Journal of
Food Chemistry & Nanotechnology

https://doi.org/10.17756/jfcn.2019-066

Review Article

Open Access

Statistical Analysis Review and Lessons Learned from Recent Outbreak Trends of Highest Population Density States in USA: Massachusetts, New Jersey and Rhode Island

Mostafa Essam Ahmed Eissa*

Microbiology and Immunology Department, Cairo University, Cairo, Egypt

***Correspondence to:**
Mostafa Essam Ahmed Eissa
Independent PhD Researcher and Candidate
Microbiology and Immunology Department
Faculty of Pharmacy, Cairo University
Cairo, Egypt
Tel: +201006154853
E-mail: mostafaessameissa@yahoo.com

Received: March 23, 2019
Accepted: March 29, 2019
Published: April 04, 2019

Citation: Eissa MEA. 2019. Statistical Analysis Review and Lessons Learned from Recent Outbreak Trends of Highest Population Density States in USA: Massachusetts, New Jersey and Rhode Island. *J Food Chem Nonotechnol* 5(1): 8-19.

Published by United Scientific Group

Abstract

The main objective from the current review study is to provide long-term quantitative study of the outbreak pattern and trend through 20 years of record and observation by Center for Disease Control and Prevention (CDC) through National Outbreak Reporting System (NORS) database web-based platform for the most heavily populated states in USA. Thorough analysis was conducted using statistical process control (SPC) software packages after data processing and segregation. Non-normally distributed data for outbreak illness cases tend to show Log-normal or Weibull (3) distribution while rare event mortality cases tend to show Negative Binomial pattern. Box plot diagram and control charts show that outbreaks pattern tends to be low in the number of the affected population with excursions in relatively fewer cases in the number of ill individuals. Excursions in outbreak rates of Massachusetts, New Jersey and Rhode Island are approximately 8, 7 and 12%, respectively. However, Massachusetts mean outbreak trend is significantly higher than that of New Jersey and not significantly different from Rhode Island at $\alpha = 0.05$. Interestingly, the threshold of out-of-control outbreaks for New Jersey is about 80 illnesses per outbreak while Massachusetts and Rhode Island are 62 and 63, respectively. Norovirus is responsible for most cases of illnesses and deaths in the three states and most outbreak cases occurred in assisted living facilities and long-term care homes among other residences which include elderly and health-compromised individuals. Outbreaks are a persistent health problem in the developed countries with intermittent excursions of illness cases.

Keywords

Log-normal, Negative binomial, Norovirus, NORS, SPC, Weibull (3)

Introduction

Epidemiological diseases are global health problem issue that impact human life and economy which may be catastrophic and devastating in extreme situations [1, 2]. Outbreaks have been reported through human history on earth with severe causalities that impacted civilizations. Many outbreak diseases are considered epidemics which are confined within specific geographical regions such as Ebola in West Africa which has caused thousands of mortality cases. On the other hand, other outbreaks are pandemic such as smallpox which caused during its 12000 years of its existence an estimated of 400 ± 100 million people deaths through human life on the earth. Table 1 shows chronologically the major historical pandemic outbreaks that have been recorded [3].

National outbreak reporting system (NORS)

Center for Disease Control and Prevention (CDC) has provided a

Table 1: Chronological record of the major pandemic outbreaks [3].

Outbreak	Pandemic Etiology	Time	Mortality
Antonine plague	Not determined	165	5000000
Plague of Justinian	Bubonic Plague	541-542	25000000
The Black Death	Bubonic Plague	1346-1353	75000000-200000000
Third Cholera pandemic	Cholera	1852-1860	1000000
Flu pandemic[a]	Influenza	1889-1890	1000000
Sixth cholera pandemic	Cholera	1910-1911	>800000
Flu pandemic	Influenza	1918	20000000-50000000
Asian flu[b]	Influenza	1956-1958	2000000
Flu pandemic[c]	Influenza	1968	1000000
HIV/AIDS pandemic	HIV/AIDS	2005-2012	36000000[d]

[a]Asiatic (Russian): Influenza A virus subtype H3N8 Flu
[b]Influenza A of the H2N2 subtype
[c]The Hong Kong Flu: H3N2 strain of the Influenza A virus
[d]Since 1981

comprehensive platform for recording and monitoring outbreaks in USA which is available for public since 2009 through National Outbreak Reporting System (NORS) website [4]. NORS database delivers useful information and data segregation and stratification could be executed in different ways according to the research focus intended from "National Outbreak Public Data Tool" [5]. Data are subjected to dynamic changes and updates. Thus, currently derived outcomes are the result of up-to-date extracted and processed data. However, the study covers 20 years of monitoring from 1998 to 2017 for outbreaks through the whole country [5].

Commercial statistical software and outbreak analysis

Statistical process control (SPC) tools are crucial for data interpretation. The analysis is rendered simple, easy and time-saving using commercial statistical software packages such as Minitab® V 17.1.0, GraphPad Prism V 6.01 and XLSTAT (add-in program) V2014.5.03, according to their electronic manual [6-8]. These programs have been used in different other studies to monitor industrial and non-industrial processes and inspection characteristics [9-12].

Ill populations in outbreaks histograms

The distribution pattern of ill population per outbreak of the three most densely populated states viz. Massachusetts, New Jersey and Rhode Island show a characteristic pattern of steep spiking followed by a gradual decline in the number of outbreaks with the rising number of sick individuals. This pattern has been reported similarly previously in other outbreaks by other investigators [13, 14]. This finding is illustrated in figure 1. The right-skewed distribution indicates that each state has historically relatively abundant outbreaks that involve specific range of population number with descending probability of occurrence of outbreaks affecting greater number individuals as the number illness cases grow with very low frequency of outbreaks which involves abnormally high morbidity which stand distant from the integrated distribution pattern.

