IBEROAMERICAN JOURNAL OF MEDICINE
iberoamericanjm
Journal homepage: www.iberoamericanjm.tk

Original article

Long-term monitoring of Cancer Mortality Rates in USA: A descriptive analysis using statistical process control tools

Engy Refaat Rashed[a], *Mostafa Essam Eissa*[b,*]

[a]*National Center for Radiation Research and Technology, Cairo, Egypt*
[b]*Independent PhD researcher and Candidate, Faculty of Pharmacy, Cairo University, Egypt*

ARTICLE INFO

Article history:
Received 28 March 2020
Received in revised form
1 April 2020
Accepted 5 April 2020

Keywords:
Cancer
Mortality ratio
SPC
USA
WHO

ABSTRACT

Background: Cancer is one of the most devastating diseases that influence humanity in the modern era. The effect is not confined to morbidity and mortality, but it is also extended to social and economic consequences despite the advancements made in the medical and healthcare fields. The present case overview will focus on the overall long-term assessment of death rates from malignancy from 1960 to 2017 as part of a global study of the mortality trend of this disease.
Methods: Statistical analysis and process control methodologies were used to study the death rate trend over this recorded period using statistical programs platform. A combination of techniques that could be used sequentially after database processing and stratification were used including distribution fitting, descriptive statistics, data fitting mathematical pattern, Gaussian Mixture Model (GMM) box plot and Individual-Moving Range (I-MR) trending chart.
Results: Two-parameter Weibull distribution (or Weibull 2) was the most appropriate that has fitted the distribution pattern of data and used for the construction of the process-behavior chart. The last two decades of the dataset showed a progressive decline in the mortality rates, which were almost linear with a higher magnitude of a negative slope than that of the initial rising pattern from 1960 until the 1990s. The hump-shaped trend showed underlying two distribution clusters: An initial distribution (I) of 81% of data of higher average mortality rates and minor one (II) that covers the last decade of monitoring record.
Discussion: A significant improvement in cancer healthcare was witnessed with a noticeable breakthrough in the last decade in the USA.

1. INTRODUCTION

Cancer is one of the most devastating diseases that affect humanity in the modern era. The effect is not confined to morbidity and mortality, but it is also extended to social and economic consequences despite the advancements made in the medical and healthcare fields [1, 2]. However, huge efforts have been effectively put in actions to enhance survivability chances of malignancy patients through extensive national and international organizations, especially in the developed nations. The present manuscript demonstrates the pattern of the survivability of cancer patients as a quantitative descriptive analysis of the total cancer mortality rates trend for modeling and assessment of the disease control over 58 years of monitoring.

* *Corresponding author.*
E-mail address: mostafaessameissa@yahoo.com

http://doi.org/10.5281/zenodo.3740610

2. METHODS

World Health Organization (WHO) internet database for the total cancer mortality ratios has been gathered for the USA as an overall death cases per 100000 of the affected individuals annually from 1960 to 2017 [3, 4]. Dataset in Excel was processed using XLSTAT V2014.5.03 built-in program that was used for the best fitting distribution analysis and data clustering by the implementation Gaussian Mixture Model (GMM) technique [5, 6]. Fitted line modeling for expressing and forecasting of mortality rates behavior was applied using Minitab® V17.1.0 [7], which was used also in drawing Box-and-Whisker plot, in addition to the construction of the control (process-behavior or trending) chart according to the closest distribution that fits data spreading.

3. RESULTS

The importance of statistical process control (SPC) methodologies for quantitative descriptive analysis of the disease would be demonstrated using statistical programs platform as the following:

3.1. STATISTICAL ANALYSIS AND QUANTITATIVE INTERPRETATION OF CANCER MORTALITY RATE

The distribution that fits best the data for the goodness of fit test is the Weibull (2) distribution. The estimated parameters were as the following: Parameter value with standard error (SE) [beta (β) 17.4016 ± 1.9222 and gamma (γ) 234.4892 ± 1.8652]. On the same line, statistics that were estimated on the input data of total mortality rates from malignancies in the USA and computed using the estimated parameters of the Weibull (2) distribution were presented as the following: Statistics (Data, Parameters) are [Mean (226.8982, 227.4390), Variance (328.1195, 260.0011), Skewness (Pearson) (-0.9484, -0.8313) and Kurtosis (Pearson) (-0.0471, 1.1410)]. Test for successful modeling was done using two means. The first goodness of fit analysis is Kolmogorov-Smirnov (KS) test [D = 0.0998,

