ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Applied Biostatistics - MATH-493

Assignment 2: A Statistical Critique on a Published Paper

Authors:

Nils Kalbfuss, Jules Mertens,
Yan Wong Wen,

Professor:

Darlene Goldstein

May 23, 2018

Contents

1 Introduction

This report is a statistical critique on a paper of Che K. et Al. titled *Survival Outcomes for Patients with Surgical and Non-Surgical Treatments in Stages I-III Small-Cell Lung Cancer*, published in the 'Journal of Cancer' in 2018 [1].

1.1 Biomedical Background

Lung-cancer is the leading cause of cancer-related death. Small-cell lung cancer (SCLC) is a specific subgroup of lung cancer, characterized by an aggressive clinical course and early metastasis. Currently, patients are mostly treated by chemotherapy and radiotherapy. The treatment by surgical resection has been discontinued after studies in the 70s questioned the effectivity of surgical resection. This claim is questioned now, because recent studies showed that surgical resection can improve survival in patients with SCLC especially in early stages. Furthermore, recent guidelines recommend surgery followed by chemotherapy for N0-1 stage and surgery followed by postoperative radiotherapy (PORT) for N1 and N2 stages.

1.2 Objectives

In order to verify the claims of other research groups and further investigate the recommendations, the authors pursued 3 objectives:

1. Define the role of surgery in patients with SCLC
2. Evaluate the outcome of PORT versus surgery treatment alone
3. Explore the outcome of surgical resection in stage III SCLC patients

2 Data

2.1 Data selection

For their study, data were selected from the Surveillance, Epidemiology and End Results (SEER) database. They retrieved data from patients who were diagnosed with stage I-III SCLC from 2004 to 2014. Since patients were treated against their disease and data on them collected and analysed by the authors, the study was observational and retrospective. In general, patients were included into the study depending on the site and morphology (lung and bronchus) of the tumour, the classification of the tumor in ICD-03, the non-reception of chemotherapy and the collection of data on the tumor's stage. Patients were excluded if surgery or survival months were unknown or if patients had N3 or Nx disease. Based on these criteria, 4780 individuals were included in the study. Variables analysed in this study were age, gender, race, primary site, surgery, radiation sequence with surgery, TNM (tumor nodes metastasis) stage, grade, number of nodes examined, number of positive nodes and year of diagnosis.

2.2 Study design

In this retrospective study, all patients were diagnosed with SCLC. Interestingly, the vast majority (89%) did not undergo surgery. Furthermore only 1.9 % were treated with PORT. Most importantly, patient data were compared between untreated patients and those that underwent surgery or PORT. Since patients were only included in the study, if they did not receive chemotherapy, it was predicted that the change of survival rates would only be due to the treatment. Therefore, untreated patients were the control group.
The goal of the study is to shed light on the effect of surgical resection and PORT in treatment of SCLC. For this, the authors contrast older studies, showing no effect of surgical resection, with latest studies showing a beneficial effect. Therefore, the author's have a clear hypothesis and try to confirm or deny it. Therefore, the study was rather confirmatory. However, the analyses on PORT were rather exploratory. They had very little numbers for PORT patients and therefore difficulties to draw conclusions. The focus of the study design was therefore more on the surgical resection of SCLC patients.

3 Analyses

The research group has done three major analyses on their data: a univariate Cox regression followed by a Kaplan-Meier survival analysis and finally Multivariate Cox regression.

These techniques are state-of-the-art to perform a survival analysis. There are two main reasons why these are more appropriate here than for example a linear regression.

- The outcome of the stastical analyses is to investigate the influence of surgery on the survival time (OS and LCSS) expressed in months. Since the survival times (or the observations) are only positive numbers, they should be transformed to make them compatible to perform linear regression, which is not necessary with the methods used in the paper to model the survival time.
- Secondly, the techniques used in the paper are able to deal with missing information in the data due to censoring, typically right-censoring. The research group selected patients diagnosed with stage I to III SCLC between 2004 and 2014, a part of them will not have died during the follow-up duration. This censored information is robustly handled by Cox regression and Kaplan-Meier survival which is not possible with for example linear regression models. The mechanism of censoring is assumed to be random and non-informative for the used modeling techniques to be valid.

3.1 Univariate Cox Regression

A Univariate Cox regression (also called Cox proportional hazards model) is performed to identify relevant variables for the overall and lung-cancer-specific survival rates (OS and LCSS). This regression was done seperately for the groups with a different stage of SCLC (I to III). The results of these Cox proportional hazards models, namely hazard ratios and p-values, are shown in tables 2 up to 4 of the paper.

Cox regression is valid under the assumption that the hazard ratio is constant over time if the predictors don't change, also called the proportional hazards assumption. Under this assumption, this technique is very appropriate to investigate the influence of variables of interest (predictors) on the survival time, expressed as hazard ratios.

In the context of the study, 'univariate' is used for Cox proportional hazards models with only one predictor and one dependent variable: the survival time. The hazard ratio shown in table 2 for the predictor "surgery" represents the risk of dying when diagnosed for SCLC in stage I comparing surgery to no surgery without taking into account any other predictor (and their possible interactions). This limits the clinical importance of the outcomes of these 'univariate' models, since confounding factors might influence the results. Excluded predictors might lead to wrong conclusions. To take into account more predictors at the same time, a 'multivariate' Cox regression is done further in the paper.