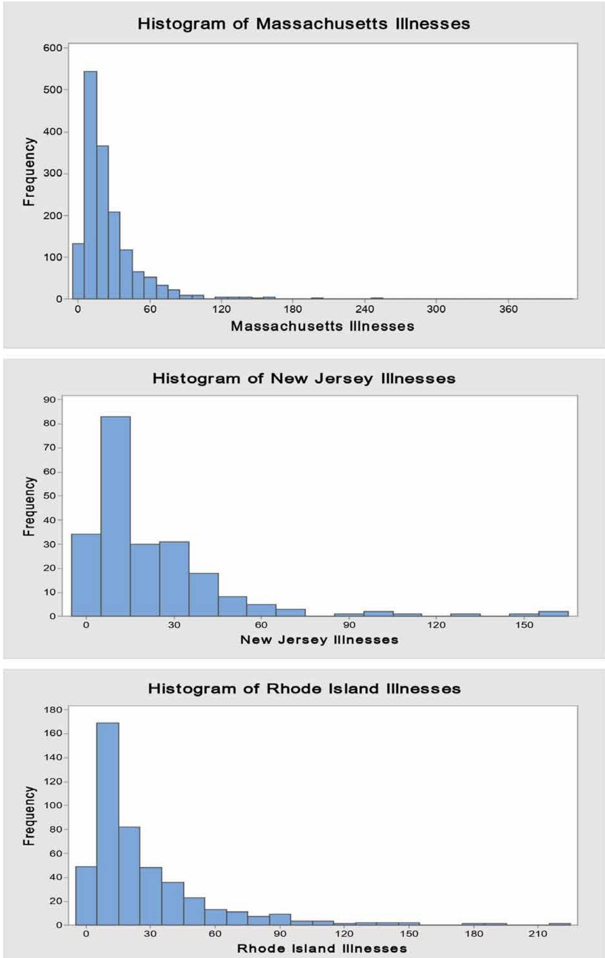


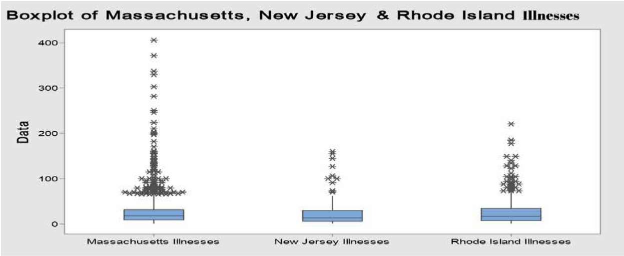
Figure 1: Histogram of ill individuals per outbreak for 20 years (1998 to 2017) in population rich states.

Mortality in outbreaks histograms

Death cases per outbreak are simpler in histogram as could be found in figure 2. Mortality rates are usually rare events with very few cases that may range from 1 to 3 usually. Mortality pattern from diseases and outbreak has been a subject of extensive study and mathematical modelling for prediction [15, 16].

Distribution pattern of outbreak illness does not meet Gaussian distribution with most of data centred toward lower illness count with many intermittent outlier outbreaks in the number of patients as could be demonstrated in table 1 by box plot diagram [17]. The non-correlated data showed rates of aberrant outbreaks illness approximately as 8, 7 and 12% for Massachusetts, New Jersey and Rhode Island, respectively. However, Massachusetts mean illness cases per outbreak is significantly higher than that of New Jersey only at $\alpha = 0.05$ using non-parametric comparison while Rhode Island falls in-between the two states. The spreading pattern of box plot

Table 2: Outliers determination by box plot, column statistics by column graph and non-parametric comparison test for outbreak record.



Method: ROUT (Q = 1.000%)	Massachusetts Illnesses	New Jersey Illnesses	Rhode Island Illnesses
Number of points Analyzed	1593	220	463
Outliers	121	16	56

Statistical Comparison (α = 0.05)					
Dunn's multiple comparisons test	Mean rank diff.	Significant?	Summary		
Massachusetts Illnesses vs. New Jersey Illnesses	152.7	Yes	**		
Massachusetts Illnesses vs. Rhode Island Illnesses	40.01	No	ns		
New Jersey Illnesses vs. Rhode Island Illnesses	-112.7	No	ns		
Test details	Mean rank 1	Mean rank 2	Mean rank diff.	n1	n2
Massachusetts Illnesses vs. New Jersey Illnesses	1161	1009	152.7	1593	220
Massachusetts Illnesses vs. Rhode Island Illnesses	1161	1121	40.01	1593	463
New Jersey Illnesses vs. Rhode Island Illnesses	1009	1121	-112.7	220	463

Descriptive Statistics			
Minimum	2.000	2.000	2.000
25% Percentile	9.000	6.000	8.000
Median	18.00	13.00	16.00
75% Percentile	31.50	29.75	34.00
Maximum	406.0	160.0	221.0
10% Percentile	5.000	3.000	4.000
90% Percentile	55.00	46.90	60.00
Std. Error of Mean	0.8321	1.722	1.374
Lower 95% CI of mean	24.73	18.62	23.52
Upper 95% CI of mean	27.99	25.41	28.92
Lower 95% CI of median	17.00	11.00	13.00
Upper 95% CI of median	19.00	16.00	18.00
Coefficient of variation	125.99%	116.04%	112.72%
Geometric mean	16.82	13.09	15.86
Lower 95% CI of geo. mean	16.06	11.40	14.45
Upper 95% CI of geo. mean	17.61	15.03	17.41
Skewness	5.015	2.841	2.610
Kurtosis	37.97	10.34	9.152

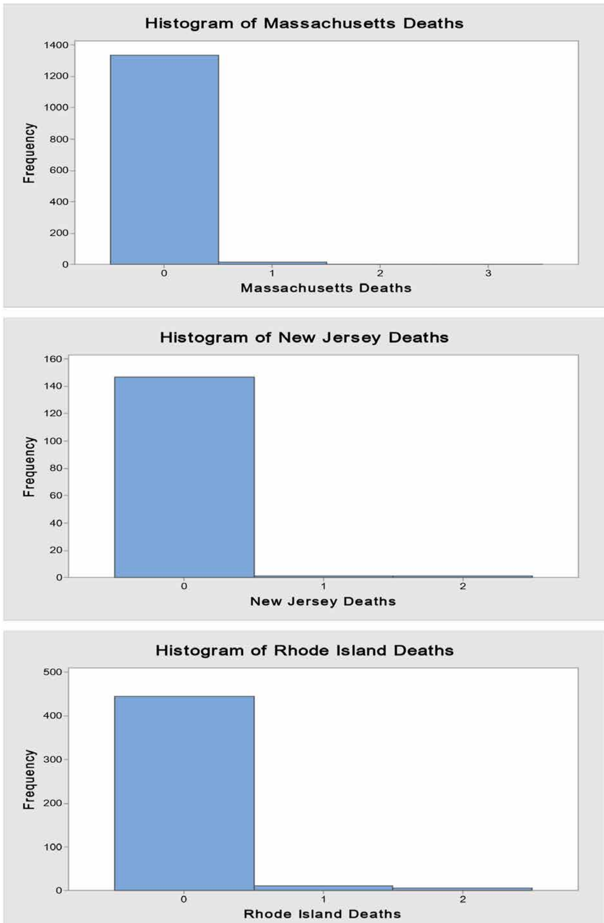


Figure 2: Histogram of mortality per outbreak for 20 years (1998 to 2017) in population rich states.