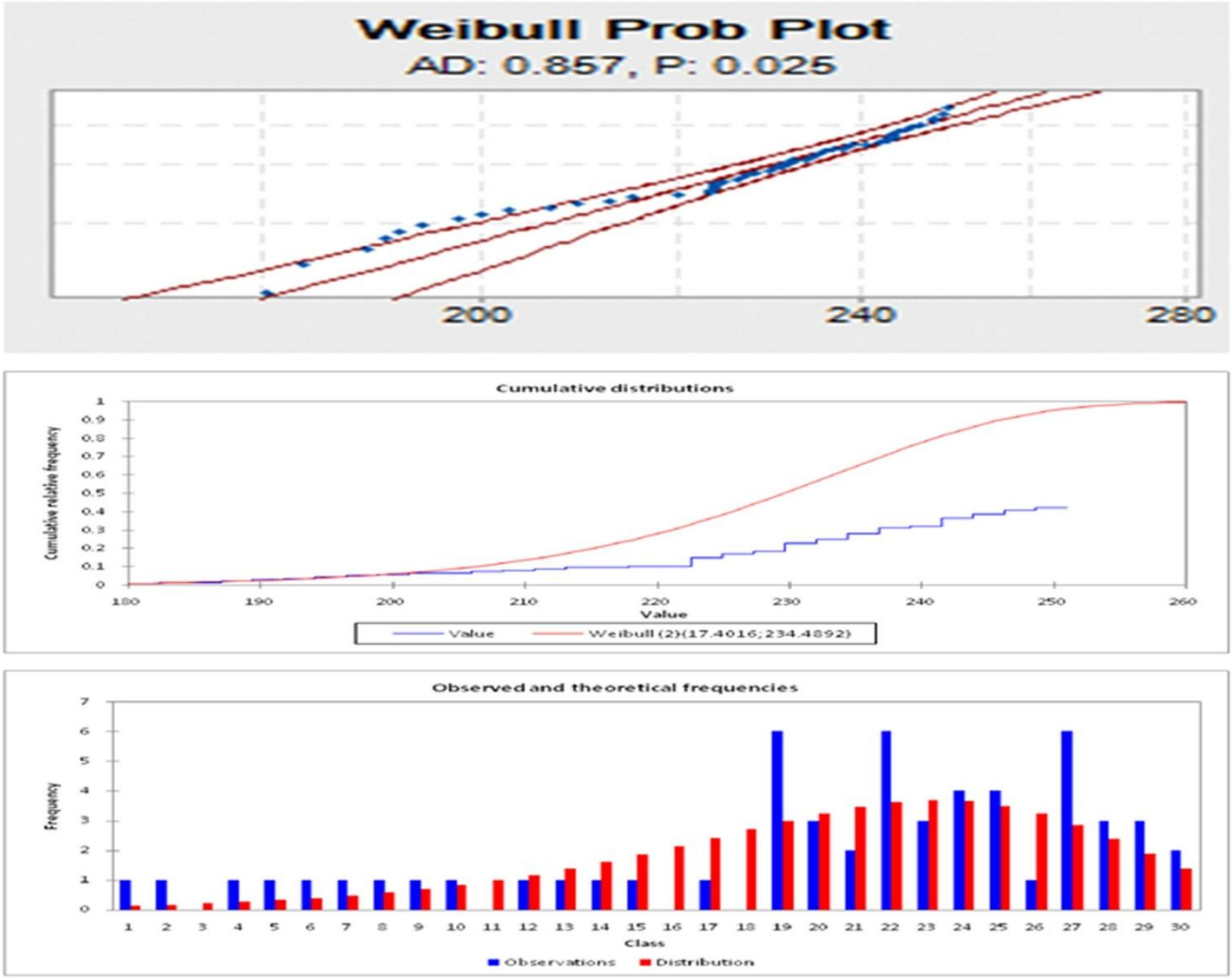

Figure 1: *Probability plot showing the degree of distribution fitness (upper graph). Cumulative distribution pattern of both practical and theoretical assumed dispersions (middle graph). best distribution fitting approach showing gap between the typical and actual dispersion pattern (lower graph).*

