control groups [35]. This result indicated that before 8 months the two survival curves

257   were very close. After 8 months, the intervention group had a higher probability of survival

258   as compared to the control group (p=0.0013 for log-rank test and p=0.0063 for Wilcoxon test;

259   Figure 1).

260

261   *4.1. Cox multiplicative model*

262   The result of Cox model showed that group, occupation, and body parts were significant

263   (Table 1). No significant interaction was observed between covariates. The intervention

264   group had a 27% lower risk of injury as compared to the control group after controlling

265   for occupation, and body parts(hazard ration (HR)=0.63; 95% CI: 0.497, 0.804;

266   p=0.0002). Nurses and nursing aides (NNA) had a 72% higher risk of injury compared to

267   Non-NNA (HR=1.72; 95% CI: 1.219, 2.416; p=0.002). The back, neck and shoulder

268   (BNS) were the most injured body parts. Compared to other body parts (Non-BNS), the

269   back, neck and shoulder (BNS) had a 115% increased risk of injury (HR=2.15; 95% CI:

270   1.618, 2.85; p<0.0001). Martingale residuals are used to check the overall fit of the

271   multiplicative hazards model for the intervention and control groups (Figures 2).

272   Martingale residuals showed that the fit of the multiplicative hazards model is

273   questionable.

274

275   *4.2. Aalen's additive model*

276   In order to visualize a covariate effect over time, the estimated cumulative regression

277   function has been examined, along with its upper and lower 95% point-wise confidence

278   limits. The plot of the estimated cumulative regression functions for group showed that

279   there was no covariate effect on the hazard up to 8 months. However, the slope was

280   negative and clear effects of decreasing hazard for the period of 8-24 months but after

281   that it was approximately constant hazard (Figure 3-a). Based on the estimated

282   cumulative regression functions, it has been concluded that intervention group had the

283   less risk of the injury event as compared to the control group. There may be time varying

284   occupation effect because the cumulative regression function shows the non-zero slope

285   over time (Figure 3-b). It has been observed that the effects of occupation have been

286   increased in hazard up to 10 months, disappearing afterwards. For body parts, Figure 3-c

287   also shows the positive slope over time and the 95% confidence limits of the covariate

288   effects did not includes zero.

289

10

290   *4.3. Lin and Ying's additive model*

291   The result of L-Y additive hazards model showed that group, occupation and body parts

292   were significant effect on the injury event (Table 2). The intervention group was

293   significantly different for the injury event comparing to the control group (p-

294   value=0.0005). The estimate is negative (-0.0025), indicating that the intervention group

295   had protection from injury as compared to the control group. This is interpretable as the

296   intervention group had 0.0025 less injuries than the control group after adjusting for

297   occupation and body parts. It means that 25 person injuries can be prevented per 10,000

298   persons by the injury prevention program. Regarding occupation, nurses and nursing

299   aides (NNA) had the significantly different on injuries than non-nurses occupations

300   (Non-NNA). NNA had 0.0024 excess risk of injuries (excess risk (ER) =0.0024; p-

301   value=0.0005; 95% C.I=0.001, 0.0038), which indicates that NNA had 24 more injury

302   compared to non-NNA per 10,000. Similarly, the body parts, combined back, neck &

303   shoulder had 0.0038 excess risk of injury than other body parts (ER=0.0038; p-value

304   <0.0001; 95% CI=0.0025, 0.0051). The Arjas plots were used for the selected covariates

305   to check the adequacy of the model. Figures 4-a shows that the plot are close to 45°,

306   indicating the group fits the model well. Notably, the Arjas plot of nurses and nursing

307   aides is not long enough, but it reasonably satisfies the model (Figure 4-b). However, for

308   the body parts, the plot is concave downwards and the deviations from the optimal fit was

309   shown (Figure 4-c).

310

311   **5. DISCUSSION**

312   We showed the differences in estimates of the coefficients from the Cox multiplicative

313   hazards model and the additive hazards models and their interpretation using an injury

314   prevention program implemented for the healthcare workers. The Cox multiplicative and

315   L-Y additive hazards models gave similar results with regard to covariates selected to be

316   significant: group, occupations, and body parts. The estimates from the models also had

317   the same signs, indicating the same directions of the covariate effects. Based on our

318   analysis, both the Cox and L-Y models, as well as Aalen's additive hazards model,

319   showed that the injury intervention program had a significant impact on reducing the

320   TLR related injures induced by patient handling among healthcare workers.

11

321

322   The parameter estimates and the standard errors from the Cox multiplicative and L-Y

323   additive hazards models are noticeably different. While the coefficients of the Cox model

324   act in a multiplicative way on unknown baseline hazards, those of the additive hazards

325   models act in an additive way on unknown baseline hazards. Because the coefficients act

326   in different ways in the multiplicative and additive hazards models, it is very difficult to

327   compare them directly. Moreover, the Cox model gives a higher estimate than additive

328   model when using a more compromised covariate profile probably due to the

329   multiplicative effect of fixed covariate on baseline function [11]. The association between

330   the covariates and the time to injuries in the additive hazards models was explained in

331   terms of the risk difference or excess risk rather than the risk ratio. Thus the different

332   models interpreted the coefficients in different ways.

