Research Article

Association between Obesity and Cancer: An Analysis Using the Competing Risk Regression Approach

Milan Bimali and Jianghua He

Department of Biostatistics, University of Kansas Medical Center, Kansas City, KS 66160, USA

Correspondence should be addressed to Milan Bimali; mbimali@kumc.edu

Received 4 October 2014; Revised 20 December 2014; Accepted 22 December 2014

Academic Editor: Peter N. Lee

Copyright © 2015 M. Bimali and J. He. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Cox model has been the commonly used method in past analyses of association between obesity and the risk estimates of cancer in situations where the subjects have also died (or could die) of noncancer events (competing events). The Cox model does not address the presence of competing events convincingly. The competing risk approach accommodates the fact that individuals who died of other causes (competing events) will never die of cancer and thus provides more realistic estimates. This study uses the competing risk approach to study the association of obesity and cancer mortality and compare the analysis results with those based on the traditional Cox model. It was seen that while the cause-specific hazard rate of cancer is significantly higher for obese population compared to normal weight population, the difference is not significant using competing risk approach. We demonstrated that higher cause-specific hazard rate does not necessarily imply higher incidence rate and in situations involving competing events we recommend using competing risk approach in addition to the Cox regression model.

1. Introduction

Over the past two decades obesity has established itself as a global health issue. Some of the staggering figures listed by the World Health Organization (WHO) such as “nearly doubled since 1980” and “fifth leading risk for global deaths” seem to indicate catastrophic consequences of obesity at global level (http://www.who.int/mediacentre/factsheets/fs311/en/). According to WHO, overweight and obesity are defined as the abnormal or excessive fat accumulation that may impair health. The association of obesity with cardiovascular disease, some types of cancer, diabetes, musculoskeletal disorders, and so forth has been reported time and again. Body Mass Index (BMI) defined as “weight in kilogram divided by square of height in meters” is a commonly used measure of obesity. Individuals with BMI exceeding 30 are categorized as obese. There have been numerous studies in the past that were designed to analyze the association between obesity and mortality and found significant results [1–5].

Besides the association of obesity and all-cause mortality, the association of obesity with specific cause of death such as cardiovascular diseases and cancer has also garnered attention [3–5]. Case-control and cohort study designs that are prospective as well as retrospective have been employed to comprehend the association between obesity and cancer [6–11]. Cox model has been commonly used for analyzing the association between obesity and all-cause mortality. When the traditional Cox model is used, deaths due to other causes were considered the same as usual censored cases, analogous to cases experiencing no event during the study period due to ending of the study, dropouts, or lost to follow-up. These usual censored cases are different from deaths due to other causes because these subjects may experience the event of interest after they were censored while individual who died of other causes such as heart diseases, AIDS, or car accidents at first can never die of cancer later. When the focus is analyzing the association with cause-specific mortality, any other event that could inhibit occurrence of the event of interest or affect its probability of occurrence is considered as a competing event. For example if the event of interest is death due to cancer, any other event such as death due to other disease that interferes with the occurrence of death due to cancer can be considered as a competing event and the associated risk as competing risk; and those subjects who have not experienced any event by the end of the study or were lost to follow-up can be considered as censored events. Lunn and McNeil have
described competing risk in very simple and comprehensive terms as "Competing risk occurs when there are at least two possible ways that a person can fail, but only one such failure type can actually occur" [12]. Such situations are commonly observed in medical research and have attracted interest over the past two decades [13]. In reality an individual who is followed over time either experiences the event of interest or experiences a competing event or is censored. The traditional Cox model does not differentiate between competing events and censored cases.