p-value = 0.5988 and alpha (α) = 0.05]. Best fitting distribution modeling was illustrated in Figure 1. The test interpretation was conducted as the following: H_0 is the assumption that the sample is following a Weibull (2) distribution and H_a is the hypothesis that the sample does not follow a Weibull (2) distribution. As the computed p-value is greater than the significance level alpha (α) =0 .05, one cannot reject the null hypothesis H_0. The risk to reject the null hypothesis H_0 while it is true is 59.88%. Secondly, Chi-square test analysis results were as the following: Chi-square (Observed value) = 33.3957, Chi-square (Critical value) = 40.1133, DF = 27, p-value = 0.1843 and alpha (α) = 0.05. Similar to KS or K-S test, the test interpretation hypothesis was as the following: H_0: The sample follows a Weibull (2) distribution H_a: The sample does not follow a Weibull (2) distribution. As the computed p-value is greater than the significance level alpha (α) =0.05, one cannot reject the null hypothesis H_0. The risk to reject the null hypothesis H_0 while it is true is 18.43%. Thus, Weibull (2) predictive distribution with p-value 0.5988 was the most appropriate descriptive pattern of the cancer death rates data in the USA. Spearman correlation r-value was -0.3088 with a 95% confidence interval (CI) of -0.5327 to -0.04448 for time (x) and death rate of cancer (y) 57 pairs. However, this negative and weak correlation was found to be significant at α = 0.05 with an approximate two-tailed P-value of 0.0194. The mean rate was 227±18, minimum and maximum numbers of deaths per 100000 individuals were 181 and 250, respectively with a range of 69.

3.2. MODELING OF CANCER DEATH RATE RECORD PERIOD FOR PREDICTION AND FORECASTING

Fitted line modeling of mortality rates values versus time (in years) was done through two perspectives as could be seen in Figure 2. First, the overall curved pattern was fitted using polynomial regression analysis and the regression equation was as the following:

