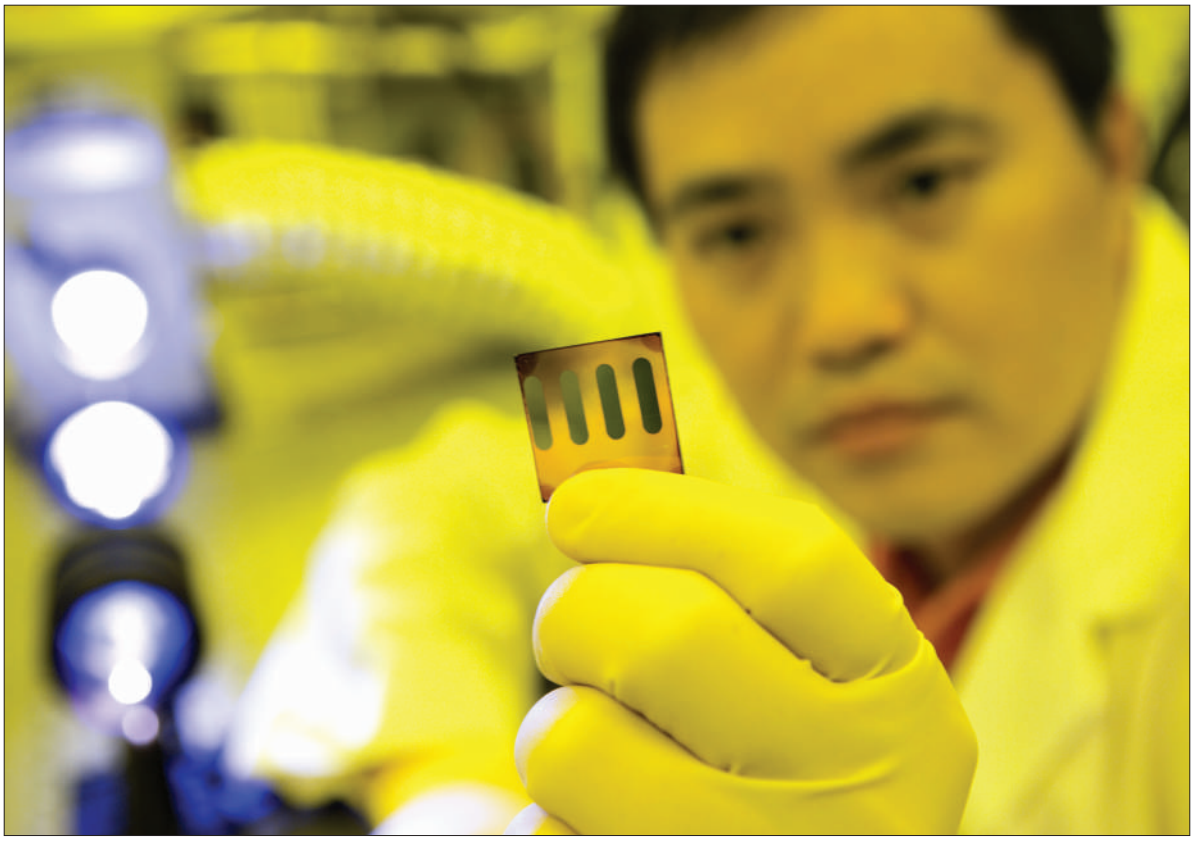


Singapore Journal of

Scientific Research

ISSN: 2010-006x

science alert

http://scialert.net/sjsr

OPEN ACCESS Singapore Journal of Scientific Research

ISSN 2010-006x
DOI: 10.3923/sjsres.2020.336.341

Technical Report
Unique Descriptive Analysis Approach of Ebola Outbreak from WHO Database Using Statistical Process Control

[1]Engy Refaat Rashed and [2]Mostafa Essam Eissa

[1]Department of Drug Radiation Research, National Centre for Radiation Research and Technology, Atomic Energy Authority, P.O. Box 29, Nasr, Cairo, P.O. 11787, Egypt
[2]Department of Microbiology and Immunology, Cairo University, Cairo, Egypt