333

334   The Cox model is most widely used; however, the proportional hazard assumption may

335   not always be satisfied in the data. In such cases, there are various solutions to consider,

336   for example, inclusion of a time-dependent covariate or stratification. In Aalen's model,

337   and the main focus was the cumulative regression plots, where the slope of the plots at

338   any given time provides information on the influence of the covariate at that moment.

339   From a practical standpoint, the graphical representation of the cumulative regression

340   functions is attractive, because it provides a direct perception of data and a picture of how

341   effects and the model fit in with change over time. Even one visualizes all covariate

342   effects over time, and a simple interpretation of the effects is not possible, which makes

343   Aalen's model less appealing in real applications than other models. However, it is still

344   useful particularly when we are interested in temporal effects. The unknown risk

345   coefficients used in Aalen's model are replaced by a constant covariate effect in the L-Y

346   model additive hazards model. A theoretical limitation of the L-Y model is that the linear

347   predictors in the model constrain to be positive [13]. Research on the additive hazard

348   model in relation to generalizing estimating function to the case of multivariate failure

349   time data as well as methods for checking the adequacy of the model is still rare. While

350   various statistical software packages are available for fitting the Cox model, the

351   procedure is limited to some software for the additive hazards model. Few macros are

352   available for the analysis of goodness of fit [17, 34].

353

354   Generally, the preference between the Cox hazards model and the additive hazards model

355   is normally a practical matter. Although in theory, either model can provide adequate fit

356   to a given time to event dataset, the more parsimonious one will unquestionably be

357   preferable to clinical investigators. One of the major advantages of using the additive

358   hazards model over the Cox multiplicative hazards model is that the resulting regression

359   parameter estimator has a closed form. In cases where both the additive and

360   multiplicative models fit the data fairly well, an additive specification may be preferred,

361   due to the easy interpretation of the regression parameters. Regression coefficients from

362   the additive model give more sensible and interpretable in public health research or

363   patient management/care, where the risk difference can be more important than the risk

364   ratio in understanding an association between a risk factor and disease occurrence [13,

365   17].

366

367   In summary, the Cox multiplicative and additive hazards models describe different

368   features of the association between the risk factors and the study outcomes. These

369   hazards models give different information and should not be viewed as alternative to each

370   other. Rather it seems desirable to use together to gain a more comprehensive

371   understanding of the data. Practitioners may benefit from these approaches, which help in

372   predicting the effect of one or more variables and in verifying their influence on the study

373   outcomes.

374

375

376   **COMPETING INTERESTS**
377   Authors have declared that no competing interests exist.

378

379

380

381

382   **REFERENCES**

383

384   1. Cox DR. Regression models and life tables (with discussion). Journal of the Royal

385   Statistical Society: Series B. 1972;34:187–220.
386

387   2. Kalbfleisch JD and Prentice RL. The statistical analysis of failure time data,
388   Wiley, New York; 2002.
389

390   3. Aalen OO. Further results on the non-parametric linear regression model in
391   survival analysis. Statistics in Medicine. 1993;12:1569–1588.
392

393   4. Andersen PK, Vaeth M. Simple Parametric and Nonparametric Models for Excess
394   and Relative Mortality. 1989;45:523-35.
395

396   5. Mau J. On a Graphical Method for the Detection of Time-dependent Effects of
397   Covariates in Survival Data. Applied Statistics-Journal of the Royal Statistical
398   Society Series C. 1986;35:245-55.
399

400   6. Mau J. A comparison of counting process models for complicated life histories.
401   Applied Stochastic Models and Data Analysis. 1988;4:283-98.
402

403   7. McKeague IW. Estimation for a Semimartingale Regression Model Using the
404   Method of Sieves. Annals of Statistics. 1986;14:579-89.
405

406   8. Huffer FW, McKeague IW. Survival analysis using additive risk models,
407   technical report. Department of Statistics, Stanford University;1987.
408

409   9. McKeague IW, Utikal KJ. Goodness-of-Fit Tests for Additive Hazards and
410   Proportional Hazards Models. Scandinavian Journal of Statistics. 1991;18:177-95.
411

412   10. Bhattacharyya M and Klein JP. A note on testing in Aalen's additive hazards
413   regression models. 2005;24:2235-40.
414