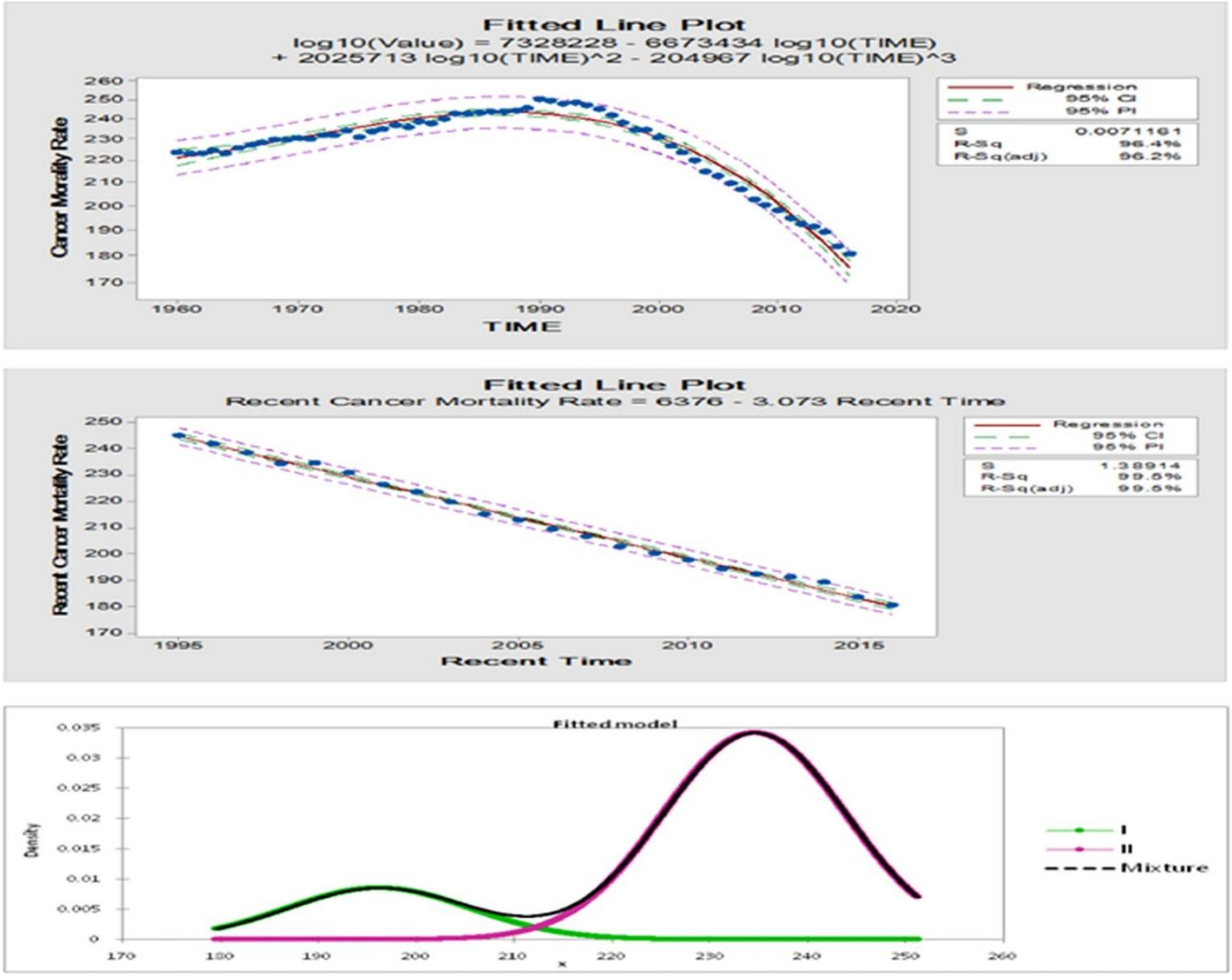

Figure 2: *Fitted line plot showing confidence interval (CI) and predictor interval (PI) at 95% confidence level: of logarithmic transformation to base ten with cubic regression model type for overall trend (upper graph), linear regression model type for the descending part of the curve (middle graph)and GMM showing cancer mortality rates trend as interfering two bell-shaped distributions analyzed as best fitted model (lower graph).*

$\log_{10}(y) = 7328228 - 6673434 \log_{10}(x) + 2025713 \log_{10}(x)^2 - 204967 \log_{10}(x)^3$.....eq.(1)

Where: y is the total mortality rate per 100000 persons from cancer and x is the time in years.

The degree of conformity of this modeling estimates could be ensured from S = 0.00711611, R-Sq = 96.4% and R-Sq(adj) = 96.2%. Analysis of Variance showed the source values of regression Degrees of Freedom/Sum-of-Squares/Mean Squares/F values/P values (DF/SS/MS/F/P) as 3/0.0710713/0.0236904/467.83/0.000 and error DF/SS/MS as 53/0.0026839/0.0000506 with a total DF/SS as 56/0.0737552. Interestingly, the last 22 years of trend demonstrated almost a linear decline in the death rate of cancer stimulating the isolation of the nearly last two decades of the record and analyzing the trending pattern. Again, the regression equation was calculated as the following:

y = 6376 - 3.073 x.....eq.(2)

Where: y is the recent cancer mortality rate per 100000 persons and x is the time in years.

The degree of conformity of this modeling estimates could be ensured from S = 1.38914, R-Sq = 99.5% and R-Sq(adj) = 99.5%. Analysis of Variance showed the source values of regression DF/SS/MS/F/P 1/8363.66/8363.66/4334.15/ 0.000 and error DF/SS/MS as 20/38.59/1.93 with a total DF/SS as 21/8402.25. Thus, forecasting the cancer death rate would be expected to diminish in the year 2075 provided constant conditions.