3.2 Kaplan-Meier Survival

Kaplan-Meier survival analysis is performed for the visualization of survival probabilities as a function of time (expressed in months). For the comparison of the Kaplan-Meier curves, log-rank tests were performed to check for significant differences between survival-rates.

In figure 1 of the paper, Kaplan-Meier curves are plotted which show the OS- and LCSS-probabilities for surgery and no surgery, once for all stages together (graphs A-B) and then for the three stages seperately (graphs C-H). A log-rank test is performed to compare the curves of surgery and no surgery. P-values lower than 0.001 were found for all stages meaning that surgery gives significantly (under $\alpha = 0.001$) higher survival chances for all stages. The Kaplan-Meier curves give a clear and easily interpretable visualisation of these findings.

In figure 2 of the paper, similar graphs show OS- and LCSS-probabilities comparing different surgery types but now only for all stages together. Again significant differences where found using the log-rank (under $\alpha = 0.001$) in favor of lobectomy. In our opinion, this analysis (and its visualization using Kaplan-Meier curves) is not appropriate to answer the question which type of surgery is favorable since this choice greatly depends on the stage of the cancer. For example, lobectomy seems for both the OS- as the LCSS-rates to be the preferable type of surgery but maybe this is just because lobectomy is mostly done on Stage I SCLC-subjects (which is maybe not true in reality but just chosen as illustration) which have higher survival-chances in general. Here, the stage of the diagnosis can be seen as a confounding factor. A solution could be to plot the Kaplan-Meier curves of the different types of surgery for every stage separately.

Figure 3 of the paper shows the Kaplan-Meier curves of OS- and LCSS-rates for surgery alone versus surgery in combination with post-operative radiotherapy (PORT) for all stages together. The comments on figure 2 apply here as well.

3.3 Multivariate Cox Regression

Finally, a 'multivariate' Cox regression is done to reduce the effect of possible confounding factors by including multiple independent variables such as age, gender and race into the model.

The results of this regression are presented in table 5 of the paper which shows the OS- and

LCSS-rate for the different predictors as hazard ratios (and their 95% Confidence Intervals) with respect to a chosen baseline.

Under the assumption of hazard ratios to be constant over time, the used method is very appropriate to evaluate the influence of a predictor taking into account a set of other predictors and their possible interactions. This is a control of the presence of confounding factors in the 'univariate' Cox regression discussed in section 3.1. Since no significant differences were found in the results, the conclusions drawn out of the 'univariate' analysis become more credible.

The word 'multivariate' is used to emphasize the fact that more independent variables are included in one and the same model to predict the survival-time. Since only one dependent variable (OS- or LCSS-time) is investigated in this study, it would be more accurate to call this a univariate multiple Cox-regression.

4 Discussion

4.1 Methods

As discussed in section 3, Cox regression and Kaplan-Meier analysis (in combination with log-rank tests) were performed to analyze the data. These are appropriate techniques for this retrospective survival-analysis with censoring.
In our opinion, as elaborated in section 3, the Kaplan-Meier Analysis and their visualization could be improved to increase the relevance of the drawn conclusion, especially for the analysis on the type of surgery.

4.2 Results

The credibility of the results of a statistical analyses is limited by the presence of confounding factors and sources of bias.

As mentioned before, the stage of diagnosis can be a confounding factor for the OS- and LCSS-rates for different surgery types. This can lead to wrong conclusions about the best surgery type for a SCLC-patient.

Another possible confounding factor might be the fact that the staging for patients that were treated without surgery are staged in a different way (clinically) than patients who underwent surgery as part of their treatment (pathologically). If clinical staging would hypothetically underestimate the stage of SCLC-patients in a systematic way, this can for example result in a so-called confounding bias favoring surgery.

No confounding factors are explicitly controlled for as far as disclosed in the paper. However, the 'multivariate' Cox regression can be seen as a method to reduce the influence of confounding factors by including multiple covariates into the model. In this way, it controls for confounding factors of the 'univariate' Cox regression for covariates as age, gender and race.

As stated in the paper, an inherent selection bias is inevitable for this retrospective study. Because of the requirements the subjects needed to meet, a total random selection of subjects is impossible resulting in a possible bias because of this selection procedure.

4.3 Power

One of the main aims of the study was to evaluate whether surgical treatment was beneficial for SCLC patients in stages I to III. This is done by comparing the Kaplan-Meier Survival curves between surgically treated patients and non-surgically treated patients (comparison done separately for each stage). Based on the information presented in the paper (Figures 1.a and 1.b), we can make use of the ***ssizeCT.default()*** function from the ***PowerSurvEpi*** package in R to estimate the minimum number of subjects required for both experimental and control groups to attain a specified minimum power. This implementation is based on a proposition by Freedman (1982) [2].

