

Research Article

Open Access



¹ National Center for Radiation Research and Technology, Cairo, Egypt

² Independent PhD researcher and Candidate, Faculty of Pharmacy, Cairo University, Egypt

Contacts of Authors



* To whom correspondence should be addressed: Mostafa Essam Eissa

Citation: Rashed ER, Eissa ME (2020). Long-Term Quantitative Assessment of Women Survivability from Cancer: A Unique Descriptive Analysis. Highlights in BioScience Volume 3. Article ID 20208. doi:10.36462/H.BioSci.20208

Received: June 4, 2020

Accepted: July 3, 2020

Published: July 5, 2020

Copyright: © 2020 Rashed and Eissa. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and supplementary materials

Funding: The authors have no support or funding to report.

Competing interests: The authors declare that they have no competing interests.

Long-Term Quantitative Assessment of Women Survivability from Cancer: A Unique Descriptive Analysis

Engy Refaat Rashed¹, Mostafa Essam Eissa^{*2}

Abstract

Statistical Process Control (SPC) methodologies are a set of statistical methods and techniques that were initially designed for industrial processes but could be adopted for non-industrial applications. The current prospective study aimed to provide a unique quantitative investigation of an epidemiological disease using the SPC program platform. The selected case herein was a long-term monitoring record of the yearly cancer mortality rates in women worldwide. Multidimensional segregation of the dataset into subgroups was conducted to visualize the clustering pattern based on nations (42 countries as boxplot), time and the Gaussian Mixture Model (two-interfering bell-shaped distributions approach). The trend of death rates versus the elapsed years would demonstrate a moderately negative correlation with the time following the theory of splines. Construction of control chart based on the fitted Weibull distribution showed a gradual steady improvement in survivability rates from malignancy. The greatest variations in the mortality ratios existed within the European countries.

Keywords: Cancer, Gaussian mixture model, Mortality rate, Weibull.

Introduction

Malignancy is one of the most devastating diseases that affect humanity in the modern era of civilization. It threatens human lifestyle quality, productivity, financial and economic status of society and country, in addition to morbidity, disability and mortality [1]. Notably, women's health and well-being are severely impacted by the rising challenge of cancer disease in the community, which hinders them from serving their crucial role in any rising and developed nations. The burden of cancer among women is high in both High Income Countries (HICs) and Low Middle Income Countries (LMICs), although the distribution of most common cancers differs. This burden is predicted to extend as populations grow and age and because the prevalence of cancer risk factors increases in some countries, especially in LMICs. The costs of cancer are considerable and even catastrophic in HICs and LMICs alike. However, this burden of disease, loss of life, and economic hardship is not inevitable [2]. All of the foremost common cancers among women worldwide, including lung, breast, cervix, liver, and colorectum, have known means of prevention and/or early detection, which may be used to cut back incidence and mortality. Furthermore, carcinoma and cervical cancer, two of the highest four cancers in women worldwide have several proven prevention measures [3]. These two cancers combined represent about 20% of all cancer deaths among women [4].

Many of those deaths may be prevented through effective tobacco control, vaccination, and screening activities. There are varieties of effective cancer control measures available to countries of all resource levels [4]. Many of those measures are extremely cost-effective given the lives saved for the cost of the intervention, especially within the case of vaccination. To stop cancer within the future, countries must prioritize policies to cut back known cancer risk factors and make prevention accessible to any or all. For those that have cancer today, effective treatments and palliative care are also needed. Additionally, to those needs, cancer surveillance and research for the prevention and treatment are indispensable for the setting of cancer control priorities and for determining the foremost effective interventions and coverings in an exceedingly given context [4]. For LMICs, all of those activities may require support and commitment from the worldwide community.

While malignant diseases are holding and exhaustive and comprehensive databases that were collected from extensive surveys and gathering centers from national and international organizations, it is the interpretation of the results that would derive useful conclusions and insight through record's pattern [5, 6]. The application of Statistical Process Control (SPC) methodologies was historically started at the end of the first quarter of the 20th century in the industrial field the monitoring, control and improvement of the manufacturing processes as well as product quality by Walter A. Shewhart at Bell Laboratories [7, 8]. In addition to the wide industrial applications, SPC techniques have been applied even in the improvement of service fields.