figure is detailed numerically in table 2 and showing highest dispersion for Massachusetts and lowest for New Jersey.

Distribution fitting and outbreak morbidity

Massachusetts and Rhode Island sickness cases per outbreak distributions are close to log-normal distribution while New Jersey distribution approximates Weibull-type of distribution pattern at significance level 0.05%. Several researchers have discussed diseases outbreak or epidemic distributions with similar behavior of data [13, 14, 18-20]. Figure 3 shows goodness-to-fit cumulative actual curves with the hypothetical ones. The estimated Log-normal parameters for Massachusetts and Rhode Island outbreak illnesses (μ ± Standard Error (SE), σ ± SE) are (2.8225 ± 0.0129, 0.9418 ± 0.4850) and (2.7640 ± 0.0194, 1.0236 ± 0.7812), respectively. While Weibull distribution parameters (μ ± SE, β ± SE, γ ± SE) for New Jerseyis as the following: (19.0405 ± 0.7805, 0.9316 ± 0.0409, 1.4062 ± 0.4803). These parameters are based on the current data trend of the past 20 years from 1998 to 2017 as found in database.

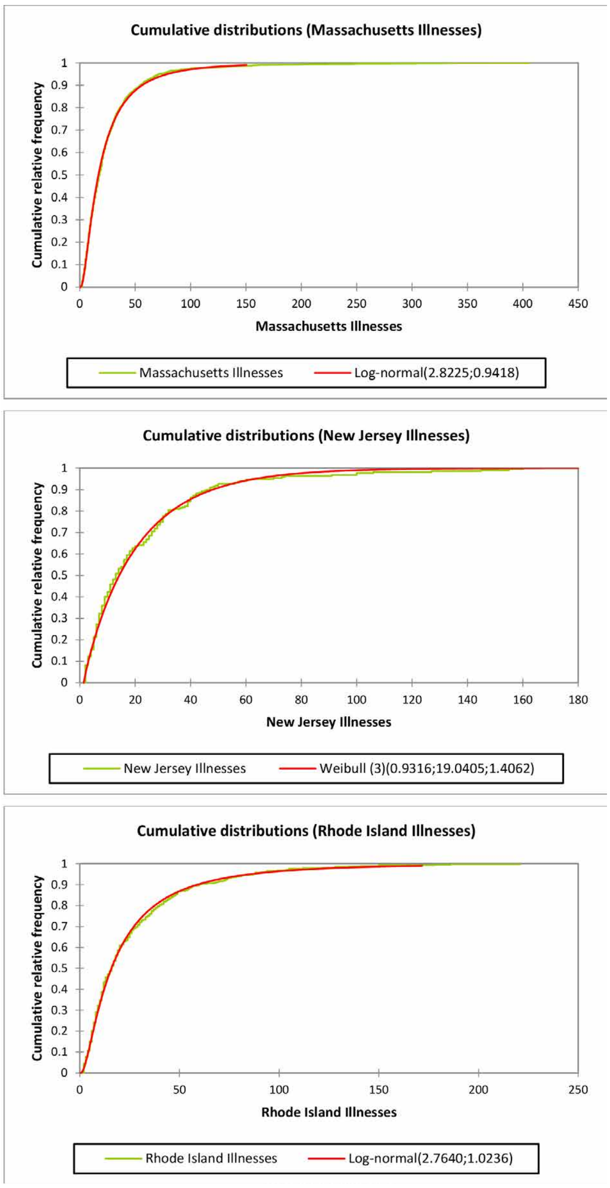


Figure 3: Cumulative distribution plot data for number of illness per outbreak of the highest population density states.

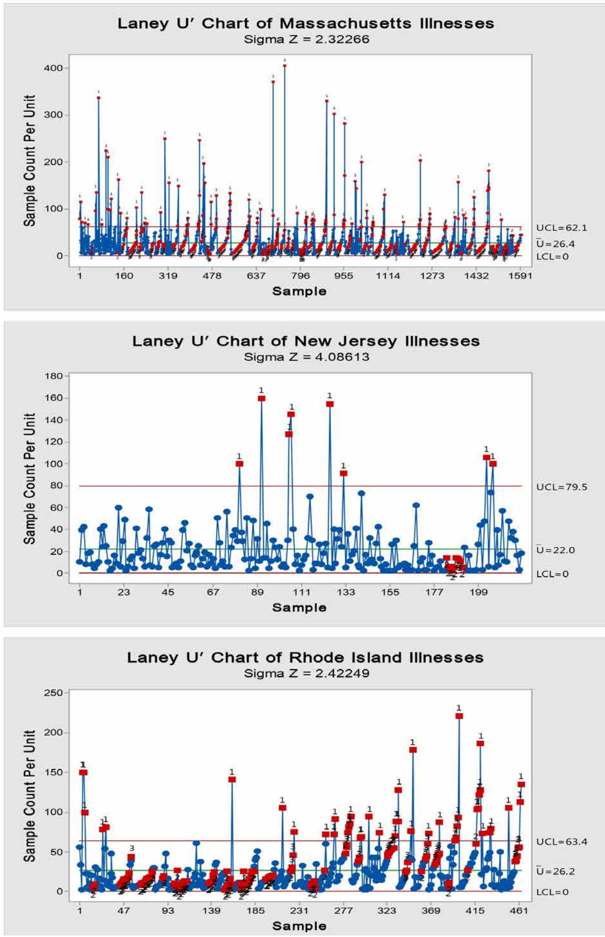


Figure 4: Attribute control charts of outbreaks during 20 years of three selected states: Massachusetts, New Jersey and Rhode Island.

