Parametric Cure Model versus Proportional Hazards Model in Survival Analysis of Breast Cancer and Other Malignancies

Shunzo Maetani¹*, John W. Gamel²
¹Tenri Institute of Medical Research, Tenri, Japan
²Department of Ophthalmology, University of Louisvill, Louisvill, USA
Email: ¹maetani@tennyorozu.jp

Received June 7, 2013; revised July 8, 2013; accepted July 15, 2013

Copyright © 2013 Shunzo Maetani, John W. Gamel. 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.

ABSTRACT

As cancer therapy has progressed dramatically, its goal has shifted toward cure of the disease (curative therapy) rather than prolongation of time to death (life-prolonging therapy). Consequently, the proportion of cured patients (c) has become an important measure of the long-term survival benefit derived from therapy. In 1949, Boag addressed this issue by developing the parametric log-normal cure model, which provides estimates of c and m where m is the mean of log times to death from cancer among uncured patients. Unfortunately, traditional methods based on the proportional hazards model like the Cox regression and log-rank tests cannot provide an estimate of either c or m. Rather, these methods estimate only the differences in hazard between two or more groups. In order to evaluate the long-term validity and usefulness of the parametric cure model compared with the proportional hazards model, we reappraised randomized controlled trials and simulation studies of breast cancer and other malignancies. The results reveal that: 1) the traditional methods fail to distinguish between curative and life-prolonging therapies; 2) in certain clinical settings, these methods may favor life-prolonging treatment over curative treatment, giving clinicians a false estimate of the best regimen; 3) although the Boag model is less sensitive to differences in failure time when follow-up is limited, it gains power as more failures occur. In conclusion, unless the disease is always fatal, the primary measure of survival benefit should be c rather than m or hazard ratio. Thus, the Boag lognormal cure model provides more accurate and more useful insight into the long-term benefit of cancer treatment than the traditional alternatives.

Keywords: Cancer Survival Analysis; Boag Model; Cox Model; Cure

1. Introduction

In recent decades, as more cancer victims have enjoyed long-term, relapse-free survival, cure has become a reality for both patients and clinicians. Thus, the primary goal of cancer therapy has shifted toward cure of the disease rather than prolongation of time to death. To achieve a cure, selecting the best regimen is vital. This is especially true with children, for whom curative treatment can yield many years of healthy life, while prolongation of life offers only a limited benefit before relapse takes the child’s life. Furthermore, cured patients are saved from cancer-associated sufferings, which could be more unbearable to patients than death itself. Hence, the proportion of cured patients (cure rate) has become an important measure of long-term survival benefit.

As early as 1949, Boag [1] addressed this issue by developing a parametric cure model that allowed him to estimate the cure rate (c) and the mean of log failure times (m) among uncured patients (failure time means time to death from cancer under study). Twenty-three years later, Cox [2] published the proportional hazards model, which uses the hazard ratio to measure the survival difference between groups. Since then non-parametric or semi-parametric methods based on his model (e.g., log-rank test and Cox regression) have remained the mainstay of cancer survival analysis. Hereafter these methods will be referred to as “standard survival analysis”. The purpose of our paper is to compare the usefulness of the Boag and Cox models, whose primary parameters are the cure rate and the hazard ratio, respec-
tively, and also to confirm the validity of the Boag model using the data from breast cancer and other neoplastic diseases.

2. Survival Models and Results of Their Applications

2.1. Proportional Hazards Model

In 1972, Cox [2] developed a unique model, paying special attention to the hazard, which is the probability that survivors at the beginning of a unit interval will die during that interval. He assumed that at any point in time the hazard in one group is proportional to the hazard in a second group at the same time point. It is important to note, however, that the absolute hazard may vary over time while the hazard ratio (HR) between the two groups remains constant. This single constant allows us to compare survival between groups. For example, if the average HR between the two groups is 0.5, it is presumed that the mortality in one group is one half of the mortality in the other group.