415   11. Lin DY and Ying ZL. Semiparametric Analysis of the Additive Risk Model.
416   Biometrika. 1994;81:61-71.
417

418   12. Lin DY and Ying ZL. Semiparametric analysis of general additive-multiplicative
419   hazard models for counting processes. The Annals of Statistics. 1995;23:1712-34.
420

421   13. Lin DY and Ying ZL. Additive regression models for survival data. Proceedings
422   of the First Seattle Symposium in Biostatistics: Survival Analysis, Springer, New
423   York. 1997;185–198.
424

425   14. Klein JP. Modelling competing risks in cancer studies. Statistics in Medicine.
426   2006;25:1015–1034.
427

428   15. Zhang X, Akcin H, Lim HJ. Regression Analysis of Competing Risks Data via
429   Semi-Parametric Additive Hazards Model. Statistical Methods and Applications.
430   2011;20:357-381.

16. Yin G and Cai J. Additive hazards model with multivariate failure time data. Biometrika. 2004;91:801–818.

17. Lim HJ and Zhang X. Semi-parametric additive risk models: Application to injury duration study. Accident Analysis and Prevention. 2009:41:211-6.

18. Lim HJ and Zhang X. Additive and multiplicative hazards modelling for recurrent event data analysis. BMC Medical Research Methodology. 2011;11:101 doi: 10.1186/1471-2288-11-101.

19. Martinussen T and Scheike TH. A flexible additive multiplicative hazard model. Biometrika. 2002;89:283-98.

20. Scheike TH and Zhang MJ. An additive-multiplicative Cox-Aalen model. Scandinavian Journal of Statistics. 2002;28:75–88.

21. Scheike TH and Zhang MJ. Extensions and applications of the Cox-Aalen survival model. Biometrics. 2003;59:1033–45.

22. Zahl PH. Regression analysis with multiplicative and time-varying additive regression coefficients with examples from breast and colon cancer. Statistics in Medicine. 2003;22:1113–27.

23. Cortese G and Scheike TH. Dynamic regression hazards models for relative survival. Statistics in Medicine. 2008;27:3563–84.

24. Cortese G, Scheike TH, Martinussen T. Flexible survival regression modeling. Statistical Methods in Medical Research. 2010;19::5–28.

25. Lim HJ, Black TM, Shah SM, Sarker S, Metcalfe J. Evaluating Repeated Patient Handling Injuries Following the Implementation of A Multi-factor Ergonomic Intervention Program among Health Care Workers. Journal of Safety Research. 2011;42: 185-191.

26. Black TM, Shah SM, Busch AJ, Metcalfe J, Lim HJ. Effect of Transfer, Lifting and Repositioning (TLR) injury prevention program on musculoskeletal injury among direct care workers. Journal of Occupational & Environmental Hygiene. 2011;8:226-235.

27. Cox DR and Oakes D. Analysis of Survival Data,. London: Chapman & Hall. 1984.

28. Aalen OO. A Linear-Regression Model for the Analysis of Life Times. Statistics in Medicine. 1989;8:907-25.

29. Klein JP and Moeschberger ML. Survival Analysis: Techniques for Censored and Truncated Data. Springer. 2003.

30. Aalen OO. Model for Nonparametric Regression Analysis of Counting Processes. Lecture Notes in Statistics. 1980;2:1-25.

31. Arjas E. A Graphical Method for Assessing Goodness of Fit in Cox Proportional Hazards Model. Journal of the American Statistical Association. 1988;83:204-12.

32. Torner A. Proportional Hazards and Additive Regression Analysis of Survival for Severe Breast Cancer. Technical Report. Stockholm: Stockholm University, Mathematical Statistics. 2004.

33. Schaubel DE and Wei G. Fitting the additive hazards model using standard statistical software. Biometrical Journal. 2007;49:719-730.

34. Howell AM. SAS Macro for the Additive Hazards Model: Medical College of Wisconsin, Biostatistics. 2007.

35. Kaplan EL and Meier P. Nonparametric Estimation From Incomplete Observations. Journal of the American Statistical Association. 1958;53:457-81.

517

518

519

520

521

522

523 Table 1. Estimation of coefficient, hazard ratio, 95% confidence interval, and p-value
524 from the Cox multiplicative hazards model.

525

Covariate	Estimate (S.E)	H.R	95% C.I	p-value
Group	-0.469 (0.128)	0.63	0.497, 0.804	0.0002
Occupation	0.540 (0.175)	1.72	1.219, 2.416	0.002
Body Parts	0.764 (0.144)	2.15-	1.618, 2.850	<0.0001

526 *S.E.: Standard Error; * HR: Hazard Ratio; * CI: Confidence Interval
527 Note: In this analysis, the reference group: Control group, , non-nurses for occupation (Non-NNA), and
528 other body parts except back, neck and shoulder for body parts (Non-BNS)
529
530
531
532
533
534

535 Table 2. Estimation of coefficient, excess risk, 95% confidence interval, and p-value from
536 the Lin and Ying's additive hazards model.

