Validation of the Lognormal Model for Prediction of Long-term Survival Rates from Short-term Follow up Data in Stages III and IV Breast Cancer

P. Tai
E. Yu, University of Western Ontario
D. Skarsgard, University of Saskatchewan
J. Tonita

Available at: https://works.bepress.com/edwardyu/56/
Validation of the lognormal model for prediction of long-term survival rates from short-term follow-up data in stages III and IV breast cancer: a 22-year follow-up study

P. Tai MD FRCP*, E. Yu MD PhD,†
D. Skarsgard MD FRCP*, and J. Tonita MSc*

1. INTRODUCTION

The purpose of the present study is to evaluate a statistical model that estimates long-term survival with short-term follow-up data in breast cancer. The lognormal distribution is defined as the distribution of a random variable whose logarithm is normally distributed. Lognormal distribution has been shown to represent the distribution of a number of variables such as the induction times of tumours in mice, growth rates of breast cancers that recurred in the scar, growth rates of breast cancers followed in serial mammograms, red blood cell volumes, and the survival times of cancer patients who die with their disease present (Table 1). Using the database of the Surveillance, Epidemiology, and End Results (SEER) program, 42 cancer sites were shown to follow the lognormal distribution.

The study team investigated the applicability of the lognormal model for the analysis of breast cancer data. Boag’s lognormal prediction modelling for long-term cancer survival rates has, in theory, been available for use for some 50 years. However, because of the lack of long-term quality follow-up data and adequate computer power in the past, the model has been validated by only a few authors. The lognormal model has been validated and found to correctly predict long-term survival from short-term follow-up data for cervical, small cell lung, larynx, prostate, breast, bladder, tongue, and thyroid cancers.

This parametric statistical model may be used prospectively for clinical trials to predict long-term survival rates several years earlier than would be possible by using the standard actuarial life table (Kaplan-Meier) methods of calculation. The nonparametric product-limit method of Kaplan-Meier is currently widely used in biomedical data analysis to deal with problems of incomplete follow-up and censoring.

The validation of lognormal model has two phases. Phase 1 tests the goodness of fit to a lognormal distribution of the survival time of those cancer patients who died with their disease present. This uncured group is represented by the fraction \((1 - C)\) of the total patients, where \(C\) is the cured fraction of all patients.

Phase 2 attempts to verify the lognormal model by using short-term follow-up data (for example 2–6 years, depending on the cancer subgroup being studied and its prognosis) to predict long-term survival rates. The latter rates are then compared with values calculated by the Kaplan-Meier life-table method from long-term available data. The second phase has been difficult to apply because of the general lack of large numbers of patients with sufficiently long follow-up information. As a result, considerably

<table>
<thead>
<tr>
<th>Cancer sites</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Head and neck cancer</td>
<td>Berg 1965, Mould et al. 1976</td>
</tr>
<tr>
<td>Nasal sinus cancer</td>
<td>Berg 1965</td>
</tr>
<tr>
<td>Mouth and throat cancer</td>
<td>Boag 1949</td>
</tr>
<tr>
<td>Mouth</td>
<td>Berg 1965</td>
</tr>
<tr>
<td>Non small cell lung cancer</td>
<td>Berg 1965</td>
</tr>
<tr>
<td>Small cell lung cancer</td>
<td>Tai et al. 8, Tai et al. 6</td>
</tr>
<tr>
<td>Intraocular melanoma</td>
<td>Gamel et al. 1993</td>
</tr>
<tr>
<td>Bone sarcoma</td>
<td>Berg 1965</td>
</tr>
<tr>
<td>Cancer of uterine cervix</td>
<td>Mould and Boag 1975, Berg 1965</td>
</tr>
<tr>
<td>Ovarian cancer</td>
<td>Berg 1965</td>
</tr>
<tr>
<td>Hypernephroma</td>
<td>Berg 1965</td>
</tr>
<tr>
<td>Bladder cancer</td>
<td>Berg 1965</td>
</tr>
<tr>
<td>Gastric cancer</td>
<td>Berg 1965, Maetani et al. 1980</td>
</tr>
<tr>
<td>Lymphoma</td>
<td>Berg 1965</td>
</tr>
<tr>
<td>Chronic leukaemia</td>
<td>Tivey 1954</td>
</tr>
<tr>
<td>Brain tumours</td>
<td>Berg 1965</td>
</tr>
<tr>
<td>For 42 SEER cancer sites</td>
<td>Tai et al. 2002</td>
</tr>
</tbody>
</table>

* Phase 2 validation also performed.