2.2. Relationship between Cure Rate and Hazard Ratio

It is generally assumed that, provided the proportional hazards model holds, the proportion of patients cured (i.e., saved from the event) can be estimated by 1-HR [3-5], which in turn is calculated from standard survival analysis. For example, consider a trial that compared adjuvant chemotherapy with or without trastuzumab and found a HR of 0.67 in HER2-positive breast cancer patients [6]. This result can be summarized either as “trastuzumab therapy is associated with a 33% (= 1 – 0.67) reduction in the risk of death”, or as “trastuzumab therapy prevents 33% of the deaths that would occur without the therapy”. Such statements are commonly found in leading medical journals [3,4,6,7]. Hence, given this information, most patients would believe their chance of being cured could be increased with the treatment if it shows a HR less than 1.

However, according to Peto et al. [8] and Clark et al. [9], 1-HR is not the proportion of patients whose deaths (or relapse) are prevented by treatment, but the proportion of deaths that are either prevented or delayed. Thus, as will be seen, the HR cannot distinguish between treatments that prevent death (curative treatment) and those that merely delay it (life-prolonging or death-delaying treatment).

Moreover, the relationship between the cure rate and HR became even more questionable when we read the Cox original paper [2], where he used the data from the Acute Leukemia Group B [10] as an example. Between 1959 and 1960, this group had conducted a randomized controlled trial to estimate the effect of 6-mercaptopurine (6-MP) versus placebo on steroid-induced remission of patients with acute leukemia (most were children with acute lymphoblastic leukemia).

When we re-analyzed the data using the Cox model, the HR of 6-MP versus placebo was 0.22 (95% CI: 0.10 to 0.49) [11]. If the relationship between the cure rate and HR were valid, this finding would indicate the following: 78% (1- HR = 100% - 22%) of the relapse that would have occurred in the placebo group were prevented by 6-MP. On the contrary, further follow-up of 6-MP-treated children revealed that almost all died from relapse [12,13].

Although the Cox model is still commonly used in cancer survival analysis, Cox [14] himself acknowledged the limitations of his model. More specifically, he stated that the model is unlikely to achieve the two objectives: 1) long-term survival study decoupled from shorter-term effects; and 2) provision of patient-specific prognostic information for clinician and an individual patient.

2.3. Parametric Cure Model

The Boag log normal model incorporates the cure rate as one of its parameters [1]. He assumed that a fraction (c) of the patients with a specific cancer are cured of the disease by treatment, while the remaining incurable patients will eventually die of the disease (Figure 1) unless they succumb to other causes. He further assumed that log failure times among uncured patients follow a normal distribution with mean (m) and standard deviation (s). In the era when computers were not available, Boag manually estimated these three parameters for various tumors such as cancers of the breast, uterus, lung, head and neck, etc.

In 1977, Farewell [15] modeled the parameter c as a dependent variable in a logistic regression, while in 1994 Gamel and McLean [16] expressed all three parameters of the Boag model as multivariate regressions on various covariates. This allows clinicians to determine the effects of treatment and other prognostic factors on both c and m—information that cannot be obtained with the hazard ratio in the Cox model.

The multivariate model was then extended to allow the analysis of grouped data where information on the cause of death for individual patients may not be available [17,18] (http://survillance.cancer.gov/cansurv/).

When the multivariate Boag model was applied to the 6 MP data, we found that the chemotherapy failed to cure the disease (Wald P = 0.99), but rather prolonged time to relapse 3.8 times longer (95% CI: 2.07 to 7.15) than in the placebo group [11]. However, since 1960 the proportion of clinically cured children with acute lymphoblastic leukemia has steadily increased to 80% due to progress in anticancer regimen (Figure 2) [19].