Moreover, SPC techniques were found to be useful in other non-industrial fields for assessment, evaluation and investigation of specific inspection properties or events [9, 10]. Some interesting fields of implementation include - but not limited to - environmental monitoring (EM), microbiological water quality, surgical site infection, epidemiological diseases and outbreaks [11-15]. The previous studies embraced control charts, Pareto diagrams, Box-and-Whisker plots, distribution identification, correlation analysis and fitted line curves, in addition to other conventional analysis tools such as descriptive statistics. While epidemiological diseases constitute a continuous threat that strike humanity, it would be useful to develop a quantitative descriptive tool that describes the progression and behavior of these diseases.

One of the important indicators that could provide insight into cancer progression is the mortality ratio of the patients that could be found in the database of national and international health organizations. The application of statistical tools and process control methodologies was sought as a unique metric means in the evaluation and comparison of the epidemiological status that could support in the decision-making and cancer management and control.

Accordingly, the present analysis might reveal new horizons for researchers to a simple, fast and useful way to study other diseases, outbreaks and pandemics.

Materials and Methods

SPC Perspective of Long-Term Global Trend of Cancer in Women

Cancer mortality rates records of different countries were obtained as a downloadable database file from the websites: <https://www.cancer.gov/> and <https://www.who.int/> [16, 17] and women data were extracted from the Excel file by filtering and arranged chronologically (**Table S1**). Three different computer programs were used in data processing combined as statistical software platform viz Prism - GraphPad v6.01, Minitab v17.1.0 and Excel built-in XLSTAT v2014.05.03 [18-21]. The Prism - GraphPad v6.01 commercial scientific software was used for the preliminary data description followed by defining the significance estimation of the change of the yearly mortality rates with time. The Excel built-in XLSTAT v2014.05.03, which is a statistical suite add-in, was used for drawing of Gaussian Mixture Model (GMM), which is a probabilistic model to spot typically disseminated subpopulations inside a general populace. Blend models, as a rule, do not require prior knowledge of datasets subordination to the subgroup, permitting the model to become familiar with the subpopulations spontaneously [22]. Thus, it will be important in this case for viewing the overall data-clustering pattern for the record from 1960 to 2017.

On the other hand, Minitab as a statistical tool package was assigned for fitting line modeling, box plot creation, histogram drawing and data trending using process behavior chart [23]. Two-dimensional data segregation was done based on GMM and individual countries found in the record to identify clustering pattern of data. Construction of the mean (\bar{X} -bar)-variation (\bar{X} -S) trending chart was executed according to the distribution identification scheme in Minitab. Shewhart charts show mean of the death rate and the standard deviation along with the event window i.e. Upper Control Limit (UCL) and Lower Control Limit (LCL).

Results and Discussion

Descriptive Statistics and Correlation Output

A preliminary examination of the database for the basic statistical analysis yielded a result in **Table 1** showing the coinciding countries with the annual values. Data are mildly positive or right-skewed [24]. Failure to comply with normality distribution was followed by the screening of the closest possible spreading pattern that might fit the record. A Weibull distribution hypothesis could not be rejected as a suitable fit for the raw dataset based on the Anderson-Darling test (AD).

Table 1: Descriptive statistics of the cancer mortality rates showing the number of deaths per 100000 individuals at CI* 95% and the corresponding countries numerically to the parameters value using Prism - GraphPad v6.01

Column Statistics	Mortality Rates (per 100000)	Concerned Department
Minimum	93.30	TUR 2009
25% Percentile	159.1	ZAF 2005
Median	183.8	AUS 1963, FIN 1984, POL 1985
75% Percentile	205.4	GBR 1962, FIN 1964, CHL 1991, NLD 1997
Maximum	286.4	ISL 1960
10% Percentile	140.9	GRC 1983
90% Percentile	227.4	LUX 1984, CZE 1996
Arithmetic Mean	183.2	SWE 1987
Std. Deviation	32.85	NA
Std. Error of Mean	0.7461	NA
Lower 95% CI of mean	181.8	AUS 1975, FIN 1980, CRI 1990
Upper 95% CI of mean	184.7	SWE 1984, SVN 1988, NOR 2003, NZL 2010, CZE 2012, GBR 2016
Lower 95% CI of median	181.8	AUS 1975
Upper 95% CI of median	185.4	SWE 1983 and 1988, ITA 1990
Shapiro-Wilk normality test W	0.9973	NA
P value	0.0017	
Data passing normality test (alpha=0.05)	No	
Coefficient of variation	17.93%	
Geometric mean	180.2	SVK 2006, ISL 2010
Lower 95% CI of geo. mean	178.7	POL 2011
Upper 95% CI of geo. mean	181.7	EST 1990
Method of outlier detection: ROUT ^c (Q = 10.00%)	No aberrant record	NA
Skewness	0.04484	NA
Kurtosis	-0.3441	NA