SEER = Surveillance, Epidemiology, and End Results Program [public-use CD-ROM (1973–1999); Bethesda, MD: National Cancer Institute; Apr 2002].
more phase 1 studies have been conducted, with the authors then assuming that the model is valid without extending their work to phase 2—for example, Berg 5 and Haybittle 14.

2. PATIENTS AND METHODS

From 1981 to 1995, 5,894 cases of breast cancer were registered with the Saskatchewan Cancer Registry. The patients were treated by surgery or by adjuvant hormonal therapy, chemotherapy, or radiotherapy in various combinations. In the present study, each new patient is counted as a single case, even if the patient had bilateral breast cancer.

As first proposed by Boag 7, the patients were assigned to one of four groups according to disease and vital status at the time of follow-up. Group 1 patients had died from breast cancer; group 2 patients had died from intercurrent disease; group 3 patients were alive and free of clinical evidence of disease; and group 4 patients were alive with persistent or recurrent disease. The methodology was described in detail by Boag 7 and has been used by various authors over the years 10–20.

2.1 Test of Goodness of Fit for Lognormality: Phase 1

To ensure the availability of adequate long-term survival data, we first collected the breast cancer data from the Saskatchewan Cancer Registry. Lognormality was tested for patients at various stages of breast cancer who died with the disease present (as distinct from patients who died of an intercurrent disease). A minimum chi-square method was used to estimate the mean \( M \) and standard deviation \( S \) of the log(10) survival time distribution of breast cancer patients who died with the disease present.

The minimum chi-square test program was run using Microsoft Excel 2000 (Microsoft Corporation, Redmond, WA, U.S.A.). A range of \( M \) values and \( S \) values were tested to reduce the chi-square to a minimum. For the minimum chi-square testing of observed data, the null hypothesis states that there is no difference between the distribution of the observed survival time and that of the lognormal distribution. The null hypothesis is rejected if \( p < 0.05 \).

Once lognormality was demonstrated and minimum chi-square estimates of \( M \) and \( S \) were obtained for patients with various stages of breast cancer, a second computer program—using a maximum likelihood method described by Boag 7—was run for the second phase of the study to estimate long-term survival rates and the cured fraction \( C \).

2.2 Validation of the Lognormal Model: Phase 2

Once it had been concluded in phase 1 that the survival time of the unsuccessfully treated cases could be represented by a lognormal distribution with mean \( M \) and standard deviation \( S \) of the log(surival time) values, then long-term survival rates were predicted by a maximum likelihood method (see Appendix A).

The present study included all breast cancer patients diagnosed from the beginning of 1981 to the end of 1985, with follow-up to February 2003, making the series ideal for validation purposes.

A cohort is defined as any designated group of people who are followed or traced over a period of time. For stage i breast cancer, after the 5-year period of diagnosis, patients were followed to December 31, 1990—that is, 5 years after 1985. The range of follow-up duration would thus be 5–10 years before censoring for calculation by the lognormal model.

The predicted cancer-specific survival rate (cssr) was then validated by the Kaplan–Meier method, using the actual follow-up data available up to February 2003. (For stage iv breast cancer, the patients were followed for 3 years after the 5-year period of diagnosis.)

3. RESULTS

The survival time of all uncured patients with stages i and ii breast cancer did not follow lognormality in the minimum chi-square test in phase i: the \( p < 0.05 \) results indicated a significant difference in the survival distribution of the early-stage cancer patients from the lognormal distribution. Subgroups of stage i patients (age < 50 years at diagnosis) and stage ii patients (age < 60 years) followed lognormality with \( p > 0.05 \) in the minimum chi-square tests. Stages iii and iv followed the lognormal distribution in the minimum chi-square tests. Table ii lists the mean \( M \) and standard deviation \( S \) values of the log(10)(survival time) obtained by the minimum chi-square method.

Given the long survival time of stage i and ii patients, the duration of follow-up in our data was not adequate for validating long-term survival up to a stable cure status.

For stage iii, a 5-year period of diagnosis with 140 patients was used, and the patients were followed as a cohort for an additional 5 years. The predicted 20-year cssr was 10%, and the cure rate (\( C \times 100\% \)) was 8% by the lognormal model. The Kaplan–Meier calculation, which used actual follow-up data to February 2003, yielded a 20-year cssr of 10% ± 3%. Eight patients were still alive at the time this paper was written.

