



## Post-Chemotherapy Survival Range Analysis of Lung Cancer Patients Through Statistical Software 'R'

### İstatistiksel Yazılım 'R' Aracılığıyla Akciğer Kanseri Hastalarının Kemoterapi Sonrası Sağkalım Aralığı Analizi

  Dhiraj Kumar Singh<sup>1</sup>,  Sunil Kumar<sup>2</sup>

<sup>1</sup> Department of Mathematics, Zakir Husain Delhi College, University of Delhi, India

<sup>2</sup> Department of Zoology, Zakir Husain Delhi College, University of Delhi, India

ORCID ID: Dhiraj Kumar Singh <https://orcid.org/0000-0002-5513-9558>, Sunil Kumar <https://orcid.org/0000-0002-2518-3443>

\*Sorumlu Yazar / Corresponding Author: Dhiraj Kumar Singh, e-posta / e-mail: [dk Singh@zh.du.ac.in](mailto:dk Singh@zh.du.ac.in)

Geliş Tarihi / Received : 12-07-2021

Kabul Tarihi / Accepted: 13-08-2021

Yayın Tarihi / Online Published: 30-08-2021

Singh D.K., Kumar S. Post-chemotherapy survival range analysis of lung cancer patients through statistical software 'R', J Biotechnol and Strategic Health Res. 2021;5(2):143-147

#### Abstract

Positive inclination of health sector towards patient data collection, management, multi-variant statistical analysis along with mathematical modelling has opened a new avenue to better understand the post-treatment health complications and survival range. It helps the researchers to predict and design better strategies to contain and cure the disease, and save the lives. Current study is an attempt to re-analyse the secondary data about the lungs cancer patients underwent chemotherapy. Statistical analysis using 'R' revealed that females showed better response towards treatment than males as assessed by the post-treatment survival range monitoring. Moreover, meal supply did not significantly affected the post-treatment body weight in both the cases.

**Keywords** Cancer, Chemotherapy, Statistical analysis using 'R'

#### Özet

Sağlık sektörünün hasta verilerinin toplanması, yönetimi, çok değişkenli istatistiksel analizin yanı sıra matematiksel modellemeye yönelik olumlu eğilimi, tedavi sonrası sağlık komplikasyonlarını ve sağkalım aralığını daha iyi anlamak için yeni bir yol açmıştır. Araştırmacıların hastalığı kontrol altına almak ve iyileştirmek ve hayat kurtarmak için daha iyi stratejiler öngörmelerine ve tasarımlarına yardımcı olur. Mevcut çalışma, kemoterapi uygulanan akciğer kanseri hastalarına ilişkin ikincil verileri yeniden analiz etme girişimidir. 'R' kullanılarak yapılan istatistiksel analiz, tedavi sonrası sağkalım aralığı izlemesiyle değerlendirildiği üzere, kadınların tedaviye erkeklerden daha iyi yanıt verdiğini ortaya koydu. Ayrıca, yemek arzı her iki durumda da tedavi sonrası vücut ağırlığını önemli ölçüde etkilemedi.

**Anahtar Kelimeler** Kanser, Kemoterapi, 'R' kullanılarak istatistiksel analiz

## INTRODUCTION

Despite six decades of active research in the field of biological sciences, our knowledge about human physiological system is very limited and requires further inter-disciplinary investigations to develop better understandings. In recent years, mathematical models has emerged as important tools that helps better understand the intricacies of various biological phenomenon such as population fluctuations, non-renewable resource management, embryonic development and drug designing<sup>1,3,6</sup>. In addition, mathematical models assist in understanding the efficacy of cancer treatments through post-treatment survival analysis of patients. These predictive models elicit profound impact on developing and designing the strategies to cure various communicable and non-communicable diseases. The statistical analysis of these models is carried out using theoretical approach as well as integrated use of SPSS and/ or Statistical Software 'R' (Open software package)<sup>6,8</sup>. Both of which provide solutions and excellent mean for graphical display of data.

Multi-factorial, post-treatment survival analysis<sup>5</sup> of patients using best applied mathematical model(s) enable the researchers and policy makers to develop deep insight about the pathophysiology, prophylaxis and treatment of various life-style, metabolic, and pathologic diseases like diabetes, hypertension, tuberculosis and cancers. However, due to lack of awareness amongst the patients, health-care sector faces severe challenge in proper collection of patient's treatment data/feedback. Continuous monitoring of pre- and post-treatment life history of patients helps better understand the causes, cure and after effects of treatment. It can further help to predict the survival range of an individual under varied socio-economic conditions. Post-treatment, survival range of patient is greatly influenced by age, sex, socio-economic status, and physical and mental health status. Inclusive analysis of multiple factors for survival analysis along with mathematical modelling helps better understand the effectiveness of treatment on patient's health and survival range. However, its success

rate and precision entirely depends on active and voluntary participation of patients, quality of data, and sample size.