3.3. CLUSTERING PATTERN OF CANCER DEATH RATE USING GAUSSIAN MIXTURE MODEL (GMM)

Segregation of long-term death rates from cancer data in the USA showed that two class patterns could be isolated and identified, denoted by I and II or simply 1 and 2. The percentage contribution of each class was 19.98% and 80.02%, respectively. The mean mortality rate values of classes 1 and 2 were approximately 196 and 235 deaths per

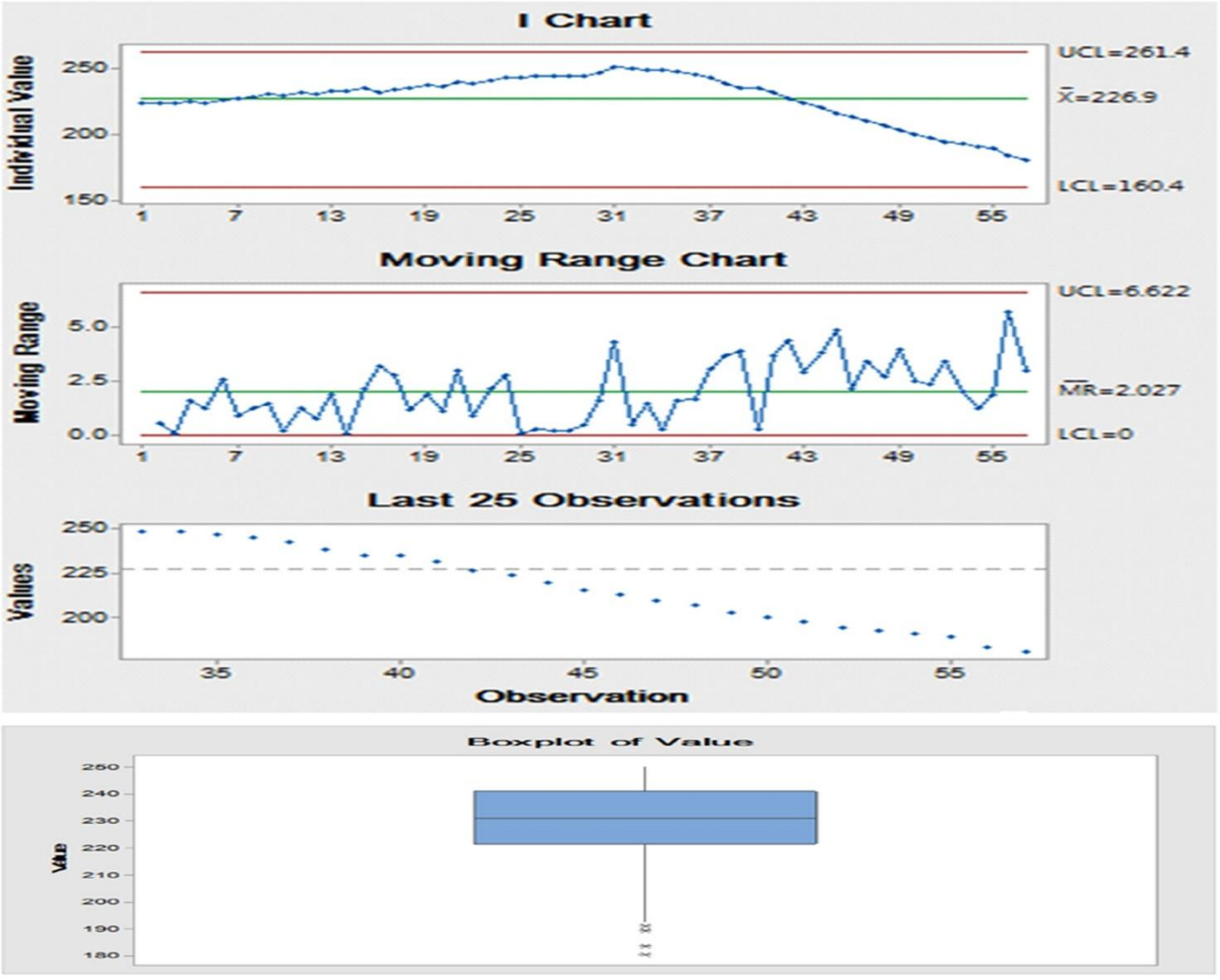

Figure 3: *Process-behavior chart of I-MR type showing the overall trend of cancer mortality rates, variability of death rate, the pattern of the latest 25 years period and box plot with exceptionally low rates of cancer mortality.*