The association between obesity and cancer has been controversial. While in general most of the studies provide evidence in the favor of increased risk of cancer for obese patients as compared to normal weight patients [3, 4, 14], there are also studies that have reported low risk of cancer death associated with obese and overweight patients [15]. The studies in the past have employed different statistical techniques such as Cox regression model, logistic regression, and chi-square test to study the association between obesity and cancer. Past studies that analyzed the associations between obesity and cancer have not taken competing risk situation into consideration. The time-to-event analysis of the association between cancer and obesity in the presence of competing risks involved Cox regression models thereby treating the competing events as censored events. Such models provide a summary of how a covariate impacts the associated risk without taking into account the presence of competing risk [16]. Cox models assume that subjects can only die of event of interest and the censored events are independent of the events of interest. In studies where the assumption of independence of competing events and events of interest is justifiable, Cox regression models when employed, in such situations, without addressing the competing event differently from censored events, are likely to produce unrealistic risk estimates [13]. In the presence of competing risk, there has also been emphasis in computation and analysis of cumulative incidence of the event of interest (the total probability of observing event of interest) [17].

An alternative approach to the traditional Cox model is to take the competing events into consideration and treat them differently from the censored events. One of the popular competing risk models is the Fine and Gray model which was specifically developed to analyze the association between covariates and risk estimates in the presence of competing events [18]. Fine and Gray approach models the total probability of observing event of interest conditional on covariates and considers that subjects who experience competing events first will never have event of interest. Beyerersmann et al. have described Fine and Gray regression model as an interpretation-friendly alternative to Cox regression model [19]. An elegant explanation of Fine and Gray model has been provided by Putter et al. [20]. Pintilie et al. and Gunkemeier et al. introduce Fine and Gray model from clinician's perspective and demonstrate its application [21-23]. In the past Fine and Gray models have been used in different fields of medical science [24-26]. However to the best of our knowledge none of the studies in the past have employed Fine and Gray regression approach to analyze the association between obesity and cancer mortality. It would be interesting to find out whether the competing risk approach provides results different from those based the traditional Cox model in terms of the association between obesity and cancer mortality.

The purpose of this paper is to analyze the association between obesity and cancer in presence of competing events using Cox regression as well as the Fine and Gray model and compare the two results. The publicly available Framingham dataset will be used for the analysis. The event of interest is death due to cancer and the competing events are death due to cardiovascular disease as well as death due to other causes.

2. Methods

In this section we will compare the model settings for the Cox model and the Fine and Gray model. We assume that event 1 is the event of interest and event 2 is a competing event. We provide a brief definition of some of the concepts in survival analysis that will be used in the rest of the paper. The definitions are intended to be evocative rather than pedagogical.

Risk Set. It refers to the number of people who are at risk at any specified time t, that is, subjects who have not experienced the event of interest at time t but can possibly experience it in future.

Survival Function. It is the probability that a subject survives (does not experience event of interest) after time t.

Hazard Function. It is the probability that the subject experiences the event of interest immediately given the subject has survived (not experienced an event of interest) until time t.

Cause-Specific Hazard. It is the hazard function associated with a specific event when there are multiple events under consideration.

Cumulative Incidence Function. It is the probability that a subject actually experiences the event of interest by time t.

The above definitions can be found in the introductory textbook on survival analysis by Klein and Moeschberger [27].

2.1. Cox Regression Model. The standard Cox regression model, also known as the proportional hazard model, is a semiparametric model that essentially models the effect of covariates on the risk (cause-specific hazard) associated with event the of interest as follows:

$$\lambda_i(t) = \lambda_{0i}(t) \exp(\beta_i^T X),$$

where $\lambda_i(t)$ is the hazard function at time $t$ for event $i$ (e.g., $i = 1$ for event of interest; $i = 2$ for competing event). These hazards function are called cause-specific hazard functions.

$\lambda_{0i}(t)$ is baseline hazard for event $i$;

$X$ is the matrix of covariates;

$\beta_i$ is a vector of regression coefficients.
Any coefficient from \( \beta_i \) (say \( \beta_{i,j} \)) when raised to power of exponential, that is, \( \exp(\beta_{i,j}) \), can be interpreted as relative change in cause-specific hazard when the corresponding covariate is increased by 1 unit.