*CI = Confidence Interval ; NA = Not Applicable; ^c Robust regression and Outlier removal

No aberrant values could be detected through the overall record. Thus, mortality rates showed apparent concatenation in the values with a range of 193.1 deaths/100000 cases with half and four-fifths of data record resides within a range of 46.3 and 86.5 deaths/100000 cases, respectively, suggesting extensive tailing. The mean and the median were almost coincided with the geometric mean very close to the average confirming that there is no extreme value(s) distant from the group collectively [25]. However, the arithmetic mean is slightly greater than that of the geometric mean as expected for the basic rule in mathematics [26]. This gap is reflected similarly on the Confidence Intervals (CIs) which showed the corresponding shift too. The relative standard deviation (RSD) showed significant dispersion of data around the mean of about 18%. In addition, Kurtosis value suggests data broadening with shoulders that is a variation from the normal bell-

shaped spreading if the normal distribution is expected, a sign of possible mixed intervening patterns [24]. This would probably be the result of the impact of annual and regional variations that was reflected in the total record of women cancer mortality. **Table 2** demonstrates the moderately negative correlation between the women mortality rates and the time in years [27, 28]. Accordingly, there was an improvement of women survivability with time passing, which might be partially attributed to the development of effective detection and control measures for the containment of cancer epidemiology [29].

Dataset Clustering Approaches: A Multidimensional Analysis

Data stratification using GMM analysis estimated two possible mixed bell-shaped spreading, indicating a global overlapping distribution of mortality rates among the studied nations in **Figure 1**.

Table 2: Non-parametric correlation of the mortality rates from malignant diseases versus time (in years) at CI 95% with two-tailed P-value using Prism - GraphPad v6.01

Spearman (r)	Time versus Death Rate Interpretation
r	-0.4110
95% CI*	-0.4484 to -0.3721
P (two-tailed)	< 0.0001
P value summary	****
True or asymptotic P value calculation**	Asymptotic
Significance at $\alpha = 0.05$	Yes
Number of paired XY	1938

*CI = Confidence Interval; **Refer to the calculation method of p-value using either the exact or approximation to the actual distribution; **** $P \leq 0.0001$

Cumulative Distribution Function (CDF) and Q-Q plots showed the convenience of this assumption. The first and the second distributions were called (Φ Phi) and (Ψ Psi) with contribution proportions 0.57 and 0.43, respectively (**Figure 1**). However, there was no clustering tendency that could show any heterogeneity according to the Normalized Entropy Criterion (NEC), which was greater than one with the selected the Integrated Completed Likelihood (ICL) criterion with the standard algorithm (EM) Expectation Maximization [30]. The means and covariances of Φ and Ψ

for malignancy mortality rates in women were 222.7, 241.0 and 1892.4, 909.9, respectively.

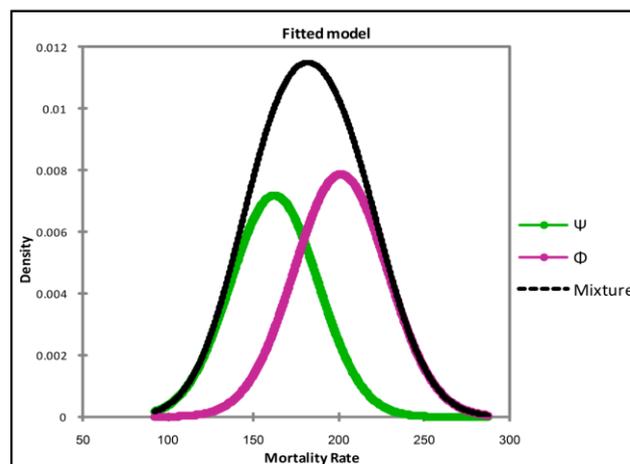


Figure 1: Dataset segregation, based on Gaussian Mixture Model (GMM) approach, showing two mixed distribution pattern with low (green) and high (pink) rates of mortality from malignancy, generated using XLSTAT v2014.05.03.

The geographically based data segregation shows that, ascendingly, Turkey, the Republic of Korea, Brazil, Mexico and Greece demonstrated the lowest average mortality rate values could be demonstrated in **Figure 2**. While countries with the ascending order from the lowest order of survivability to the highest means were Denmark, Hungary, Ireland, Czechia, GBR and Netherlands (**Figure 2**).