Attribute process-behavior charts and Laney modification choice

Trending charts are useful in process visualization and monitoring of the inspection characteristic [20, 21]. In the present study, the x-axis of the process-behavior chart is showing the number of recorded outbreaks during 20 years of monitoring from 1998 to 2017 arranged chronologically and the y-axis is showing the number of the ill individuals per each observed outbreak. Laney correction of data dispersion is useful for data that could not fit presumed distribution of the control chart used and the degree of over or under-dispersion is indicated by σZ value on the chart [22-24]. Mean value, Upper Control Limit (UCL) and alarming out-of-control outbreak illness numbers (indicated by red dots) are shown for each state in figure 4. Lower Control Limit (LCL) is zero for the three states. However, the greatest concern is devoted to the outbreak threshold i.e. UCL where excursions in the number of sick populations is the signaling for warning alarm. While as the number of outbreak cases in the lower part of the chart are going to zero are desirable. Interestingly, the threshold of out-of-control outbreaks for New Jersey is about 80 illnesses per outbreak while Massachusetts and Rhode Island are 62 and 63, respectively. Despite it showed the lowest outbreaks rate for 20 years and average number of illness cases per outbreak 22 versus about 26 individuals in the other two states. Out-of-control point marked by "1" are due to extraneous abnormal events that influenced outbreaks which are outside CL. Aberrant points denoted by "2" are showing outbreaks that are progressively changing in the average value of sickness cases. On the other

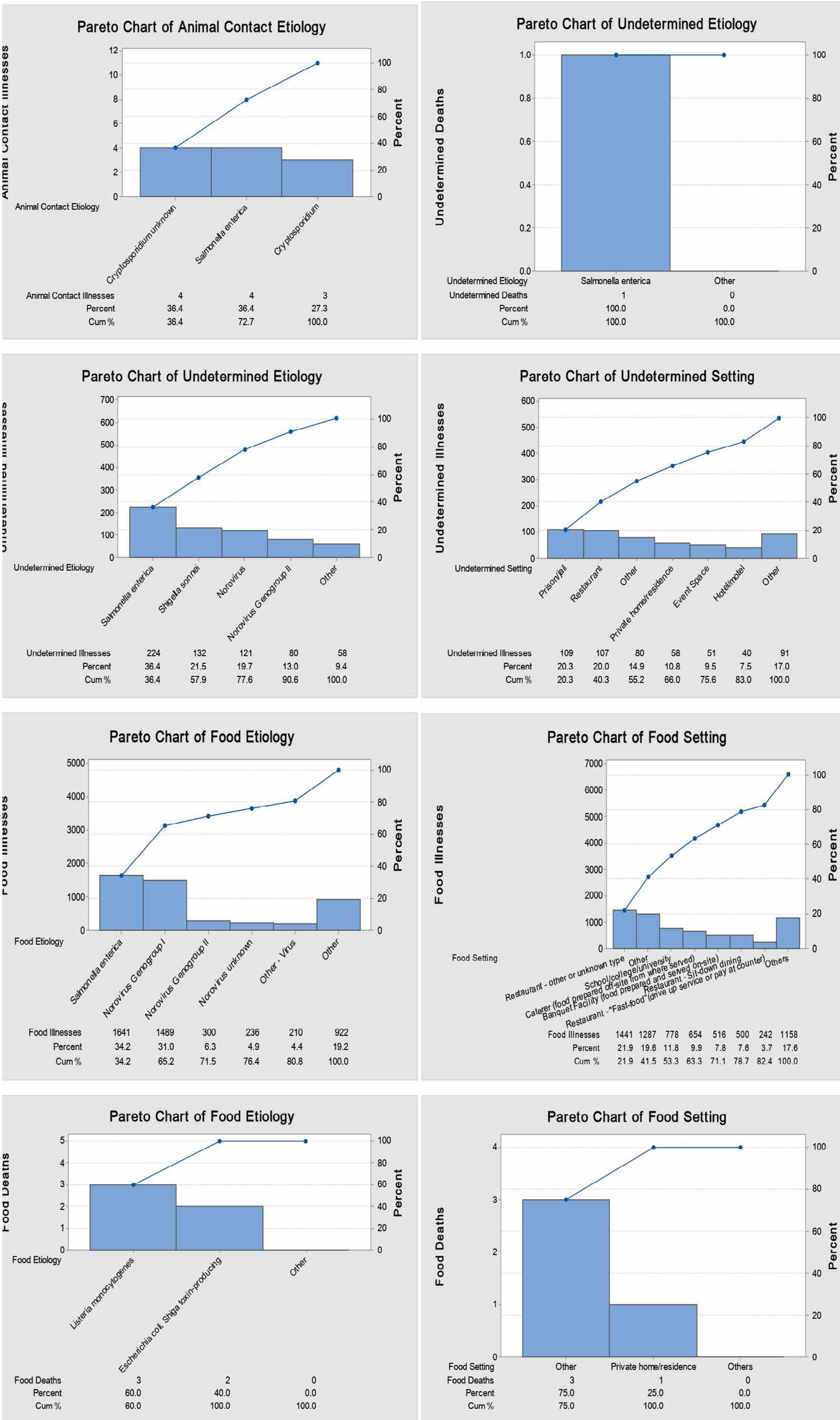


Figure 5: Pareto diagram of animal contact, food and other undetermined modes for outbreaks in Massachusetts.