Abstract

Background and Objective: Ebola outbreak is a lethal devastating viral epidemiological disease that had impacted the West Africa region recently, particularly Guinea, Liberia and Sierra Leone. Ebola Virus Disease (EVD) has been comprehensively reported by World Health Organization (WHO). However, data analysis using Statistical Process Control (SPC) would provide a different unique view for disease monitoring, trending and modeling for the outbreaks to set and derive appropriate preventive measures and identifying major characteristics or patterns of the epidemic In addition, it would provide a mean for adoption of a novel way for quantitative risk analysis. **Materials and Methods:** Data processing and stratification from Comma-Separated Values (CSV) file database about Ebola outbreaks in its official website were performed using Excel sheet functions. Modified results were then subjected to distribution modeling using Excel built-in program known as XLSTAT. While statistical analysis tools were applied using SPC techniques of Minitab. **Results:** The EVD outbreak illnesses showed negative binomial distribution pattern with more than half of the impacted populations were from all age groups rather than from specific ones and more than one-third of the affected individuals were from sierra leone alone. Process behavior charts showed the pattern and trend of the outbreak events. The attribute chart demonstrated the magnitude of illnesses reported for outbreaks and rare events or G chart illustrated the occurrence rate for the events. **Conclusion:** The SPC methodologies are useful quantitative and unique tools that provide an important approach to study epidemiological diseases using simple, fast and inexpensive techniques. In addition, it would provide a mean for adoption of a novel way for quantitative risk analysis.

Key words: Ebola outbreak, statistical process control, epidemics, viral disease

Citation: Rashed E.R. and M.E. Eissa, 2020. Unique descriptive analysis approach of ebola outbreak from WHO database using statistical process control. Singapore J. Sci. Res., 10: 336-341.

Corresponding Author: Mostafa Essam Eissa, Department of Microbiology and Immunology, Cairo University, Cairo, Egypt

Copyright: © 2020 Engy Refaat Rashed and Mostafa Essam 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.

Competing Interest: The authors have declared that no competing interest exists.

Data Availability: All relevant data are within the paper and its supporting information files.

INTRODUCTION

Ebola virus epidemics in West Africa constitute fatal devastating outbreaks in Africa, especially for Guinea, Liberia and Sierra Leone[1]. Despite extensive efforts made by international health organizations in the monitoring and control of Ebola Virus Disease (EVD), there are many traceability data missing in the records of the World Health Organization (WHO) which may reflect some important gaps in detection, monitoring and control of the disease[2]. However, EVD disease has received greater attention recently due to recent catastrophic outbreaks that impacted the public health and economy of many countries through the last decades.

The EVD is actually a dangerous pathogenic agent with high mortality rates which was involved in multiple epidemics in the African continent particularly recently[3]. Comprehensive records from the internal organizations have been provided and rendered available to the public about EVD epidemics. However, it is the effectiveness of data mining and processing that could provide a quantitative mean for the analysis of the outbreaks and determining their pattern and behavior in addition to the assessment of the encountered risks.

Hence, Statistical Process Control (SPC) techniques emerge as useful statistical analysis methodologies that were initially developed for monitoring of the industrial processes and product quality[4]. Nevertheless, their usefulness could be demonstrated for the inspection properties study in other non-industrial fields such as microbiological quality of environment and Surgical Site Infection (SSI) rate[5,6]. The use of these tools might show usefulness for epidemiological analysis for quantitative description, analysis and modeling of outbreaks such EVD endemics in some African countries.

Accordingly, the following descriptive-analytical study was aimed to investigate the application of SPC in the statistical analysis of the virus disease epidemics using Ebola as a working example from the WHO dataset.

MATERIALS AND METHODS

Excel sheet derived from Comma-Separated Values (CSV) dataset called "CASES, DEATHS" could be obtained for disease outbreaks e.g., Ebola from the official web sites of the international organizations such as WHO official internet page https://apps.who.int/gho/data/view.ebola-sitrep.ebola-summary-latest?lang=en. Data reported for three EVD infested West African countries viz. Guinea, Liberia and Sierra Leone would focus on the outbreak sections from May 2015-May 2016. The produced Excel sheet was filtered processed and stratified using internal functions of Microsoft Office.