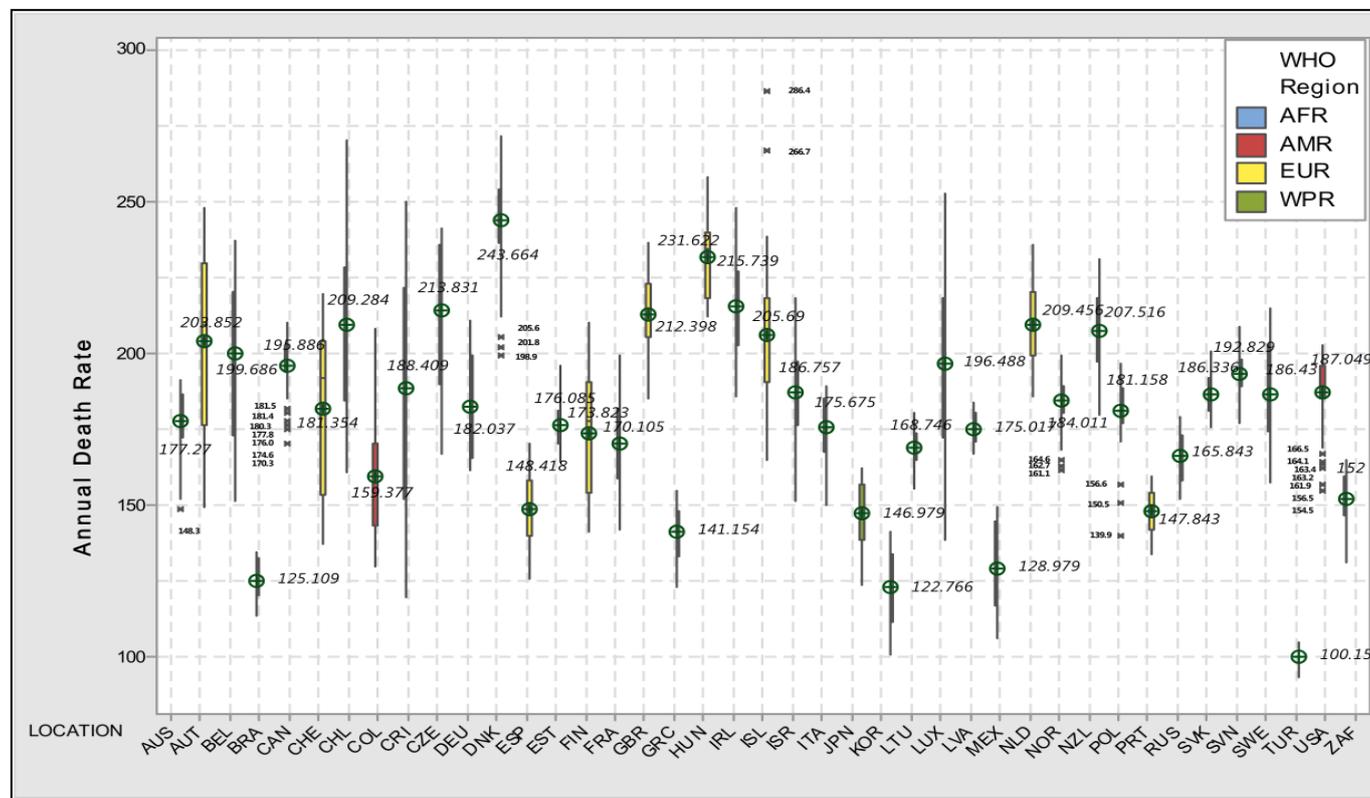


Figure 2: Dataset segregation, based on WHO geographical distribution by countries shown as Box plot, indicating the means, data spreading and the aberrant annual mortality rates records generated using Minitab v17.1.0.

The highest mortality rates from cancer occurred primarily in the EUR region with the domination of economically rich nations suggesting that the overall survivability of women from cancer depends on other factors that seemed to be important to be controlled in addition to the improvement of the healthcare quality. This assumption might be partially in agreement with previous analysis provided by some researchers that had linked the prevalence of some types of cancer in the high-income countries with some adverse lifestyles such as smoking, obesity and alcohol consumption [31, 32]. Nevertheless, it should be noted that beforehand rigorous control, awareness and actions taken about cancer detection and diagnosis associated with controlled monitoring might be one of the important reasons for the comprehensiveness of data records of high-income and developed nations if compared with those from the relatively lower income-countries. This hypothesis may be a factor that contributed to a possible

underestimation of the actual overall estimation of the mortality rates in the developing and low-income countries.