100000, respectively with a variance of about 87. The selection criterion for the proposed model with DF of four and Log-likelihood of -234.1184 would be -484.4091, -476.2369, -489.6470, 0.2304 and 2.6190 for Bayesian Information Criterion (BIC)/Akaike Information Criterion (AIC)/Integrated Classification Likelihood (ICL)/Normalized Entropy Criterion (NEC)/Entropy, respectively. Accordingly, the Normalized Entropy Criterion (NEC) criterion is lower than one; there is a clustering structure in the data. These distributions were expressed graphically in Figure 2.

3.4. PROCESS-BEHAVIOR CHART AND BOX PLOT FOR TRENDING OF THE OVERALL CANCER MORTALITY RATES IN THE USA

Box-and-Whisker showed normally distributed data that passed the test at 99% confidence with a tendency of the lowest recent records to fall outside the common pattern of data. This aberrant and exceptional decline in the mortality rates would be evident from the behavior of the last 25 readings that illustrated a continuous and steady decrease in the number of death cases in 100000 affected individuals. Individual-Moving Range (I-MR) control chart was constructed based on the underlying assumed dispersion pattern according to the best distribution fitted to the data. In the first component MR chart, the inter-annual general trend variation of death rates showed a tendency to rise in values, especially in the latest years within the last two decades of the record observations. Complementarily, The mortality rates - in I chart - were changing within the window of the Upper Control Limit (UCL) and Lower Control Limit (LCL) of the monitored disease. This burden represents the tolerance range of the death rate from the general long-term record that had fallen between about 160 and 261 death cases in 100000 individuals. These findings were demonstrated visually in Figure 3.

4. DISCUSSION

The present study illustrated one of the serious challenges of the post-world war II period that influenced the life of modern man [8, 9]. Malignancy impacts the human lifestyle adversely with primarily negative social and economical consequences in any affected nation [10-12]. SPC methodologies and statistical analysis tools are useful means in data mining and the assessment of the diseases quantitatively when suitable and appropriate datasets are available as well as in the main application in the industrial field [13-15]. Total mortality rates showed significant improvement in cancer healthcare was witnessed with a noticeable breakthrough in the last decade in the USA, which was evidenced as exceptionally low ratios of the total cancer mortalities [16]. The gradual rise in the death rate from the sixties until the nineties in the twentieth century was followed by a relatively rapid decline in mortalities ratios that should theoretically through the seventies of twenty-one century if the existing healthcare conditions were maintained for the cancer victims as from 1991 to 2017. The hump-shaped curve could be actually divided into three regions: an initial gradual increase in the mortality rates with a small positive slope from 1960 to 1991 [17], the transition period from 1991 to 2000 with a little or no increase and a start of descending pattern of the death rates and the decline phase line with statistically significant improvements during the last decade viz 2007 to 2017 [18]. Thus, the mixed nature of data had lead to the polynomial approach of the fitting. However, the initial and last segments were best fitted to a linear plot. The continuous improvement in cancer patients' survivability was indicated by outliers points in the boxplot diagram, control chart, negatively skewed distribution. Thus, the disease mortality ratios over the study period showed good fitness to the two-parameters Weibull distribution, which was demonstrated by other researchers in a previous work [19]. GMM detection of this new trend was evident as new minor distribution with a low average value of mortality. The control of malignant diseases could be sensed significantly during recent years. It is the positive outcome of the national directed efforts that started at 1971 following the signing of the National Cancer Act to stimulate extensive scientific work and research to combat malignancy in an effective fashion [20]. Nevertheless, enhancing survivability chances would be questionable due to other unexpected factors that may intercept cancer treatment, especially with the failure risk of containment of the recent Corona Virus Disease 2019 (COVID-19) viral global pandemic that could affect severely the life of the already health-defected populations.