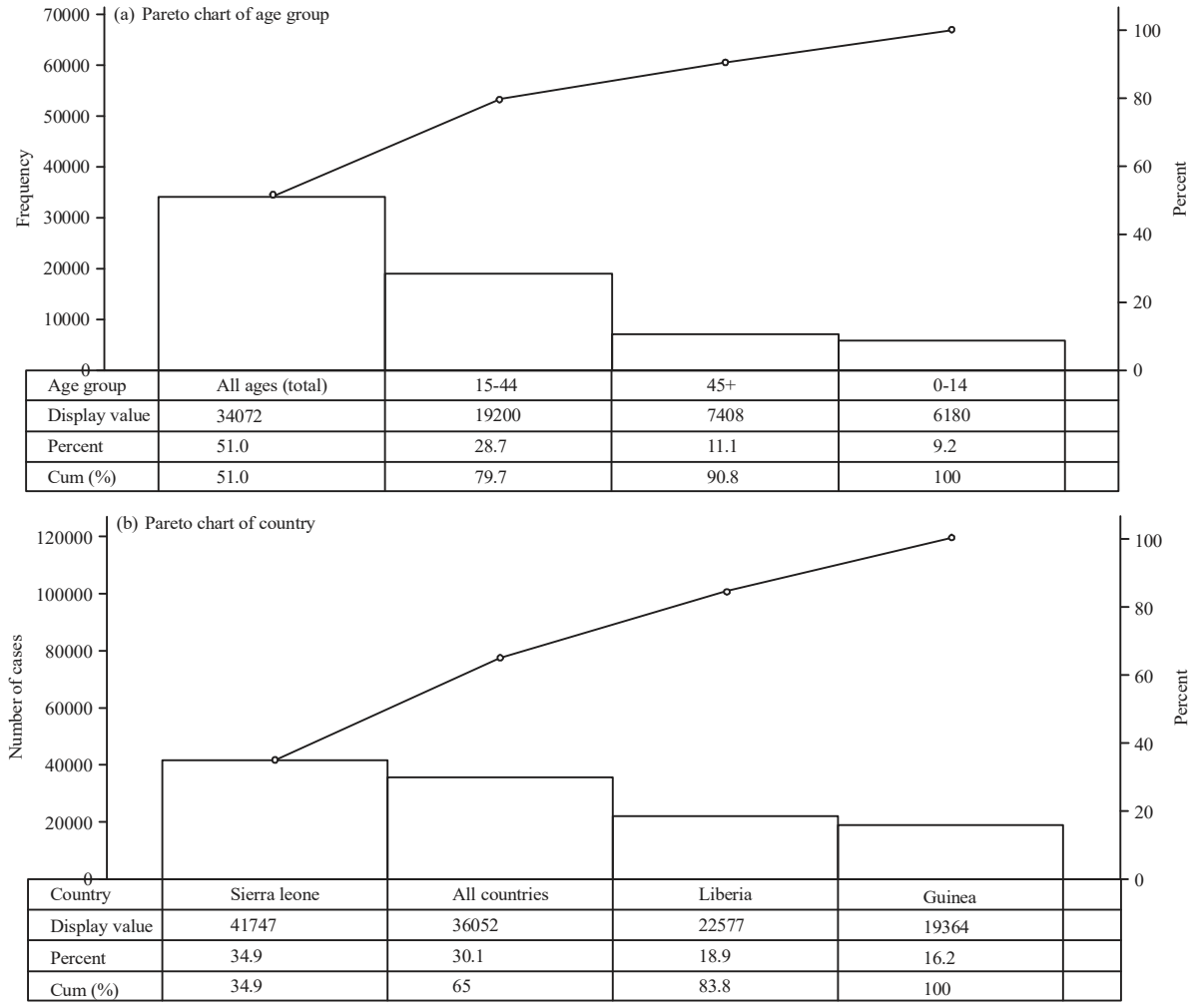
Output results were further processed using two platforms. Important tools that were involved were Pareto diagram, distribution fitting and process-behavior (Shewhart, trending or control) charts. The systematic approach of the study was adopted as previously demonstrated in other investigations previously[7,8]. First, Minitab® 17.1.0 software was used for Pareto plot and construction of process behavior charts namely attribute trending (Laney-modified for correction of non-conforming to the assumed distribution) and rare events (G) charts for both numeric cases numbers and the reported outbreak events occurrence, respectively. Second, Excel integrated built-in XLSTAT 2014.5.03 program was used for distribution fitting modeling of the displayed values of the record.