537

Covariate	Estimate (S.E)	E.R	95% C.I	p-value
Group	-0.0025 (0.0007)	-0.002	-0.0039, 0.0010	0.0005
Occupation	0.0024 (0.0006)	0.002	0.0010, 0.0038	0.0005
Body Parts	0.0038 (0.0006)	0.003	0.0025, 0.0051	<0.0001

538 S.E.: Standard Error; * ER: excess risk; * CI: Confidence Interval
539 Note: In this analysis, the reference group: Control group, , non-nurses for occupation (Non-NNA), and
540 other body parts except back, neck and shoulder for body parts (Non-BNS)
541
542
543
544
545
546
547
548

549
550
551
552
553
554
555
556
557
558 **Figure 1: Estimated survival probability curve by group.**
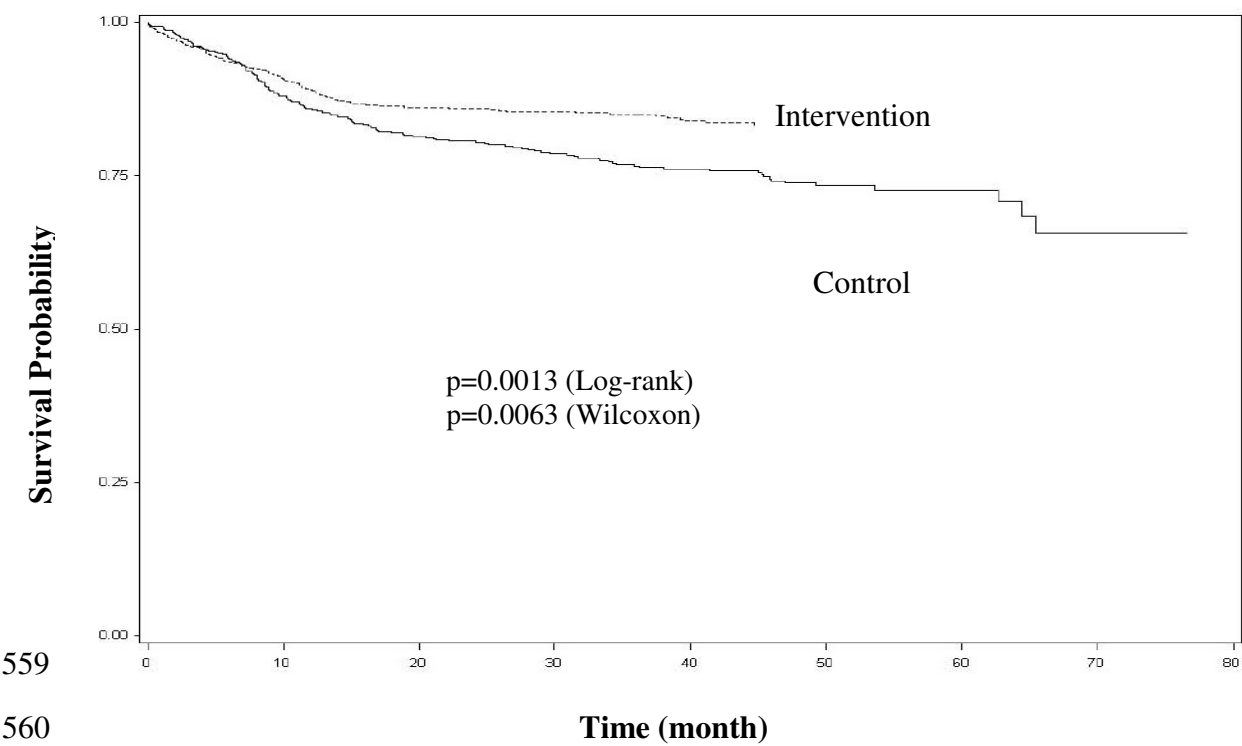
559
560 **Time (month)**
561
562
563

Figure 2: Martingale residuals plot for the multiplicative model.

Plot of rm*Time_Mth. Legend: A = 1 obs, B = 2 obs, etc.