In current study, survival analysis secondary data for a group of 250 patients underwent chemotherapy for lung cancer<sup>4</sup> was statistically re-analysed through 'R'<sup>7</sup>. It is of interest that chemotherapy treatment is designed to kill the mitotically active cells. However, drug used in the treatment is non-selective, and thus, along with cancerous cells; it kills various other mitotically active cells required for normal physiological functions. As a consequence, other physiological system in patients underwent chemotherapy are greatly compromised and works at sub-optimal level. Healing of the patient from deleterious effect of chemotherapy is greatly depends upon various factors including age, sex, access to healthy food, socio-economic and mental health status. In order to develop the best suited mathematical model, patient's data about their age, sex, calorie intake, body weight and the survival range was re-analysed using statistical software 'R'. All-inclusive analysis was performed to understand the cumulative effect of various selected parameters on survival range of the individuals.

## MATERIALS and METHODS

Statistics is the science of making decisions based on collection, analysis, interpretation, and expressive representation of data. Usually, these decisions are made under conditions of uncertainty or under various defined conditions. Often situations or problems faced in real world have no precise or definite solution. Statistical methods help to make scientific decisions in such situations. In today's world, data analysis and mathematical modelling help the researchers to assess the effectiveness of an invented medical treatment towards a disease.

Survival analysis is a collection of data, its statistical analysis with respect to a time until an event occurs. Time may include years, months, weeks, or days from the beginning of follow-up of an individual until an event occurs. Alter-

natively, time can refer to the age of an individual when an event occurs. An event could be the death, disease incidence, relapse from remission, recovery or any designated experience of interest that may happen to an individual. In a survival analysis, time variable is usually referred as survival time. It provides information about survival of an individual post-treatment over follow up time. Further, death or relapse of disease was referred failure of an event. In the current paper, Statistical Software 'R' with the below mentioned packages were used:

```
install.packages("ggplot2")  
library("ggplot2")  
install.packages("survival")  
library("survival")  
install.packages("survminer")  
library("survminer")  
data("lung")
```

## RESULTS

### 1. Age related effect on survival time, body weight loss and calorie intake in post-chemotherapy male patients

During post-chemotherapy monitoring period, a positive correlation between age of the individual and survival time is observed in the age groups ranging from 40 to 60 years (Fig. 1.). With further progression of patient's age, survival time started to decline and attained nadir at the age group 80-85 years. On the contrary, a negative correlation was observed with regard to calorie intake. Amount of calorie intake successively reduced with the increase in age of individuals. On the other hand, inverse correlation between age and body weight loss was observed till the age group 52. Thereafter, with the progression of age, body weight of the patients largely remained unchanged.

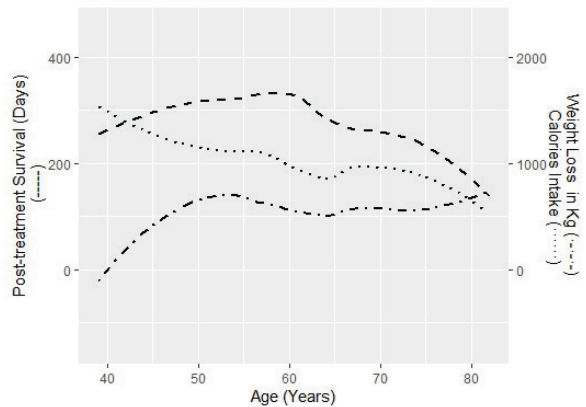


Figure 1. Post-treatment survival analysis of lung cancer male patients underwent chemotherapy. Secondary data for the post-treatment survival days, loss of body weight and calorie intake was obtained from [2], statistically analysed through software 'R'.

### 2. Age related effect on survival time, body weight loss and calorie intake in female patients post-chemotherapy

In female patients, double-bell shaped survival pattern was observed. While age group 45-50, and 67-70 years showed longer survival time periods, lowest was observed in age group 58-62 years. On the contrary, calorie intake in females was declined with the increase in age. A progressive loss in body weight of the patients was observed at all the ages, however, maximum weight loss was observed at the age group 58.