RESULTS AND DISCUSSION

The present report provided a worked example case for the application of SPC tools in viral outbreak incidents. Achieving suitable control measures for EVD outbreaks were difficult in the affected wide geographical distribution area of the epidemic for the stricken countries in West Africa during a period[9] from 2014-2016. Several Pareto analysis showed that more than half of EVD outbreaks affected all population ages followed by age group from 15-44 years with a total contribution of 80% of the cases supported by previous observations (Fig. 1a) and more than the third of illnesses occurred in Sierra Leone alone and more than 30% of the cases were affecting the three countries (Fig. 1b) simultaneously[9]. Complementarily, 2 types of control charts were used for outbreaks monitoring viz., attribute (Laney-modified) and rare event (G) control charts. The G chart demonstrated five intermittent episodes of the recorded outbreak bursts which occurred harmoniously together separated by placid weeks suggesting that EVD endemics occur as consecutive waves (Fig. 1c). It should be noted that similar observations about this pattern were observed by Coltart *et al.*[10]. The probability of occurrence of the EVD outbreak during the time period of the major epidemic was 0.161. Unfortunately, not all outbreak cases showed a full record of the number of the affected individuals which is evident in the attribute chart of the number of cases per outbreak. This kind of limitation has been analyzed by Ilesanmi *et al.*[11]. Thus, parameters obtained from the trending chart such as upper and lower control limits (UCL and LCL) were partial estimates from the available

data (Fig. 1d). This gap may stimulate awareness and perception of the healthcare providers and observers for the importance of setting appropriate measures for timely data collections before losing the ability of the traceability (Fig. 1).

Despite the limitation of some records in the dataset for the impacted countries (Fig. 2a) which affect this study, a problem of concern that has been reported by Awini *et al.*[12], estimation of the outbreak illnesses distribution through data modeling could be studied using distribution fitting analysis. A 2-parameter negative binomial pattern of data was observed for the Ebola outbreak of the impacted countries as could be found in Fig. 2b. The hypothetical testing of distribution suitability and parameters (k and p) determinations±Standard Errors (SE) along with brief statistical analysis were shown appending the observed vs. theoretical distribution values in Table 1. This distribution pattern of outbreaks was also modeled and reported in previous studies by Tuite and Fisman[13], Althaus[14] and Toth *et al.*[15]. The availability of the commercial statistical software platform has rendered this SPC analysis applicable at a convenient time with high precision. Finally, the red dots points in the trending charts are an indication of the excursions in the monitored process either time intervals in G chart or numeric values in the Laney attribute chart which can provide a focus investigation group of an opportunity for learning lessons in the improvement of the EVD control measures[16-18]. The present worked example is evidence for the usefulness of 6 Sigma tools as an approach to assess outbreaks in a unique statistical means from which a quantitative risk analysis could be derived.



Age group	All ages (total)	15-44	45+	0-14
Display value	34072	19200	7408	6180
Percent	51.0	28.7	11.1	9.2
Cum (%)	51.0	79.7	90.8	100

Country	Sierra leone	All countries	Liberia	Guinea
Display value	41747	36052	22577	19364
Percent	34.9	30.1	18.9	16.2
Cum (%)	34.9	65	83.8	100