565

566   **Time (month)**
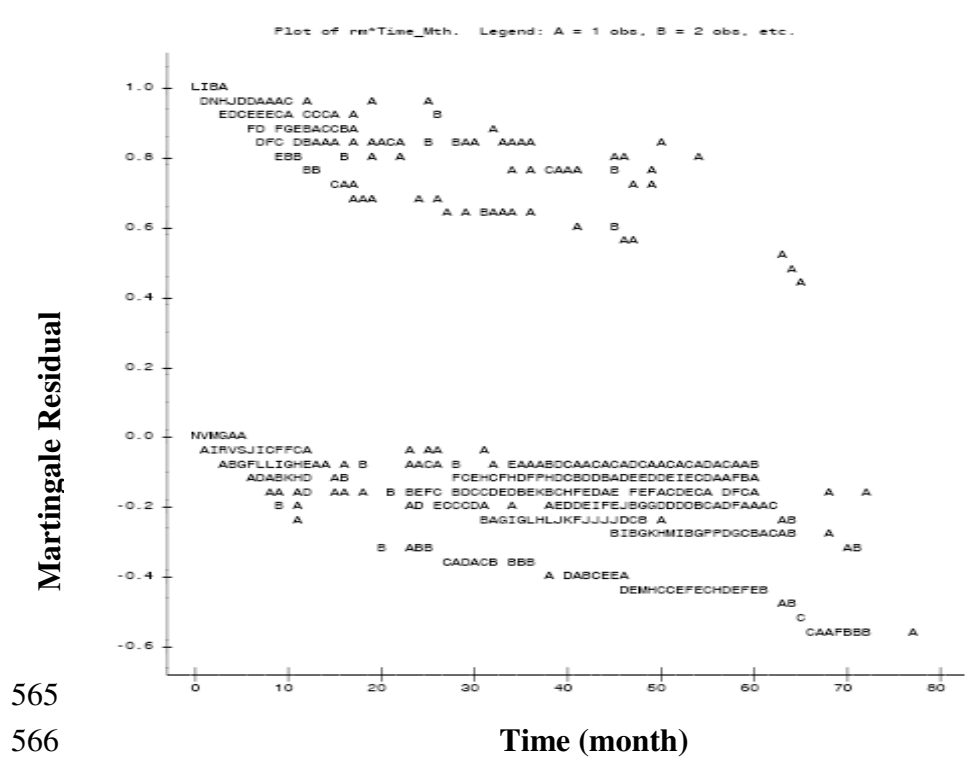
567   **Figure 3-a: Estimated cumulative regression function by Aalen's additive model**
568   **with its upper and lower 95% point-wise confidence limits for group.**