These results show that at least two types of anticancer treatments are available: a curative one that increases the
fraction of patients cured of the disease and a life-prolonging one that merely delays tumor-related death. To illustrate the differential effects of these two regimens on survival and hazard curves, these curves were simulated by increasing one of the two parameters (either \( c \) or \( m \)) of the Boag model (Figure 3).

The red curves show the base-line values of the control group (Group 1), while the blue curves represent the effects of therapies (Group 2). The middle and right panels show the effects of curative and life-prolonging regimens, respectively.

It is readily seen that increase in parameter \( m \) alone
Figure 3. Survival and hazard curves derived from the Boag model and Cox model. Red: control (Group 1); blue: test treatment (Group 2); c: cure rate; m and s: mean and standard deviation of log failure times in month. To meet the Cox proportional hazard assumption, non-mixture model was used (left side panels). Thus, the survival rates were estimated as $c \times F(t)$, where $c$ represents the cure rate; $F(t)$ the cumulative lognormal distribution, and the hazard rate is estimated as $\ln(c) \times F(t)$. From 2.5 to 3.0 results in crossing of the hazard curves (right lower panel), so the proportional hazards assumption does not hold. The left panels show the survival and hazard curves satisfying the proportional hazards assumption. Note that these curves are very similar to the corresponding ones derived from the Boag model (middle panels) in which the parameter $c$ alone is increased from 0.3 to 0.4 while $m$ is kept unchanged.

2.4. Cox Model versus Boag Model in Randomized Controlled Trials of Cancer Therapy

To further illustrate the difference between these models, two examples of randomized controlled trials will be shown in which both models are applied.

2.4.1. Effect of Adjuvant Chemotherapy for Stage 2 Breast Cancer

Using the three parametric versions of the Boag cure model (lognormal, log logistic and Weibul) plus the log-rank statistic, Gamel et al. [20] re-analyzed the data from five randomized controlled trials. These trials, published from 1981 to 1992 by Bonadona et al. evaluated adjuvant chemotherapy for stage 2 breast cancer [21-25]. The chemotherapeutic regimens included standard treatment (i.e., mastectomy alone) or mastectomy plus intravenous CMF (cyclophosphamide, methotrexate and fluorouracil) with or without doxorubicin.

The results showed that in three of the five trials there were statistically significant survival differences between the treatment and control groups. However, a curative effect was found only in the trial with doxorubicin plus CMF, whereas in the other two positive trials (with CMF regimens), the treatment merely prolonged the time to relapse. The stepwise log likelihood ratio test and chi-square statistics showed that the lognormal distribution provided the better fit to the pooled data than the log-logistic or Weibul versions of the Boag cure model.

2.4.2. Effect of D2 versus D1 Lymphadenectomy in Gastric Resection for Cancer

From 1989 to 1993, the Dutch Gastric Cancer Group [26] conducted a randomized controlled trial to compare the effects of limited lymphadenectomy (D1) versus extended lymphadenectomy (D2) in 711 patients undergoing potentially curative gastrectomy for gastric adenocarcinoma. After a median follow-up of 6 years the hazard ratio for relapse between D2 and D1 groups was 0.84 (95% CI: 0.65 to 1.09). These findings suggest no significant difference, whereas the postoperative mortality and morbidity were significantly higher with D2 dissection.

Such negative results did not agree with the clinical experience of Japanese surgeons, so we applied the Boag model to the same data. Our findings showed a significant difference in cure rate (11.5%; 95% CI: 3.1 to 20.0).
between the two groups [11]. Later the Dutch group, by extending their follow-up to a median of 15 years, found that D2 lymphadenectomy is associated with a significantly lower disease-related death rate (37% versus 48%) with a hazard ratio of 0.74 (95% CI:0.59 to 0.93) [27]. This difference in the death rate (11%) is close to the difference in cure rate (11.5%) found with the Boag model approximately 10 years earlier.