Fig. 1(a-d): Continue

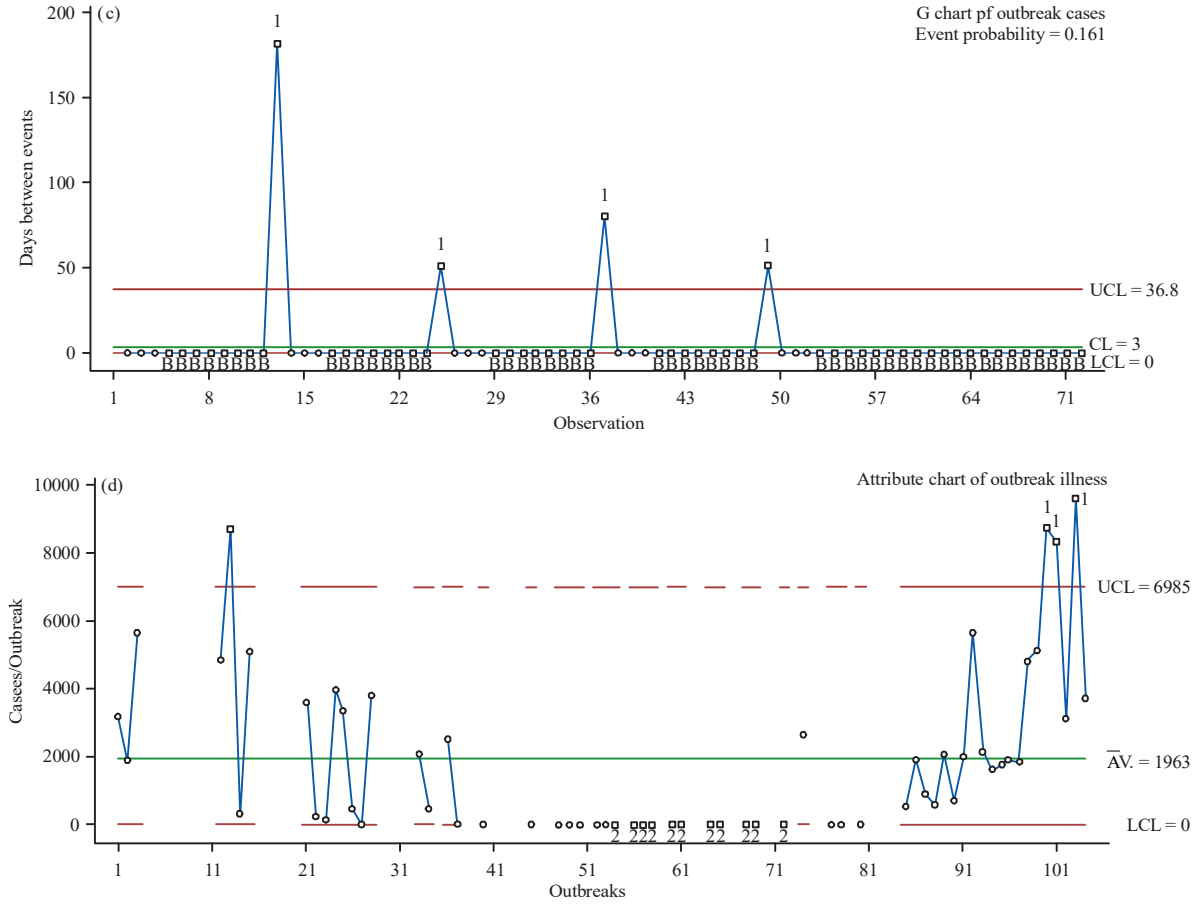


Fig. 1(a-d): Pareto charts of (a) Affected age groups, (b) Countries, (c) Trending chart of recorded interval periods and (d) Magnitudes of cases

Table 1: Statistical interpretation of the distribution pattern of EVD outbreak incidences in the recorded series

Estimated parameters	Values	Standard error
Parameters		
k	0.2237	0.0326
p	14039.4891	4171.8291
Statistics estimated on the input data and computed using the estimated parameters of the Negative binomial (2) distribution		
Parameters	Data	Statistic
Mean	3141.0149	3141.0217
Variance	23671450.4089	44101480.3331
Skewness (Pearson)	2.7140	4.2283
Kurtosis (Pearson)	9.6609	26.8183
Chi-square test	Data	Test interpretation
Chi-square (Observed value)	43.9922	H0: The sample follows a Negative binomial (2) distribution
Chi-square (Critical value)	57.8576	Ha: The sample does not follow a Negative binomial (2) distribution
DF	27	As the computed p-value is greater than the significance level
p-value	0.0207	alpha = 0.0005, one cannot reject the null hypothesis H0. The risk
Alpha	0.0005	to reject the null hypothesis H0 while it is true is 2.07%