56

570   **Time (month)**

571
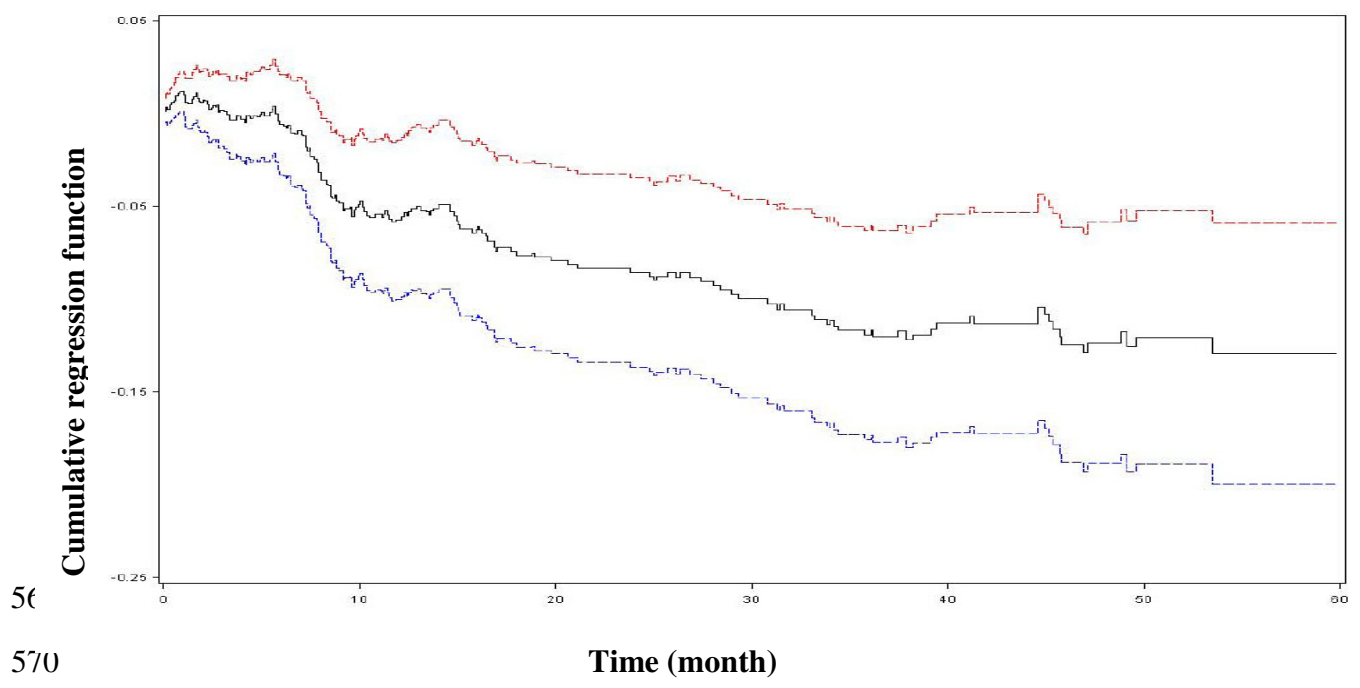
572   **Figure 3-b: Estimated cumulative regression function by Aalen's additive model**
573   **with its upper and lower 95% point-wise confidence limits for occupation.**
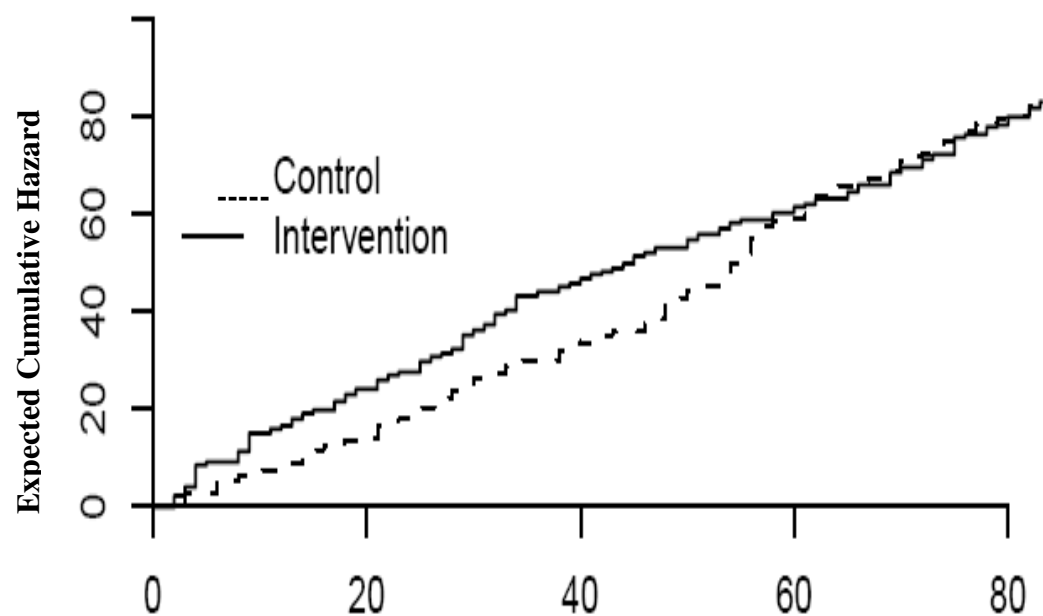
19
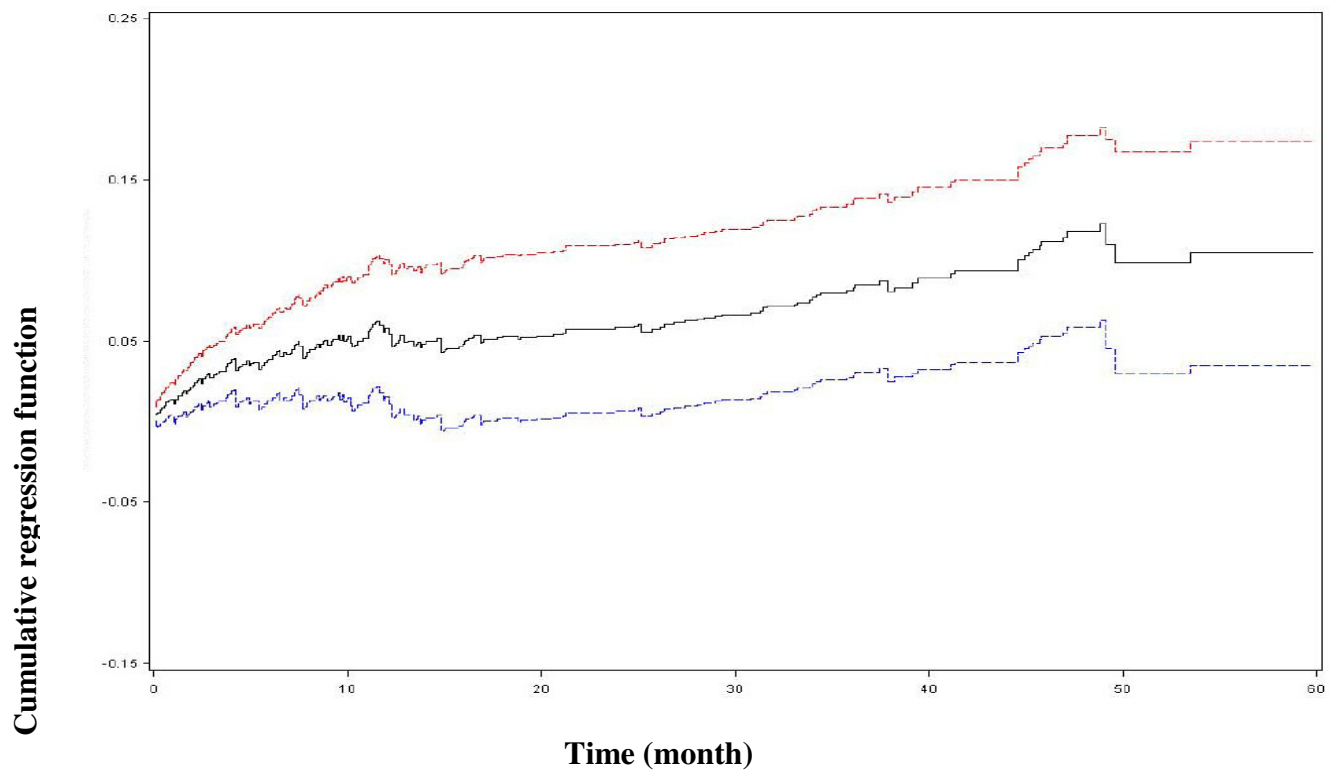
Figure 3-c: Estimated cumulative regression function by Aalen's additive model with its upper and lower 95% point-wise confidence limits for body parts.
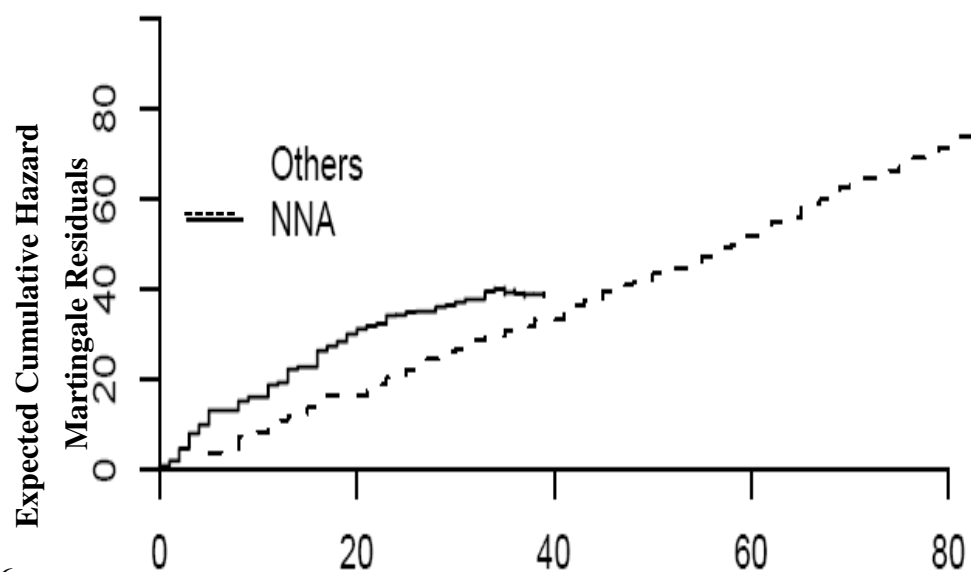