For stage iv, a 5-year period of diagnosis with 109 patients were used, and the patients were followed as a cohort for an additional 3 years. The predicted 15-year cssr was 1%, and the cure rate (\( C \times 100\% \)) was 0% by the lognormal model. The 15-year cssr calculated by the Kaplan–Meier method, which used actual follow-up data up to February 2003 was 2% ± 1%. All patients died of their cancer or intercurrent diseases in 18 years.
SURVIVAL RATE PREDICTION IN BREAST CANCER

<table>
<thead>
<tr>
<th>Stage</th>
<th>Diagnosis period</th>
<th>Patients (n)</th>
<th>M values</th>
<th>S values</th>
<th>p Value a</th>
</tr>
</thead>
<tbody>
<tr>
<td>i (age&lt;50)</td>
<td>1981–1995</td>
<td>66</td>
<td>1.89</td>
<td>0.31</td>
<td>0.54</td>
</tr>
<tr>
<td>ii (age&lt;50)</td>
<td>1981–1995</td>
<td>448</td>
<td>1.77</td>
<td>0.37</td>
<td>0.29</td>
</tr>
<tr>
<td>iii</td>
<td>1981–1995</td>
<td>312</td>
<td>1.55</td>
<td>0.43</td>
<td>0.54</td>
</tr>
<tr>
<td>iv</td>
<td>1981–1985</td>
<td>115</td>
<td>1.59</td>
<td>0.51</td>
<td>0.25</td>
</tr>
<tr>
<td>iv</td>
<td>1981–1995</td>
<td>341</td>
<td>1.27</td>
<td>0.49</td>
<td>0.08</td>
</tr>
<tr>
<td>iv</td>
<td>1981–1985</td>
<td>103</td>
<td>1.32</td>
<td>0.48</td>
<td>0.79</td>
</tr>
</tbody>
</table>

a The null hypothesis states that there is no significant difference between the observed and expected data estimated from the lognormal distribution. The hypothesis is rejected if \( p < 0.05 \).

4. DISCUSSION

The lognormal distribution has a long history and has been shown to give a good approximation for time to death associated with many cancers. In studies by Gamel et al., Mould and Boag, and Rutqvist (alone and with co-workers), the lognormal model was found to be the most optimum model as compared with other parametric models.

4.1 Practical Implications

In the present study, the lognormal model was validated in the prediction of long-term survival rates for patients with advanced-stage breast cancer. In the literature, many cancer sites have been shown to follow a lognormal distribution. The validation procedure requires actual long-term and good-quality follow-up data; however, few cancer centres have those data. The Saskatchewan Cancer Registry satisfies those requirements, giving us the ability to investigate the utility of the model for additional cancer sites.

The practical value is that the lognormal model could be used to predict the results of prospective trials earlier than the Kaplan–Meier method can. Once accrual is complete, use of the lognormal model can considerably shorten the subsequent follow-up years. As a result, cancer treatment could potentially make more rapid advances. A reduction in the cost of cancer research may be another potential benefit.

5. ACKNOWLEDGMENTS

The present study was supported by Saskatchewan Cancer Agency Research Grant Award 2785. Results were also part of a podium presentation at the annual scientific meeting of the Clinical Oncological Society of Australia (COSA); Sydney, Australia; November 28–30, 2002.

6. REFERENCES


Correspondence to: Patricia Tai, Allan Blair Cancer Centre, 4101 Dewdney Ave, Regina, Saskatchewan S4T 7T1.
E-mail: ptai@scf.sk.ca
* Saskatchewan Cancer Agency, Regina, Saskatchewan.
† London Regional Cancer Centre, London, Ontario.

APPENDIX A

Validation of the lognormal model, phase 2

Boag’s maximum likelihood method was used to estimate the long-term survival rates. Using the parametric lognormal model, the standard deviation $\sigma$ was fixed, and only the two remaining parameters, $M$ and $C$, were kept floating. Multiple iterations converged to a stable solution for $C$.

The cancer-specific survival rate (CSSR) at time $t$ equals

$$\left[ C + (1-C) \times Q \right] \times 100\%,$$

where $Q$ is the integral of the lognormal distribution between the limits $t$ and infinity; $t$ is long-term survival time; $C$ was estimated by the maximum likelihood method in phase 2.