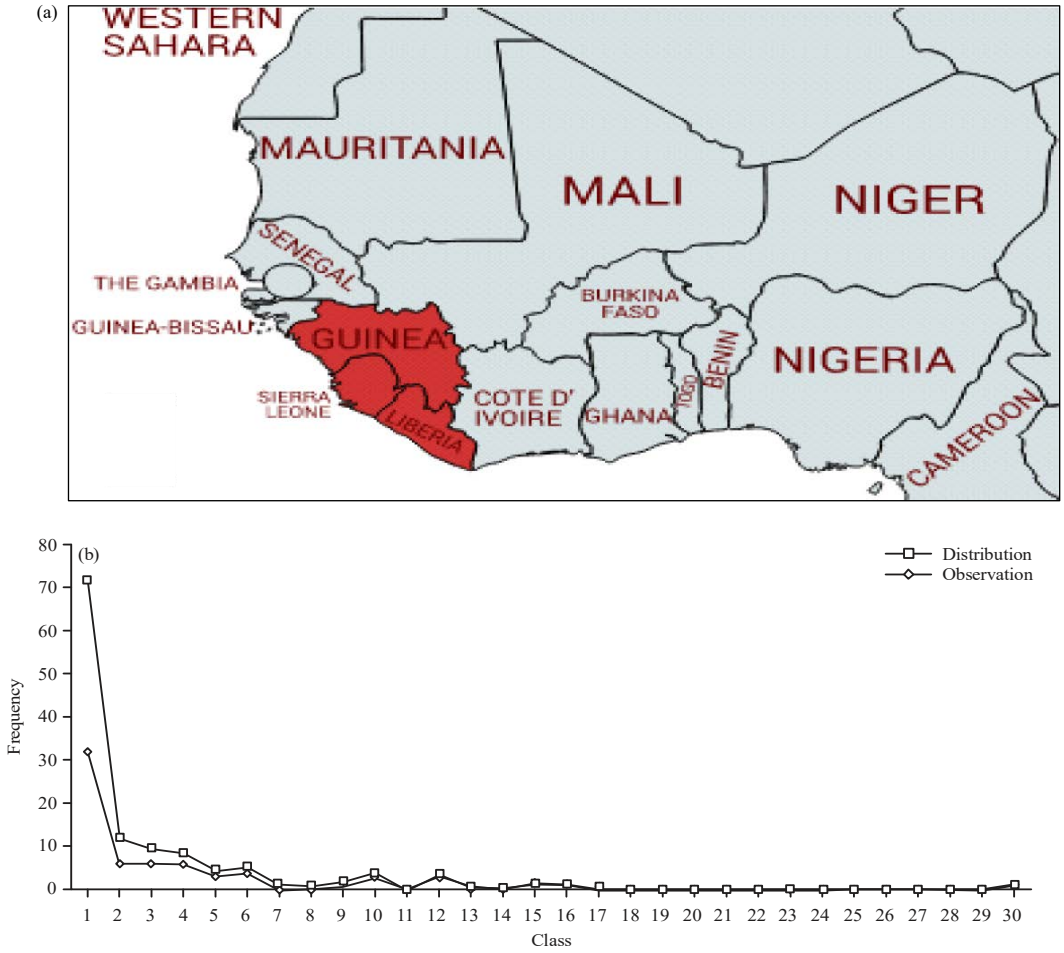


Fig. 2(a-b): West African region of (a) Most affected countries and (b) Distribution modeling showing real and theoretical distribution

Class refers to the ascending rank numbering for grouping of the number of cases between upper and lower bound values increasing by a fixed increment

CONCLUSION

The SPC techniques are useful means for outbreak study as in the current descriptive analysis of EVD. They provide insight into the pattern of the epidemic diseases with quantitative results. These numerical data may be used in the future as a method for adopting a unique way for risk assessment and analysis to monitor the kinetics of epidemic diseases locally or globally. Nevertheless, the accuracy of the derived information is limited by the level of neatness for the reported dataset and the degree of comprehensiveness of the records.

SIGNIFICANCE STATEMENT

This study discovers the possible application of SPC methodologies in the study of viral outbreaks taking EVD as an example that can be beneficial for unique monitoring, trending and modeling of the epidemics. This study will help the researcher to adopt a novel mean of quantitative risk and hazard assessment that many researchers can be simply and timely implement for study of the outbreaks and comparison for outbreaks. Thus, a new perspective can be deduced in the outbreak monitoring, control and prevention and possibly other theories may be arrived at.