The use of Cox regression models in estimation of cumulative incidence proportion has been considered cumbersome and difficult to interpret [16, 28].

When the traditional Cox model is used to analyze competing events, each event is analyzed separately and the competing events are considered as censored. When competing events and censored events are independent of the event of interest, the estimated total risk \( (F_i(t)) \) for event \( i \) by a certain time point \( t \) is

\[
F_i(t) = \int_0^t \sum_{j \neq i} \left( \lambda_j(u) \times \exp \left\{ - \int_u^t \lambda_j(v) \, dv \right\} \right) \, du.
\]

This probability is the total probability of having event \( i \) by time \( t \) in a hypothetic world where the event \( i \) is the only possible way of failure. The estimated risk tends to be greater than the observable risk in a real world with competing events [20–22]. Using this approach, it is possible that the sum of the estimated probabilities of multiple competing events is greater than 1.

### 2.2. Fine and Gray Model

The Fine and Gray regression model is based on the idea of subdistribution hazard. The subdistribution hazard is defined as the probability of observing the event of interest in a subject under the assumption that the subject is alive at time \( t \) and either has experienced competing event or has not experienced any event at all. Table 1 illustrates the difference between hazard and subdistribution hazard. Lau et al. have given a nice pictorial representation of the difference between cause-specific and subdistribution hazard [16].

At time 0, we assume that 10 subjects are being followed where the two possible outcomes are either event of interest or competing event (censored events have been excluded for tabular convenience). At time 1, let us assume 1 subject experiences event of interest and two subjects experience competing event. The cause-specific and subdistribution hazards are both 0.1. However, at time 2, only the subjects that have not experienced any events \( (10 - 1 - 2 = 7) \) are included in risk set in the estimation of cause-specific hazard whereas in case of subdistribution hazard the subjects who have experienced competing event are also included \( (7 + 2 = 9) \) in the risk set. Thus if we were to observe an event of interest at time 2, the hazard estimates would differ (Table 1).

The logic behind inclusion of the subjects who died of competing risk is that these subjects can be considered as representative of the part of population that cannot have the event of interest [18].

Owing to the difference in risk sets, the risk estimates based on Fine and Gray model cannot exceed the risk estimates based on Cox model. Table 1 illustrates it.

Fine and Gray model essentially links the subdistribution hazard to the cumulative incidence function as follows:

\[
I_i^*(t) = 1 - \exp \left( - \exp \left( \beta_i^T \mathbf{X} \right) \times \int_0^t \lambda_{i,0}^*(s) \, ds \right),
\]

where \( \lambda_{i,0}(s) \) is the baseline subdistribution hazard function.

Past studies have shown association of age, smoking, and gender with cancer [29–35]. These covariates were thus included as the main effects in the model. In the interest of parsimony, stepwise regression technique was employed in selecting interaction effects to be included in the model. Thus while age, smoking, and gender were always included, only the interactions that were identified as significant (\( P \) values below 0.05) were included in the model. The analysis was done performed using the statistical software—SAS and R. The "crr" package together with the "base" package in R was used. The analysis algorithm is summarized in Figure 1. Scrucca et al. provide step-by-step guidance of conducting competing risk modeling in R [36]. Readers, on request, may obtain the dataset as well as the R and SAS codes from the authors.

### 3. Dataset

A dataset of “Framingham Heart Study” is used for analysis. The Framingham Heart Study is a long term prospective study of the etiology of cardiovascular disease among a population of free living subjects in the community of Framingham, Massachusetts. The original cohort of the study consisted of 5209 men and women from Framingham, Massachusetts. Extensive examinations were carried out every other year on this cohort since 1948. The dataset consists of follow-up information of 4526 individuals (2005 men and 2521 women) who were alive at Exam 4 of the study and all the variables were measured at Exam 4.