The time-based data segregation shows that, Europe WHO region was the most versatile with divergent mortality rates between nations, as it is rich with country data, in contrast to the African area that includes only South Africa. This could also be seen on the extend of the outlier figures (denoted by asterisks) above and below the box plot graphs in **Figure 3**. South Africa data record only started in 1993, which is later than the other three regions by 33 years and reaching its peak in 1999 (**Figure 3**). The annual spreading of death cases rates for AMR and WPR was less than that of EUR. However, the distribution of WPR values was generally more compact than that of AMR, where greater pervasion could be observed between annual records. The general profile can be seen in **Figure 3** with the yearly steady and almost regular decline in death rates started with AMR in late 1980s, EUR and WPR at almost 1995.

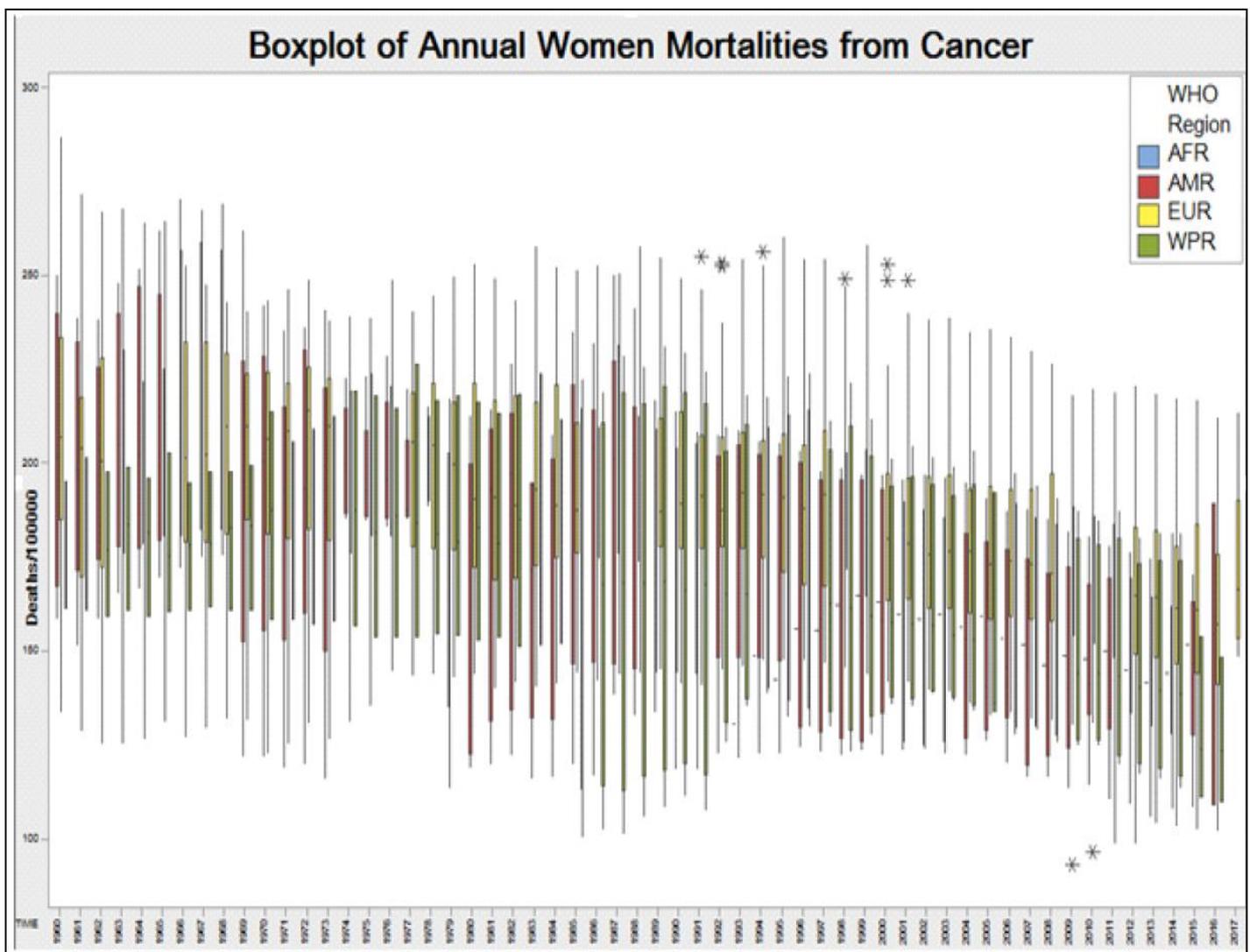


Figure 3: Dataset segregation based on WHO geographical distribution by time order shown as Box plot, indicating data spreading and the aberrant mortality rates records generated using Minitab v17.1.0.