5. REFERENCES

1. Weatherall D, Greenwood B, Chee HL, Wasi P. Science and technology for disease control: past, present, and future. In: Jamison DT, Breman JG, Measham AR, Alleyne G, Claeson M, Evans DB, et al, editors Disease control priorities in developing countries, 2nd ed. Washington DC: World Bank; 2006:119-38.

2. Alwan A. Global status report on noncommunicable diseases 2010. Geneva: World Health Organization; 2011.

3. Comprehensive Cancer Information [Internet]. National Cancer Institute. 2020. Available from: https://www.cancer.gov/ (accessed March 2020).

4. Who.int. 2020 [Internet]. Available from: https://www.who.int/ (accessed March 2020).

5. Gaussian mixture models [Internet]. XLSTAT, Your data analysis solution. 2020. Available from: https://www.xlstat.com/en/solutions/features/gaussian-mixture-models (accessed March 2020).

6. XLSTAT Support Center [Internet]. Help.xlstat.com. 2020. Available from: https://help.xlstat.com/s/article/gaussian-mixture-model-clustering-in-excel-tutorial?language=en_US (accessed March 2020).

7. Minitab LL. Getting Started with Minitab 17. PA, USA: Minitab Inc., State College; 2014:73.

8. Falzone L, Salomone S, Libra M. Evolution of cancer pharmacological treatments at the turn of the third millennium. Front Pharmacol 2018;9:1300. doi: 10.3389/fphar.2018.01300.

9. Zarros A, Tansey T. Pharmaceutical innovation after World War II: from rational drug discovery to biopharmaceuticals. Front Pharmacol. 2019;10:834. doi: 10.3389/fphar.2019.00834.

10. Braveman P, Gottlieb L. The social determinants of health: it's time to consider the causes of the causes. Public Health Rep. 2014;129(suppl 2):19-31. doi: 10.1177/00333549141291S206.

11. Miller KD, Nogueira L, Mariotto AB, Rowland JH, Yabroff KR, Alfano CMet al. Cancer treatment and survivorship statistics, 2019. CA: Cancer J Clin. 2019;69(5):363-85. doi: 10.3322/caac.21565.

12. Allart-Vorelli P, Porro B, Baguet F, Michel A, Cousson-Gélie F. Haematological cancer and quality of life: a systematic literature review. Blood Cancer J. 2015;5(4):e305. doi: 10.1038/bcj.2015.29.

13. Eissa ME, Seif M, Fares M. Assessment of purified water quality in pharmaceutical facility using six sigma tools. Int. J Pharm Qual Assur. 2015;6(2):54-72.

14. Eissa ME. Variable and attribute control charts in trend analysis of active pharmaceutical components: Process efficiency monitoring and comparative study. Experimental Medicine (EM) 2018;1(1):32-44. doi: 10.31058/j.em.2018.11003.

15. Eissa M. The attribute control charts for outbreak trends o f selected states in the USA: a brief report of the insight into the pattern. Int Med. 2019;1(1):11-4. doi: 10.5455/im.31744.

16. Sheikh K. Cancer Death Rate in U.S. Sees Sharpest One-Year Drop [Internet]. Nytimes.com. 2020. Available from: https://www.nytimes.com/2020/01/08/health/cancer-deaths-decline.html (accessed March 2020).

17. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. CA Cancer J Clin. 2020;70(1):7-30. doi: 10.3322/caac.21590.

18. American Cancer Society. Cancer mortality continues steady decline, driven by progress against lung cancer: Drop of 2.2 percent from 2016 to 2017 is largest ever-reported [Internet]. ScienceDaily. 2020. Available from: https://www.sciencedaily.com/releases/2020/01/200108074809.htm (accessed March 2020).

19. Nadler DL, Zurbenko IG. Developing a weibull model extension to estimate cancer latency. ISRN Epidemiology 2012;2013. https://doi.org/10.5402/2013/750857.

20. Arteaga CL, Adamson PC, Engelman JA, Foti M, Gaynor RB, Hilsenbeck SG, et al. AACR Cancer Progress Report 2014. Clin Cancer Res. 2014; 20(19 Suppl):S1-S112. doi: 10.1158/1078-0432.CCR-14-2123.