### 4. Results

#### 4.1. Variables

Table 2 summarizes the variables in the Framingham data set.
Table 2: Summary of variables in the Framingham dataset.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Percentage</th>
<th>Median follow-up time (yrs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>44.3</td>
<td>26.76</td>
</tr>
<tr>
<td>Female</td>
<td>55.7</td>
<td>32.92</td>
</tr>
<tr>
<td>Current smokers</td>
<td>52</td>
<td>29.53</td>
</tr>
<tr>
<td>Current Nonsmokers</td>
<td>48</td>
<td>30.87</td>
</tr>
<tr>
<td>Underweight (BMI &lt; 18.5)</td>
<td>1.3</td>
<td>31.82</td>
</tr>
<tr>
<td>Normal weight (18.5 ≤ BMI &lt; 25)</td>
<td>43.6</td>
<td>33.36</td>
</tr>
<tr>
<td>Overweight (25 ≤ BMI &lt; 30)</td>
<td>41.7</td>
<td>29.11</td>
</tr>
<tr>
<td>Obese (BMI ≥ 30)</td>
<td>13.4</td>
<td>25.59</td>
</tr>
<tr>
<td>Cancer death</td>
<td>15.5</td>
<td>20.56</td>
</tr>
<tr>
<td>CVD death</td>
<td>26.1</td>
<td>19.72</td>
</tr>
<tr>
<td>Other cause of death</td>
<td>19.5</td>
<td>26.17</td>
</tr>
<tr>
<td>Censored</td>
<td>38.9</td>
<td>35.59</td>
</tr>
</tbody>
</table>

Summary statistics for age (years)

<table>
<thead>
<tr>
<th>Min</th>
<th>Median</th>
<th>Max</th>
<th>Mean</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>34</td>
<td>49</td>
<td>69</td>
<td>49.82</td>
<td>8.49</td>
</tr>
</tbody>
</table>

\[ n = 4526 \text{ (after removing 15 obs. with missing data).} \]

Table 3: Summary of analysis for cancer as event of interest.

<table>
<thead>
<tr>
<th>Cause-specific hazard modelling (Cox model) with cancer as event of interest</th>
<th>Hazard</th>
<th>Std. error</th>
<th>( P ) value</th>
<th>95% conf. int.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Underweight [0.705 \pm 0.413]</td>
<td>0.397</td>
<td>(0.314, 1.583)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overweight [1.051 \pm 0.084]</td>
<td>0.552</td>
<td>(0.891, 1.24)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Obese [1.336 \pm 0.12]</td>
<td>0.016*</td>
<td>(1.056, 1.69)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age [1.051 \pm 0.0007]</td>
<td>6.11 \times 10^{-13}*</td>
<td>(1.037, 1.065)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Smoking [1.574 \pm 0.084]</td>
<td>5.7e \times 10^{-8}*</td>
<td>(1.336, 1.855)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender [0.364 \pm 0.484]</td>
<td>0.037*</td>
<td>(0.141, 0.934)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age \times gender [1.028 \pm 0.0094]</td>
<td>0.003*</td>
<td>(1.009, 1.047)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Fine and Gray modelling with cancer as event of interest</th>
<th>Hazard</th>
<th>Std. error</th>
<th>( P ) value</th>
<th>95% conf. int.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Underweight [0.641 \pm 0.419]</td>
<td>0.29</td>
<td>(0.282, 1.46)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overweight [1.043 \pm 0.085]</td>
<td>0.62</td>
<td>(0.884, 1.23)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Obese [1.087 \pm 0.120]</td>
<td>0.49</td>
<td>(0.858, 1.38)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age [1.017 \pm 0.006]</td>
<td>0.0073</td>
<td>(1.005, 1.03)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Smoking [1.406 \pm 0.081]</td>
<td>2.7 \times 10^{-6}*</td>
<td>(1.199, 1.65)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender [0.589 \pm 0.441]</td>
<td>0.23</td>
<td>(0.248, 1.40)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age \times gender [1.014 \pm 0.009]</td>
<td>0.10</td>
<td>(0.997, 1.03)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Statistically significant.