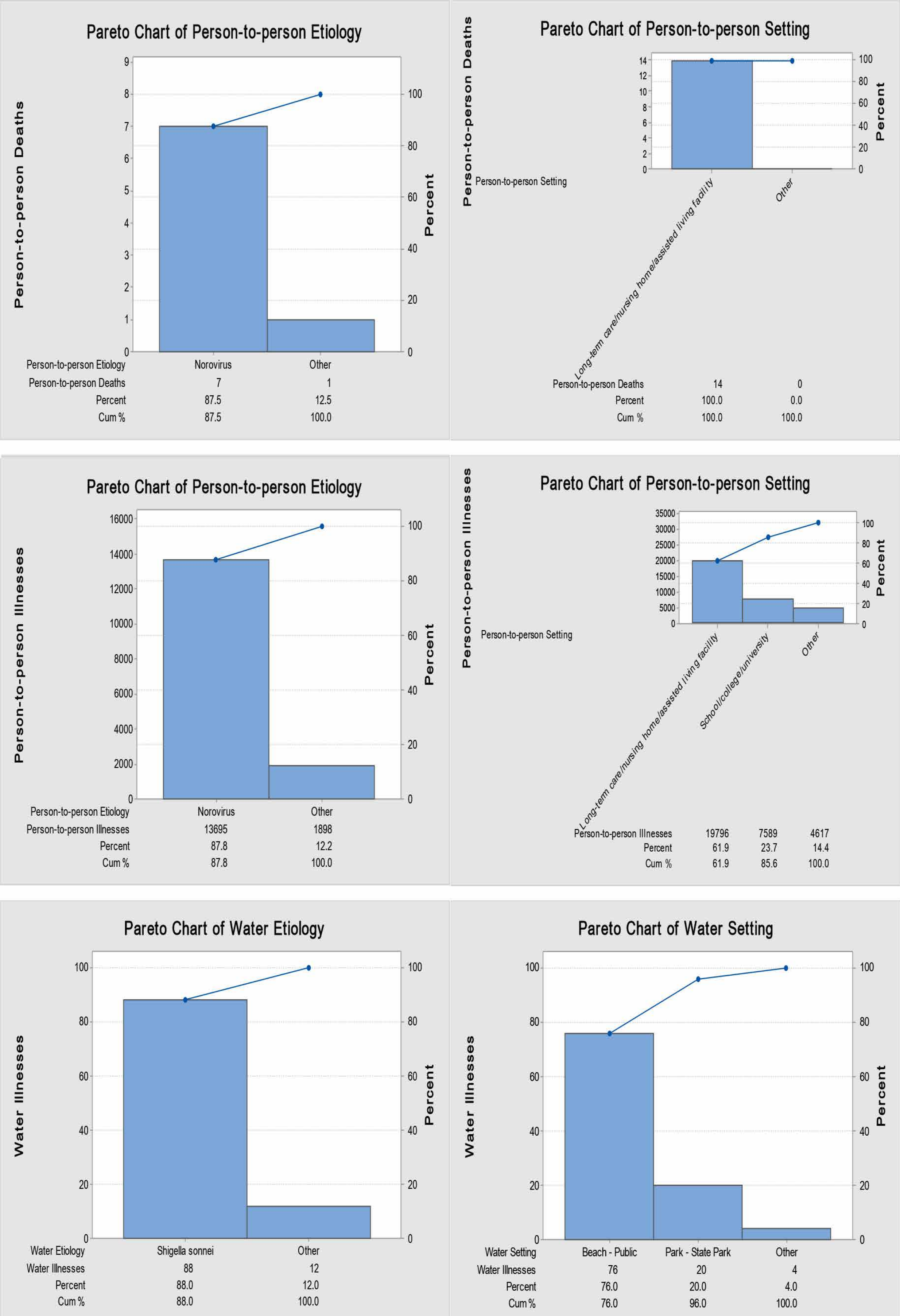


Figure 6: Pareto diagram of water and person-to-person modes for outbreaks in Massachusetts.