20
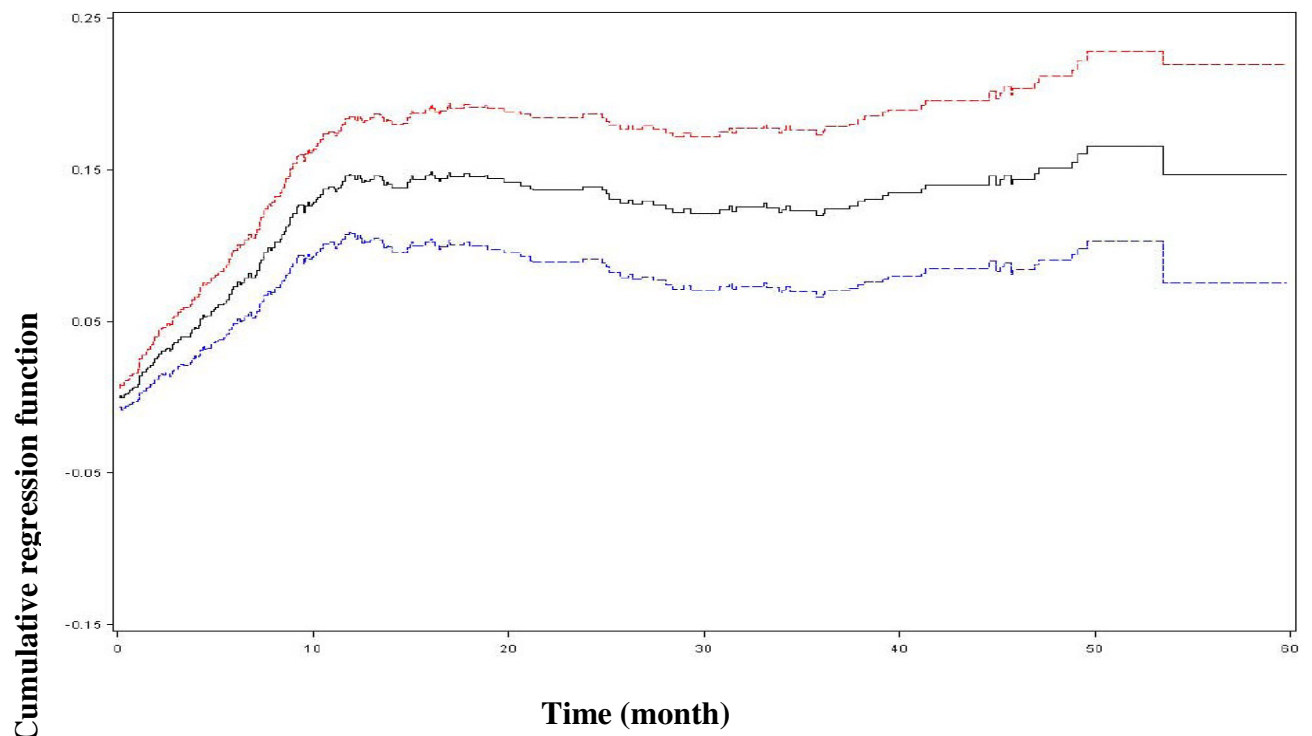
582 **Figure 4-a: Arjas plot of the estimated cumulative hazard by group.**