REFERENCES

1. Omoleke, S.A., I. Mohammed and Y. Saidu, 2016. Ebola viral disease in West Africa: A threat to global health, economy and political stability. J. Public Health Afr., 7: 27-27.
2. Senga, M., A. Koi, L. Moses, N. Wauquier and P. Barboza *et al.*, 2017. Contact tracing performance during the Ebola virus disease outbreak in Kenema district, Sierra Leone. Philos. Trans. R. Soc. B: Biol. Sci., Vol. 372. 10.1098/rstb.2016.0300.

3. Malvy, D., A.K. McElroy, H. de Clerck, S. Gunther and J. van Griensven, 2019. Ebola virus disease. Lancet, 393: 936-948.
4. Montgomery, D.C., C.L. Jennings and M.E. Pfund, 2010. Managing, Controlling and Improving Quality. John Wiley & Sons, Hoboken, NJ., USA., ISBN-13: 978-0471697916, pp: 2-3.
5. Eissa, M.E., A.M. Mahmoud and A.S. Nouby, 2015. Evaluation and failure risk of microbiological air quality in production area of pharmaceutical plant. RGUHS J. Pharmaceut. Sci., 5: 155-166.
6. Eissa, M.E., 2019. Application of control charts in monitoring of surgical site infection trending records using statistical software. Asian J. Applied Sci., 12: 76-84.
7. Eissa, M.E.A.M., 2019. Surveillance of *Cyclospora cayetanensis* epidemics in USA from long-term national outbreaks reporting system-based monitoring: An Observational study using statistical process control methodologies. Iberoamerican J. Med., 2: 4-9.
8. Eissa, M., 2019. Avian schistosomes impact on public health: A long-term disease observation. Ann. Med. Res., 26: 1723-1724.
9. McNamara, L.A., I.J. Schafer, L.D. Nolen, Y. Gorina and J.T. Redd *et al*., 2016. Ebola surveillance—Guinea, Liberia and Sierra Leone. MMWR Suppl., 65: 35-43.
10. Coltart, C.E., B. Lindsey, I. Ghinai, A.M. Johnson and D.L. Heymann, 2017. The Ebola outbreak, 2013-2016: Old lessons for new epidemics. Philos. Trans. R. Soc. B: Biol. Sci., Vol. 372. 10.1098/rstb.2016.0297.
11. Ilesanmi, O.S., O. Fawole, P. Nguku, A. Oladimeji and O. Nwenyi, 2019. Evaluation of Ebola virus disease surveillance system in Tonkolili district, Sierra Leone. Pan Afr. Med. J., Vol. 32. 10.11604/pamj.supp.2019.32.1.14434.
12. Awini, E.A., J.H. Bonney, J.A. Frimpong, W.K. Ampofo and K.A. Koram, 2017. Information gaps in surveillance data and effects on the Ghanaian response to the Ebola outbreak in West Africa. Ghana Med. J., 51: 115-119.
13. Tuite, A.R. and D.N. Fisman, 2018. The IDEA model: A single equation approach to the Ebola forecasting challenge. Epidemics, 22: 71-77.
14. Althaus, C.L., 2015. Ebola superspreading. Lancet Infect. Dis., 15: 507-508.
15. Toth, D.J., A.V. Gundlapalli, K. Khader, W.B. Pettey, M.A. Rubin, F.R. Adler and M.H. Samore, 2015. Estimates of outbreak risk from new introductions of Ebola with immediate and delayed transmission control. Emerg. Infect. Dis., 21: 1402-1408.
16. Levine, D.M., 2006. Statistics for Six Sigma Green Belts with Minitab and JMP. Prentice-Hall Inc., Upper Saddle River, NJ., USA., ISBN-13: 978-0137017126, Pages: 400.
17. Henderson, G.B., 2011. Six Sigma Quality Improvement with Minitab. 2nd Edn., John Wiley and Sons Ltd., USA., ISBN-13: 9781119976189, Pages: 528.
18. Bass, I., 2007. Six Sigma Statistics with Excel and Minitab. McGraw-Hill, New York, USA., ISBN-13: 9780071542685, Pages: 374.