Stages	No Surgery (%)	Surgery (%)
I	66.2	33.8
II	79.7	20.3
III	96.7	3.3

(a) Proportion of subjects

Stages	OS - Hazard ratio of surgery	LCSS - Hazard ratio of surgery
I	0.369	0.335
II	0.549	0.506
III	0.477	0.456

(b) Hazard Ratios

Figure 1: Information regarding SCLC patients in stages I to III

We decided that for the conclusions to be convincing, the power for each Cox-Proportional Hazard Model should at least have 95% power (conventional values are 80%, 90%, 95%). For both OS and LCSS, we calculated the minimum number of patients required in each Cox Regression Model (Figure 2).

Stages	OS		LCSS	
	Surgery	No Surgery	Surgery	No Surgery
I	34(344)	66(674)	34(344)	66(674)
II	69(60)	268(235)	60(60)	233(235)
III	48(116)	1405(3351)	46(116)	1326(3351)

Figure 2: Estimated minimum sample size required to achieve 95% power (numbers in braces are actual number of subjects used in study)

From these estimations, it is obvious that the Cox-Proportional Hazard Models used for OS and LCSS in Stage I have the highest power (number of subjects used exceeds minimum requirement by more 10 times), and thus conclusions derived from these two models should be regarded as most convincing. The number of subjects used in the Hazard Model for OS and LCSS in Stage III still exceeds the minimum requirement by more than 2 times. So, the conclusions for stage III are credible as well. The conclusions drawn from the Cox regression for OS and LCSS in Stage II are the most questionable from the three: the model for OS did not reach the minimum sample size , and the model for LCSS barely exceeds the requirement on the sample size to have a power of at least 95%.

It is difficult for us to perform the same estimation of minimum sample size requirement for analyses regarding effectiveness of PORT, and comparison of different types of surgical treatments because specific values of cumulative survival probabilities were not mentioned in the paper (eyeballing from the Kaplan-Meier survival curves provided in the paper is an option, but our estimations would be inaccurate, thus we decided not to proceed).

Stages	OS (%)	LCSS (%)
I	100	100
II	92.3	95.3
III	99.9	99.9

Figure 3: Estimated power of comparison of survival curves between surgically and non-surgically treated patients

The paper did not mention any information related to the power of analyses carried out in the study. To estimate the power of comparison between survival curves, we can utilize the ***power.default()*** function from the ***PowerSurvEpi*** package in R, similar to the approach used to estimate minimum sample size in the previous section. The implementation of this function is based on the same proposition by Freedman (1982) [2]. Figure 3 shows the estimations of power using the ***power.default()*** function. As mentioned in the previous section, Cox Regression models for stage I and Stage III have high power (close to 100%) and thus conclusions derived from these 4 models are highly credible. Cox regression models for stage II have lower power, but are still acceptable.

Based on these estimations, one should be able to confidently claim that surgical treatments are in general beneficial for SCLC patients, especially for patients in Stages I and III.

5 Conclusion

In summary, the authors conclude that patients with stage I-III SCLS who underwent lung resection had significantly better survival than patients treated without resection. This does not only include patients of stage I and II but also patients of stage III. They further conclude that patients undergoing lobectomy have the best outcome. In contrast, patients who underwent PORT did not have a higher overall survival and lung cancer-specific survival.

In general, the design was satisfactory for the analysis of survival rates in patients undergoing surgery:

- The statistical methods used are state-of-the-art and appropriate
- Analyses were done on patients who did not undergo chemotherapy. This decreased the number of other factors that might be influential on the survival rate. The study was therefore well designed in order to determine the effects of surgery alone on the survival rate.
- They were able to retrieve data from a large number of patients.

However, there were also some limitations of the study:

- The analysis of the effect of PORT was done on a very low number of PORT patients and compared to an overwhelmingly large number of patients who did not undergo PORT treatment. Furthermore, in the analysis of the effect of PORT patients were of different SCLC stages were pooled. This could lead to a Simpson paradox.

- Apparently, the used dataset also included some limitations: chemotherapy information was incomplete and not explicit, for some patients it was not known whether chemotherapy treatment was actually done or not and some data were not sufficiently described (for example that PORT represented patients who underwent surgery including surgery to other regional and distant sites.)

- In addition, the authors neglect the cellular and molecular diversity of SCLC and do not go into greater molecular depth. They also fail to include or describe the decision of a doctor when to perform surgery. This is especially true for stage III patients when there are already metastases. In fact, a bias might occur because doctors might only perform a surgery of stage III patients if their disease is not far progressed.

- Sometimes SCLC stages were summarized to compare survival rates of different surgery types or to compre PORT treatment. This might have lead to the Simpson paradox with another stage as a confounding factor.

References

[1] Keying Che, Hongchang Shen, Xiao Qu, Zhaofei Pang, Yuanzhu Jiang, Shaorui Liu, Xudong Yang, and Jiajun Du. Survival Outcomes for Patients with Surgical and Non-Surgical Treatments in Stages I-III Small-Cell Lung Cancer. *Journal of Cancer*, 9(8):1421–1429, 2018.

[2] L. S. Freedman. Tables of the number of patients required in clinical trials using the logrank test. *Statistics in Medicine*, 1(2):121–129.