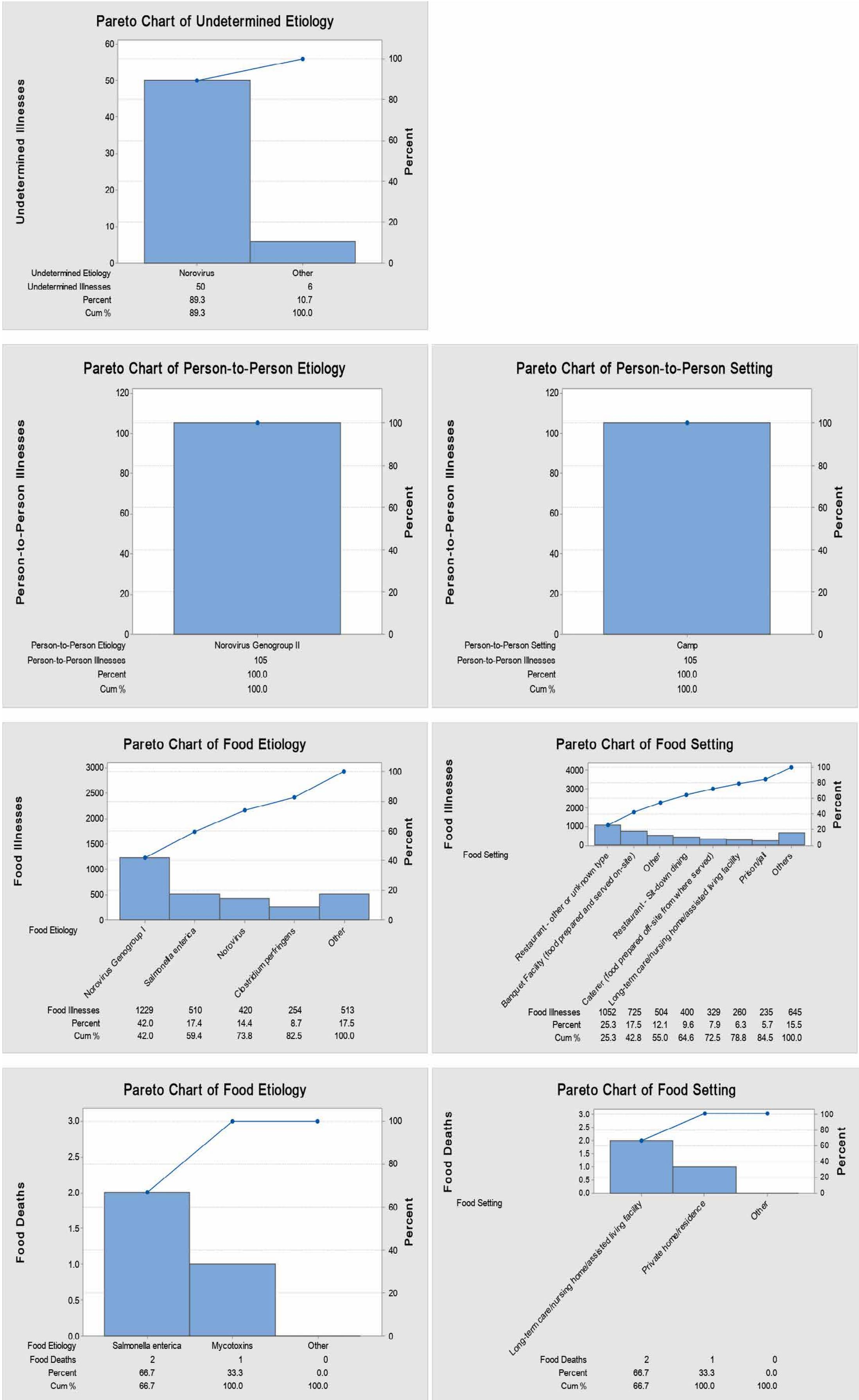


Figure 7: Pareto diagram of person-to-person, food and other undetermined modes for outbreaks in New Jersey.

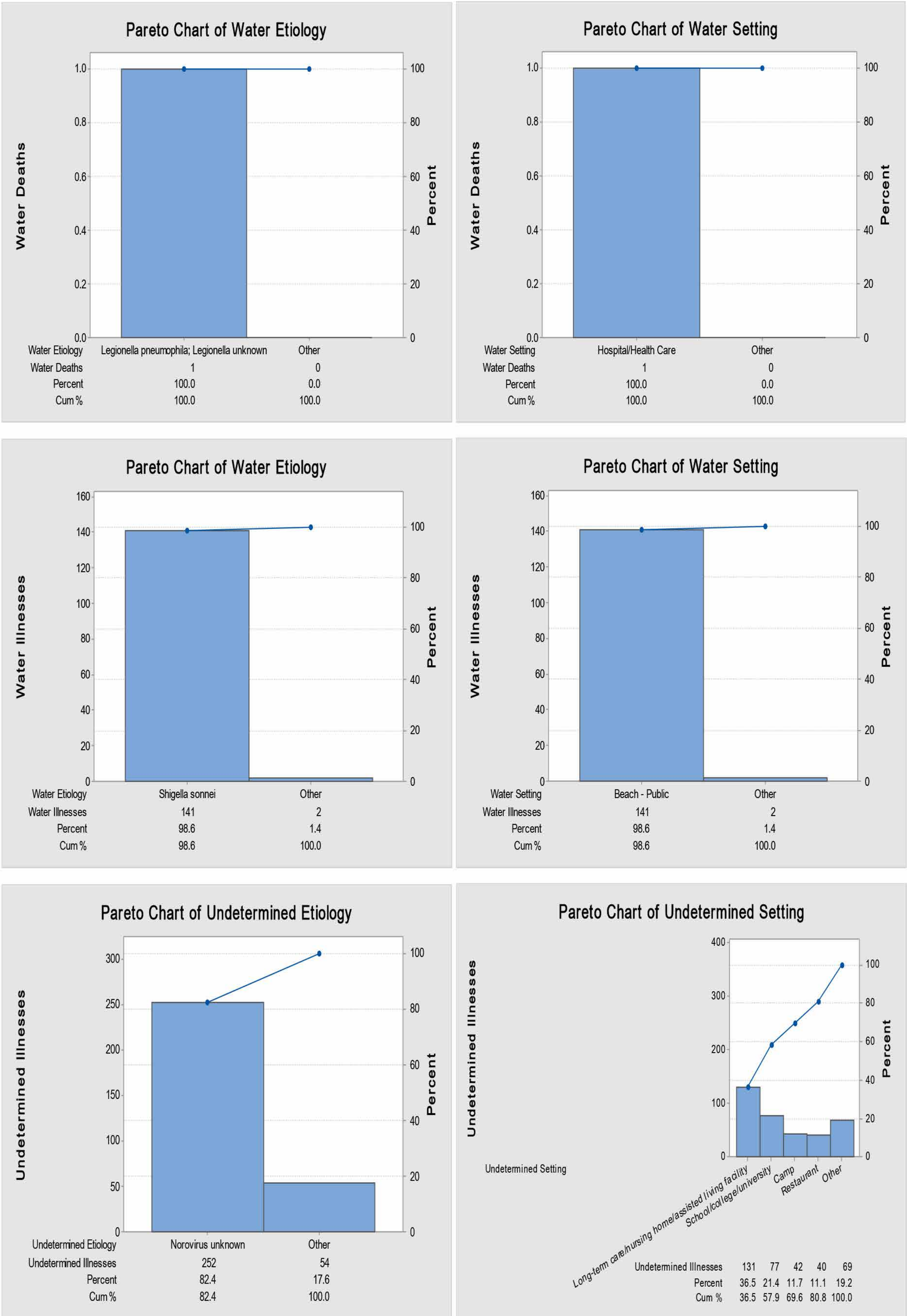


Figure 8: Pareto diagram of water and undetermined modes for outbreaks in Rhode Island.