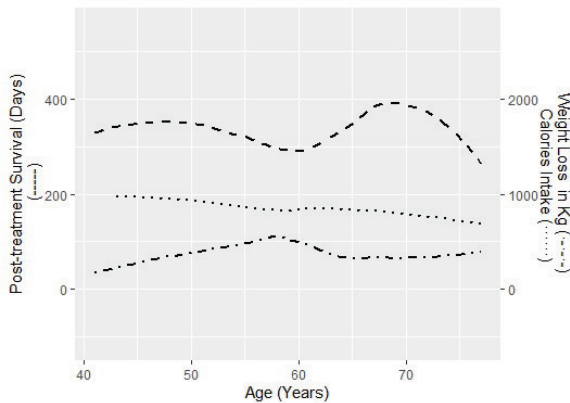


Figure 2. Post-treatment survival analysis of lung cancer female patients underwent chemotherapy. Secondary data for the post-treatment survival days, loss of body weight and calorie intake was obtained from [2], statistically analysed through software 'R'.

## DISCUSSION

Lung cancer is one the most fatal and common type of cancers in the world. By compromising the functioning of vital respiratory system, lung cancer poses high threats of survival to the patients. Numerous medico-clinical methods including chemotherapy treatment has always been considered as an effective mean to control the growth of cancerous cells. However, this treatment is non-selective that along with killing dividing cancerous cells, it non-selectively kills mitotically active cells of mucus membrane of digestive tract, hematopoietic stem cells, and hair follicles to name a few. As a consequence, chemotherapy adversely affects the digestive and hematopoietic functions of the body. This result in the chemotherapy side effects like anaemia, disturbed digestive functions, loss of appetite and body weight, and hair fall. In post-chemotherapy time, availability of balanced meal, socio-economic and mental status, and proper health care markedly influence the recovery rate of patient from harmful side effects of chemotherapy.

In current study, paired re-analysis of age-dependent effect on survival range, calorie intake and body weight loss

revealed that post-chemotherapy, mean survival time is higher in female as compared to that of males. This indicates the higher responsiveness of females toward chemotherapy, and recovery from disease. Though premature to comment, intake of calorie did not positively affect the survival range of the individuals as body weight appeared to decline in both males as well as females. It appears that post-treatment, calorie supply to the body largely is utilized to decrease the deleterious effect of chemotherapy and rejuvenation of damaged cells and tissues, instead of body mass increase and hence body weight. Further, detailed exploration considering numerous in depth parameters are required to strengthen the predictions for survival range in the cancer patients. Thus, it can be concluded that using appropriate statistical analysis, data can be re-analysed from multi-dimensions to make new predictions and design better methods to control/cure disease.

## Conflict of Interest

None declared by the authors.

## Financial Disclosure

None declared by the authors.

## Ethical Approval

In the current study, secondary data from Biecek et al., 2017 [4] has been used to re-analyse through statistical software 'R'. Thus, no ethical clearance is required for the study.

#### References

1. Barnes. B. and Fulford G. R. Mathematical Modelling with Case Studies using Maple and MATLAB, Third Edition, CRC Press, 2015.
2. Emmert-Streib F. and Dehmer M., Introduction to Survival Analysis in Practice, Machine Learning and Knowledge Extraction, 1, 1013–1038, 2019.
3. Jones D. S., Plank M. J. and Sleeman B. D. Differential Equations and Mathematical Biology, Second Edition, CHAPMAN & HALL/CRC Mathematical and Computational Biology Series, 2009.
4. Kassambara A., Kosinski M., Biecek P., Survminer: Drawing Survival Curves using 'ggplot2', R Package Version 0.3 2017, 1.
5. Kleinbaum D. G. and Klein M., Survival Analysis A Self-Learning Text, Second Edition, Springer New York, 2005.
6. Narayanachar Tattar P., Ramaiah S. and Manjunath B. G., A Course in Statistics with R, John Wiley & Sons, Ltd, 2016.
7. Therneau T. M. A package for survival analysis in S. R Package Version 1977, 2, 280. Available online: <https://www.semanticscholar.org/paper/A-Package-for-Survival-Analysis-in-S-Therneau/aba73f6957bce1648f066935f7ea85a99119be5d> (accessed on 1 July 2019).
8. Verzani J. Using R for Introductory Statistics, Second Edition, Chapman & Hall/CRC The R Series, 2014.