583

584 **Number of Injury**

585 **Figure 4-b: Arjas plot of the estimated cumulative hazard by occupation.**

586

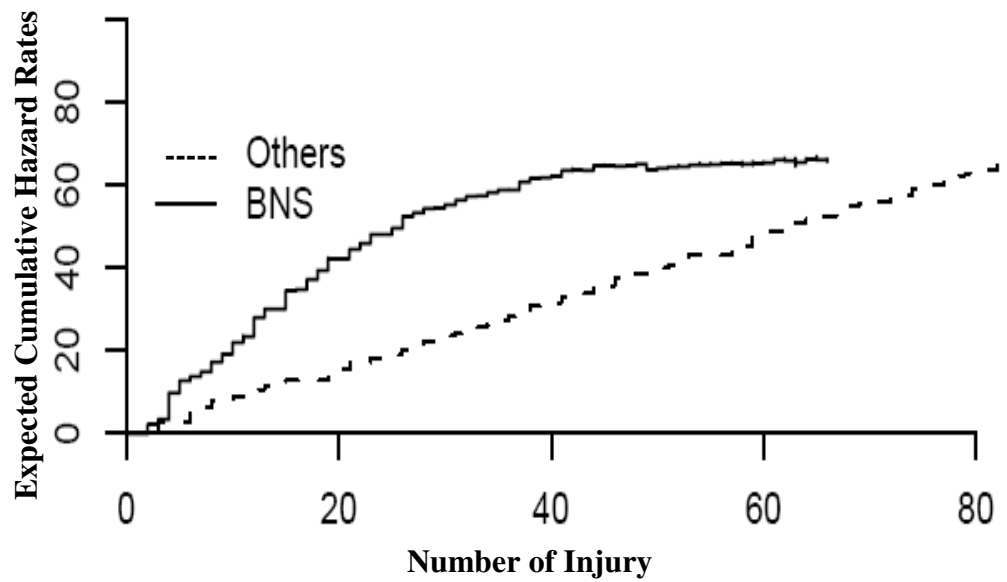587 **Number of Injury**

588 *NNA: Nurses and nursing aides

589 *Others include therapists, technicians, unit supporters, paramedics, etc.

590

591

592

593 **Figure 4-c: Arjas plot of the estimated cumulative hazard for the additive model by**
594 **body parts.**

595

596
597
598   * BNS: Back, neck and shoulder

599   * All other body parts include abdomen, chest, face, etc.