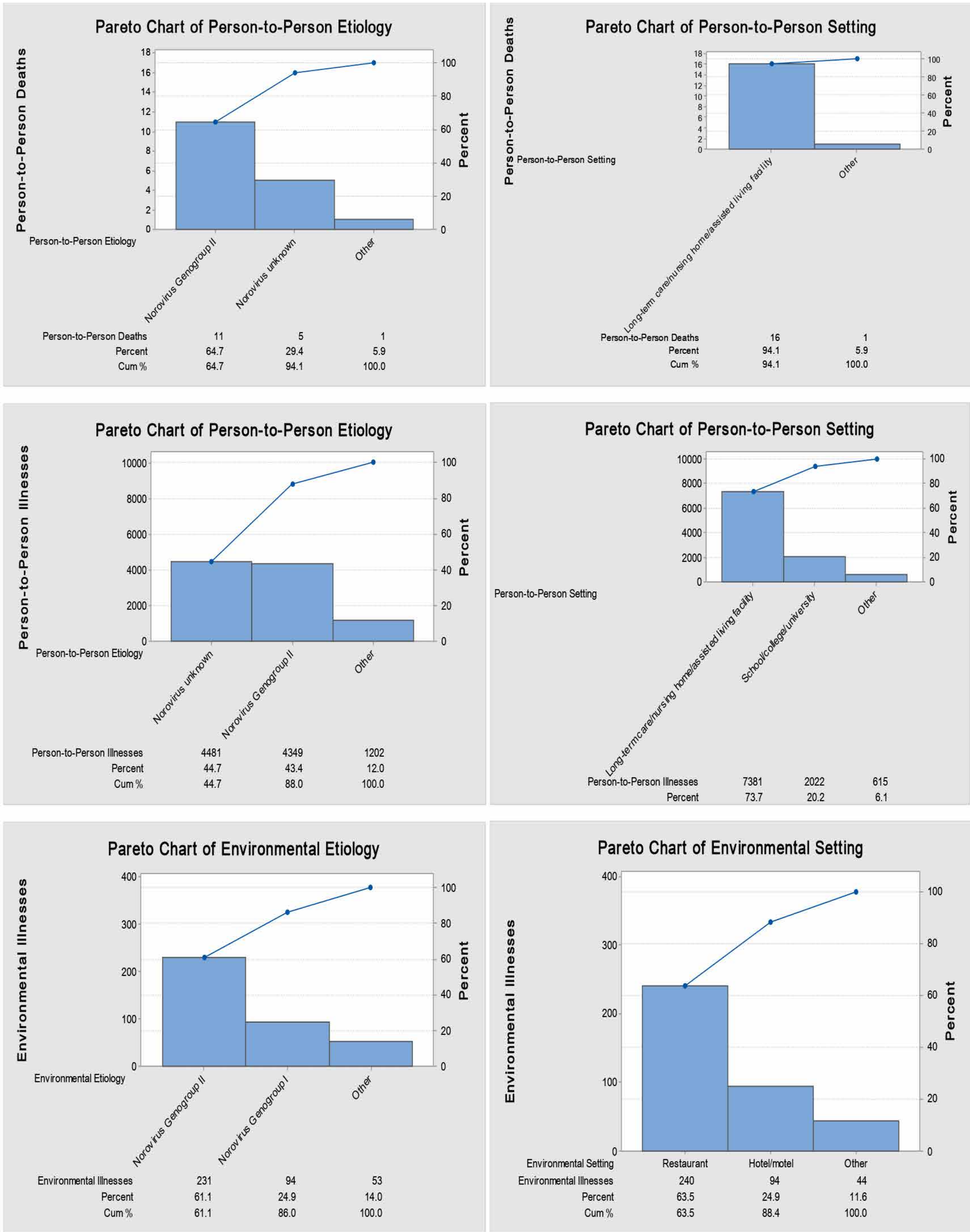


Figure 9: Pareto diagram of person-to-person and environmental modes for outbreaks in Rhode Island.

hand, outbreaks labelled by "3" demonstrate trend shifting in the number of illnesses [25]. It should be noted that in many cases of outbreaks in Massachusetts and Rhode Island (not New Jersey) spiking "1" in the number of illnesses are predisposed by "2" and/or "3" outbreak cases.

Pareto diagram and root cause prioritization

Pareto charts are useful in prioritizing the root causes from a different perspective, basically primary modes, etiological agents and settings of the outbreaks. In Massachusetts, the main etiological agent of more than 60% of outbreaks by animal contact was *Cryptospordium spp.* followed *Salmonella enterica*. For non-identified mode of outbreak spreading *Salmonella enterica* caused single death case but caused more than 90% of outbreak morbidity along with *Shigella sonnei* and Norovirus with over 65% probability in prisons, jails, restaurants, hotels, motels and private residences or homes. *Listeria monocytogenes* accounts for 60% fatality from ingestion of contaminated food vehicles. Norovirus and *Salmonella enterica* contribute by more than 75% of food outbreaks with 60% of the affected settings are restaurants, caterers, schools, colleges and universities. Detailed demonstration could be visualized in figure 5. More than 87% of mortalities and morbidities occurred from person-to-person contact due to Norovirus in assisted living facilities, nursing homes and long-term care houses followed by educational settings. Outbreaks sourced from water in more than 75% of settings in public beaches occurred primarily due to *Shigella sonnei* more than 85% of morbidity with restaurants contributed by more than 60% of cases. Mortality and morbidity were primarily from Norovirus by person-to-person communication in assisted living facilities, nursing homes and long-term care houses.

Overall summary of etiology and location for mortality and morbidity in the three most population-crowded USA states

Norovirus is responsible for over 80% of the morbidities and above 60% of the mortalities followed in lethality by *Salmonella enterica* and *Listeria monocytogenes* which share equal effect on the population death cases 12.8% based on Pareto diagram in figure 10. Based on the affected setting types, it appears that elderly and health-defective populations represent more than three-quarters of the death causalities in long-term care houses, nursing homes and assisted-living facilities. Also, about half of illnesses come from these types of settings, followed by educational institutions then restaurants and finally hospitals.