Fitted Line of Yearly Average Death Rates from Cancer for the Affected Women Populations versus Time

Dataset pattern showed a steady decline in the global mortality rates from cancer after the initial high plateau that could be demonstrated by the quadratic relation with significant correlation (Figure 4). The regression equation was expressed through Polynomial Regression Analysis of average mortality rates from cancer for women versus the time (in years).

$$\text{Log}_{10}(Y) = -10245 + 6222 \text{Log}_{10}(T) - 944.5 \text{Log}_{10}(T)^2 \dots \text{eq.(1)}$$

Where: Y and T are the years and death rates from cancer worldwide were recorded for women, respectively, with $S = 0.00577903$, $R\text{-Sq}$ and $R\text{-Sq(Adj)} = 97.0\%$. Figure 4 illustrates this relation between death rates and years graphically.

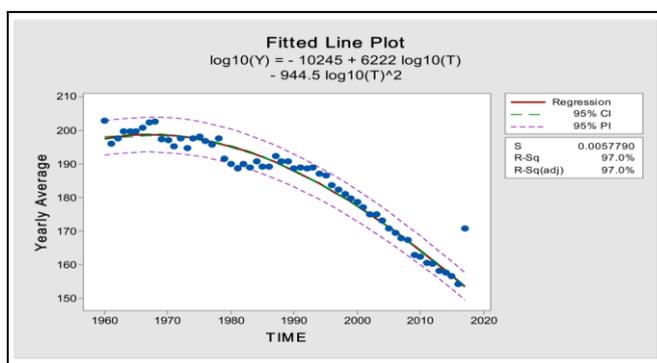


Figure 4: Polynomial regression analysis graph showing quadratic equation modeling, Confidence Interval (CI) and Prediction Interval (PI) along with R^2 and the Standard Error of the Regression (S), generated using Minitab v17.1.0.

Trending of Global Women Survivability from Malignancy Using Process-Behavior Chart

In Figure 5, Histogram showed a pattern of dual hump that could be explained from GMM model. Boxplot graph is indicative of the pattern for the studied WHO regions (Figure 5). AFR demonstrated the lowest mean and spreading due to the inclusion of ZAF only for relatively a short period from the region, in contrast of EUR with upper and lower aberrant records from some countries in the European region from Island (1960) and Turkey (2009, 2010 and 2011). The survivability rate in WPR is better than AMR. Analysis of the distribution fitting showed the convenience of the Weibull distribution based on the probability plot examination as could be observed in Figure 5. Plots the method mean (Xbar chart) and method standard deviation (S chart) over time for variable information in subgroups (Figure 5). This dual type of control chart is extensively used to observe the steadiness of procedures in lots of fields. Both the Xbar and S charts are displayed together due to the need to interpret each chart to decide if the trend is stable [32]. The procedure variation in s chart was almost stable within the control limits and tendency toward lower annual variation since 1979. The Xbar chart in Figure 5 showed initially intermittent excursions in the first few years followed by a successive shift in the annual mortality means leading to a declining curve with the improvement of survivability over time until exceeding the lower bound, which is a desirable outcome [33]. The last record of 2017 was incomplete during data registration that may explain the apparently sudden rise in mortality ratio due to the inclusion of few countries with high values in the process-behavior chart of Figure 5.