4.2. Cancer as Event of Interest. Deaths due to cancer are treated as events of interest and deaths due to other causes are the competing events. Figure 2 presents the nonparametric estimates of cumulative incidence function based on cause-specific (Cox model) and subdistribution (Fine and Gray model) hazard approach for each BMI category. It is evident that the estimates of risk based on Cox model are in general higher than the risk estimated by Fine and Gray (subdistribution hazard) model. The overestimation in case of Cox model is due to the identical treatment of censored events and other competing events. Thus the separation of curves between normal and obese group based on Cox model is more pronounced than that of Fine and Gray model.

The three BMI categories (obese, underweight, and overweight) were included in the model and based on the results from stepwise regression “age,” “smoking,” “gender,” and “age \times gender” were included the model. The results based on the Cox regression model and the Fine and Gray model are summarized in Table 3.

After adjusting for the effect due to age, smoking status, gender, and the interaction between age and gender, the results from Cox regression show that the cause-specific hazard rate of cancer for “obese” group is approximately 4 times higher than “normal” group (Table 3). This difference is found to be statistically significant (\( P \)-val 0.016). The Fine and Gray analysis shows that the incidence rate of cancer for “obese” group is roughly 3 times higher than “normal” group (Table 3). However this difference is not statistically significant (\( P \)-val 0.49). Such inconsistent results were also observed in case of the “gender” and in interaction between “age” and “gender.” For the remaining covariates the two models do not show conflicting \( P \) values; that is, the \( P \) values are above or below the conventional \( \alpha \)-level of 0.05 for both models.

5. Discussion

Our work essentially compared the association between obesity and risk of cancer using the Cox model for cause-specific analysis and the Fine and Gray model of the competing
risk approach. Our findings show that while the cause-specific hazard rate estimated with Cox model for cancer in obese population is significantly higher than that of normal population, the incidence rate of cancer estimated with the Fine and Gray model in obese population is not significantly higher than that of normal population. The analysis thus demonstrates that significantly higher cause-specific hazard rate in a hypothetical world without competing events does not necessarily imply a significantly higher incidence rate in a real world with competing events. These findings are in agreement with the previous studies [37, 38]. We carried out an additional analysis with deaths due to cardiovascular diseases as the event of interest. In this case the results from Cox model also showed higher mortality risk than Fine and Gray model but the two results were in agreement as the effects of obesity on mortality due to cardiovascular diseases were significant using both approaches.

Which approach should be used for research? From implementation aspect, Cox model is easy to use and is flexible in accommodating time-dependent covariates and had been widely used; the Fine and Gray model is not as flexible as the Cox model. For example, it does not address time-dependent covariates and it also does not allow stratified model [39]. Its extension to interval censored, as well as truncated data, is a work in progress thereby limiting its application primarily to right-censored data. For a given data, Fine and Gray and Cox regression models are not comparable by any likelihood-based model selection techniques because these two models use different data values as in the Fine and Gray model the event times due to competing events are replaced with the largest observed event time in the data [18]. Criteria such as research goals and the relationship among competing events can be used in comparing and/or selecting the two models. When there is only one event of interest and censored events and events of interest are independent, the Cox regression model is very useful. Due to identifiability dilemma, that is, the independence assumption being not testable [27], often times researchers have to make the decisions about independence based on clinical considerations and prior experience. When there are competing events but the occurrence of competing event can be assumed independent of events of interest, the Cox model provides reasonably close estimates. When the independence assumption is obviously violated and the effect of competing risk is not negligible, the competing risk approach is recommended. Even in situations where independence of competing events and events of interest is a safe assumption, Fine and Gray modeling can still be considered as a supplementary analysis as it provides additional information. If both models indicate significant association between covariates and the risk estimates, it suggests a real effect; if the two models provide different results, the researchers should examine the research goals and the relationships among the events more carefully to determine which model is more relevant.
Conflict of Interests
The authors declare no conflict of interests regarding the publication of this paper.

References