3. Discussion

Boag is the first scientist who attached special importance to cure in the analysis of cancer-related survival. The multivariate extension of his model estimates the impact of treatment and other prognostic variables on the likelihood of cure, thus providing both patients and clinicians the information they need to make vital decisions [28]. Although the hazard ratio in the Cox model is closely related to the cure rate in the Boag model [29], HR is less comprehensible to non-statisticians than c, which more accurately depicts the long-term survival benefit of a given treatment.

In contrast, the parameter m of the Boag model plays a subordinate role unless all patients being studied are incurable. If only time to death is studied to assess the effect of treatment, this could be likened to counting only the coins in a cash transaction while leaving the bills uncounted.

Clinical trials and simulation studies have shown that standard survival analysis suffer a number of failings; 1) they cannot distinguish between curative and life-prolonging treatments [21,30,31]; 2) they are more sensitive to an increase in failure time than to an increase in cure rate, especially when the follow-up is limited; 3) as a result, they may favor a death-delays treatment over one that is curative; 4) they tend to loose power with increasing follow-up [30,32,33].

These limitations are graphically shown in Figure 3. In the right lower panel, the two hazard curves representing the control (red) and the test treatment (blue), respectively, separate at the beginning, but then come closer and cross. If follow-up is truncated earlier than this crossing (for example, at 2 years), the hazard curve of the life-prolonging treatment remains lower than that of the control throughout the study. Thus the HR of treatment versus control would be overstated, while violation of the proportional hazards assumption may be missed. For these reasons, standard analysis might mislead clinicians and patients into selecting a less effective regimen. Though pharmaceutical industries may benefit from the consumption of such drugs, the welfare of patients and their families will suffer; if they were correctly informed, they might have refused to have a regimen that merely delays death. When Heyland et al. [34] interviewed 278 elderly patients who were at high risk of dying in the next 6 months. Only 12% preferred life-prolonging care.

On the other hand, the parametric cure models also suffer limitations. They are less sensitive to difference in failure time during the early period [31], although they gain power with longer follow-up. Furthermore, under certain conditions, they may provide a poor fit to the observed data. To avoid this problem, some statisticians recommend alternative parametric models.

However, the cause of the poor fit may not be use of the wrong model (misspecification) but misclassification of events. For example, patients who actually died from therapeutic complications in the early postoperative period may be misclassified as dead from cancer (failure). It must be kept in mind, however, that early failure is very rare, since most clinical trials in the adjuvant setting require that participants are in remission or have undergone potentially curative surgery (i.e., these are the eligibility criteria for candidates to be enrolled in the trial). So it is unlikely that failure occurs shortly after its cause has been eliminated.

Even if the trial is conducted in the non-adjuvant setting, imminently fatal cases should have been excluded from most trials. Consequently, the actual hazard curve should begin at zero, rise to a peak and then gradually decline to zero (Figure 1). Such a unimodal hazard curve is seen only in the lognormal or log-logistic model, but not in the Weibul or gamma model. If death occurs very early, the cause of death should be checked carefully. If the cause is ambiguous, such cases should be classified as censored at the time of death. Otherwise, the maximum likelihood estimation may fail to converge, or the result may be biased toward a shorter estimate of m.

Another criticism against the parametric cure models is that they rely on “extrapolation of a survival curve outside the available data” [5,35]. It is important to note, however, that predicting events beyond observed data has long served many branches of sciences. An excellent example is meteorology, where predicting the course of hurricanes has enjoyed great success. In addition, few models have been reported to provide a better fit to observed survival data than the lognormal or log-logistic models [1,20,35-39]. Nevertheless, we must continue our effort to find a better model using large, accurate, life-long follow-up data sets.

4. Conclusion

In conclusion, unless the disease is always fatal, the primary measure of survival benefit should be the proportion of patients cured rather than hazard ratio or median
time to failure. Thus, the Boag lognormal cure model provides more accurate and more useful insight into the long-term benefit of cancer treatment than the standard non-parametric alternatives.

REFERENCES