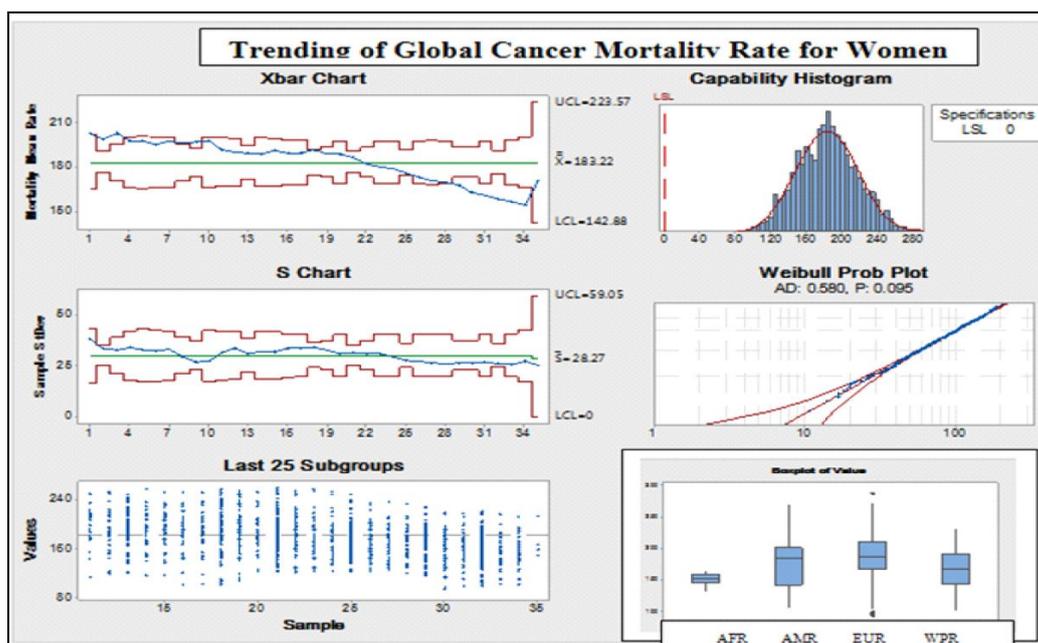


Figure 5: SPC analysis tools, (control chart, histogram and box plot) of the global cancer mortality rates in women with Weibull probability fitting plot, showing the overall trend of the annual death rates as \bar{X} -S control chart and the pattern of WHO regions as box and whisker plot, generated using Minitab v17.1.0.

Conclusion

SPC methodologies could provide a unique insight into the characteristics and patterns of epidemiological diseases. Quantitative assessment of mortality and/or morbidity levels may be a useful indispensable tool in management and future decision-making. Further studies are crucial to access the absolute number of deaths since using the ratio record alone may not reflect the real progression of the disease. For instance, the annual increase in the number of cancer patients may counterbalance the efforts to control deaths and the actual yearly mortality number could be a rising challenge to the efforts of cancer healthcare sector due to exacerbation of cancer epidemiology. In addition, the accuracy of the investigation and analysis is largely dependent on the Impartialness and comprehensiveness of the database. However, cancer researchers could adopt similar techniques to track different types of malignancies in their countries over time to access the changes in the epidemiological pattern and the effectiveness of the control measures. The application of SPC methodologies would be encouraged to be projected in future studies of other diseases such as Coronavirus disease (COVID-19) global cases.

Supplementary Files

The Supplementary Material for this article can be found online at: <https://doi.org/10.36462/H.BioSci.20208>

Supplementary Table S1: List of countries whose data have been used in this study and their corresponding abbreviation.

References

- Schmidt H. Chronic disease prevention and health promotion. In *Public health ethics: Cases spanning the globe 2016* (pp. 137-176). Springer, Cham.
- Boufford JI, Cassel CK, Bender KW. The future of the public's health in the 21st century. *An Institute of Medicine Report*. 2003.
- Nagai H, Kim YH. Cancer prevention from the perspective of global cancer burden patterns. *Journal of thoracic disease*. 2017 Mar;9(3):448.
- Torre L, Islami F, Siegel R, Ward E, Jemal A. Global Cancer in Women: Burden and Trends. *Cancer Epidemiology Biomarkers & Prevention*. 2017;26(4):444-457.
- Donaldson M, Lohr K. *Health data in the information age*. Washington, D.C.: National Academy Press; 1994.
- Krosnick JA, Presser S, Fealing KH, Ruggles S, Vannette DL. The future of survey research: Challenges and opportunities. The National Science Foundation Advisory Committee for the Social, Behavioral and Economic Sciences Subcommittee on Advancing SBE Survey Research. 2015 May: 1-5.
- Barlow RE, Irony TZ. Foundations of statistical quality control. *Lecture Notes-Monograph Series*. 1992 Jan 1:99-112.
- Bergman B. Conceptualistic pragmatism: a framework for Bayesian analysis?. *IIE Transactions*. 2008 Nov 7;41(1):86-93.
- Olson LW. The application of statistical process control in non-manufacturing activities. *AIR FORCE INST OF TECH WRIGHT-PATTERSON AFB OH*; 1988 Jan.
- Thor J, Lundberg J, Ask J, Olsson J, Carli C, Härenstam KP, Brommels M. Application of statistical process control in healthcare improvement: systematic review. *BMJ Quality & Safety*. 2007 Oct 1;16(5):387-99.
- Morrison LW. The use of control charts to interpret environmental monitoring data. *Natural Areas Journal*. 2008 Jan;28(1):66-73.
- Baker AW, Haridy S, Salem J, Ilieş I, Ergai AO, Samareh A, Andrianas N, Benneyan JC, Sexton DJ, Anderson DJ. Performance of statistical process control methods for regional surgical site infection surveillance: a 10-year multicentre pilot study. *BMJ Qual Saf*. 2018 Aug 1;27(8):600-10.
- Eissa ME. Determination of the microbiological quality of feed city water to pharmaceutical facility: distribution study and statistical analysis. *Athens J Sci*. 2017;4(2):143-60.
- Eissa M. A Long-Term Impact Study of Bacterial Outbreak Using Control Chart-Risk Assessment Combination. *Worldwide Medicine*. 2019;1(4):117-122.
- Rashed ER, Eissa ME. Global assessment of morbidity and mortality pattern of CoVID-19: Descriptive statistics overview. *Iberoamerican Journal of Medicine*. 2020 Apr 8;2(2):68-72.
- El-Attar IA. Cancer databases in the Arab world. *Ethn Dis*. 2005;15(1):1-3.
- Chen J, Pee D, Ayyagari R, Graubard B, Schairer C, Byrne C, Benichou J, Gail MH. Projecting absolute invasive breast cancer risk in white women with a model that includes mammographic density. *Journal of the National Cancer Institute*. 2006 Sep 6;98(17):1215-26.
- Motulsky H. *Analyzing data with graphPad prism*. San Diego: GraphPad Software Inc.; 1999.
- XLSTAT. *Statistical Software for Excel* [Internet]. XLSTAT, Your data analysis solution. 2014 [cited 28 May 2020]. Available from: <http://www.xlstat.com/en/download.html>
- Minitab 17. *Getting Started with Minitab 17* [Internet]. Minitab.com. 2016 [cited 28 May 2020]. Available from: https://www.minitab.com/uploadedFiles/Documents/getting-started/Minitab17_GettingStarted-en.pdf
- Patil KN, Kaushik SC, Garg SN. Performance prediction and assessment of energy conservation potential for a

- light pipe system in Indian composite climate of New Delhi. *Journal of Solar Energy Engineering*. 2018 Oct 1;140(5): 051012 (9 pages).
22. McGonagle J, Pilling G, Tembo V, Chumbley A, Ross E, Khim J. Gaussian Mixture Model | Brilliant Math & Science Wiki [Internet]. Brilliant.org. 2020 [cited 29 May 2020]. Available from: <https://brilliant.org/wiki/gaussian-mixture-model/>
 23. Kenett RS, Zacks S, Amberti D. *Modern Industrial Statistics: with applications in R, MINITAB and JMP*. John Wiley & Sons; 2013 Dec 16.
 24. Shanmugam R, Chattamvelli R. *Statistics for scientists and engineers*. John Wiley & Sons, Incorporated; 2015 Jul 27. <https://doi.org/10.1002/9781119047063.ch4>
 25. Banas T. Differences in Arithmetic & Geometric Mean [Internet]. Sciencing. 2017 [cited 29 May 2020]. Available from: <https://sciencing.com/differences-arithmetic-geometric-mean-6009565.html>
 26. Steele J. *The Cauchy-Schwarz master class*. Cambridge: Cambridge University Press; 2010.
 27. McDonald JH. *Handbook of biological statistics*. Baltimore, MD: sparky house publishing; 2014 Aug.
 28. Weaver KF, Morales VC, Dunn SL, Godde K, Weaver PF. *An introduction to statistical analysis in research: with applications in the biological and life sciences*. John Wiley & Sons; 2017 Sep 5.
 29. 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 [cited 30 May 2020]. Available from: <https://www.sciencedaily.com/releases/2020/01/200108074809.htm>
 30. Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*. 1977 Sep;39(1):1-22.
 31. Nagai H, Kim YH. Cancer prevention from the perspective of global cancer burden patterns. *Journal of thoracic disease*. 2017 Mar;9(3):448.
 32. Western Electric Company. *Statistical quality control handbook*. Western Electric Company; 1956.
 33. Siegel RL, Miller KD, Jemal A. *Cancer statistics, 2020*. CA: A Cancer Journal for Clinicians. 2020 Jan; 70(1):7-30.