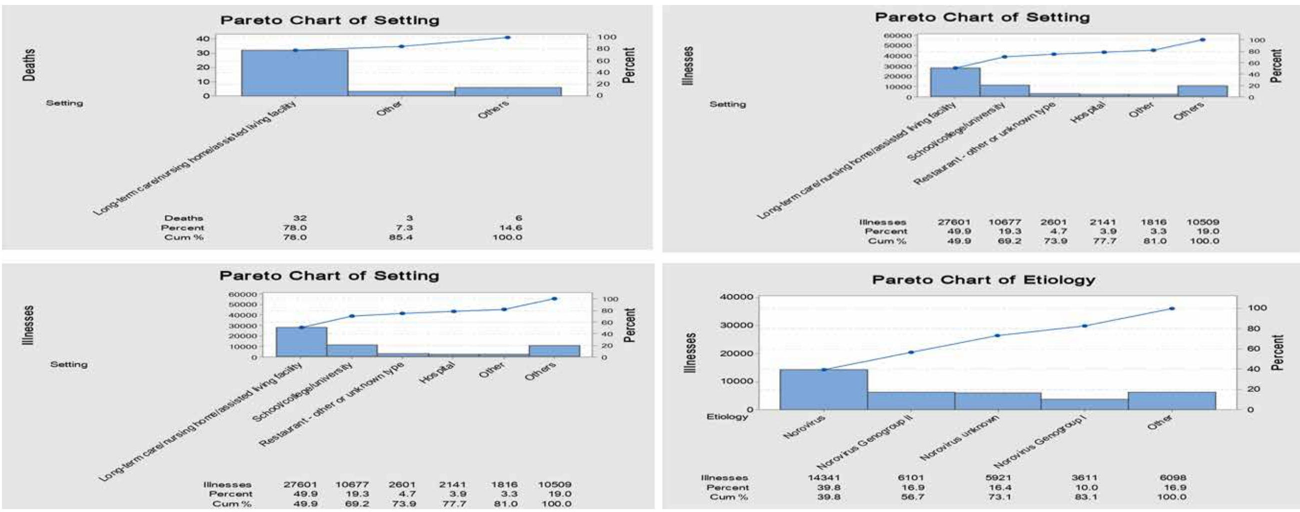


Figure 10: Pareto plot for overall deaths and illness cases from main etiological agents and location sources of outbreaks.

in 88% of the cases as could be seen in figure 6. In New Jersey, Norovirus is the key etiological factor of illness in camps from person-to-person communication other undetermined modes and this is evident in figure 7. Also, Norovirus with *Salmonella enterica* and *Clostridium perfringens* contribute by more than 80% of food-borne illnesses with about 60% of incidents occur in restaurants and caterers. Deaths from food consumption occurred due to *Salmonella enterica* and mycotoxins primarily in assisted living facilities, nursing homes and long-term care houses. Figures 8 and 9 are Pareto charts for outbreaks in Rhode Island showing predominance of Norovirus as etiological agent for affected settings which resides mainly in camps, educational facilities, nursing, assisted living and long-term care houses for unidentified primary modeof transfer cases. Interestingly, it shares with Massachusetts etiology and location type of water-borne outbreak illnesses but there is a case of death in hospital/ healthcare from water due to *legionella spp*. On the other hand, environmental illnesses from Norovirus were responsible for

Final overview on the overall data distribution of morbidity and mortality from the three states in USA

The closest distribution that fits to the morbidity cases in the outbreaks from the three states is Lognormal ($\mu = 2.7864 \pm 0.0087$, $\sigma = 0.9718 \pm 0.3975$) while mortality cases distribution is following Negative Binomial [26, 27]. During the long-term monitoring, the most predominant pattern was closest to Negative Binomial ($k = 0.0325 \pm 0.0108$, $p = 0.7667 \pm 0.2964$) when considering the overall outbreaks from the three states: Massachusetts, New Jersey and Rhode Island [27]. Graphical presentation showing the actual histograms and the assumed distributions are shown in figure 11. Thus, outbreak modelling may be of great value in epidemics prediction [28, 29].

Integration between SPC and failure mode and effect analysis (FMEA) for quantitative risk assessment

The principle of scoring system of FMEA (Risk

Probability Number (RPN) = Severity (S). Occurrence (O). Detectability (D)) [30] can be applied with modification to be quantitative (and hence qRPN) a mean for the risk assessment based on the control charts. The concept that can be applied in such instance is by considering mean as corresponding to S, UCL of the chart as D and the rate of outbreak occurrence as O. The risk assessment can be adopted based on the properties being inspected and this would be reflected through the process-behavior charts. An example could be illustrated when comparing the outbreaks from different states during the same period from figure 4 to determine which states at higher risk from outbreak excursions than.

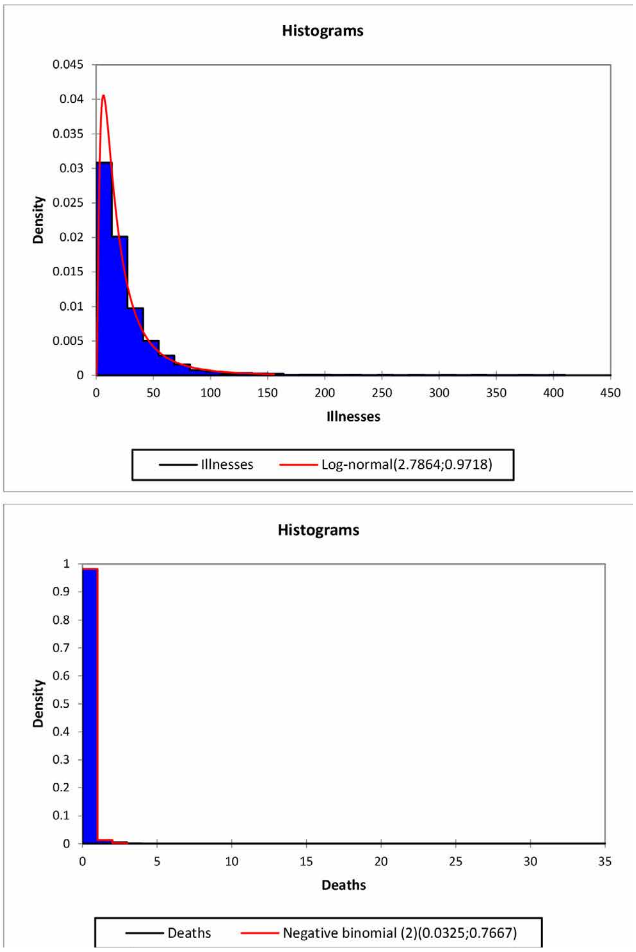


Figure 11: Distribution fitting of the overall illness and death cases from the outbreaks during 20 years of observations.

Conclusion

SPC tools provide useful lessons from data-rich outbreak records - despite some gaps in database - by delivering thorough quantitative and visual analysis of database trend. The comprehensive records derive critical information required by healthcare and regulatory agencies to focus on major contributing factors and parameters in the outbreaks based on the long-term historical dataset. Continuous updating of data records would show numerically the existing state of outbreaks and the impact of the actions taken which may be either improving, deteriorating or have no significant influence. This could be helpful in decision making for further corrective and preventive actions (CAPA), in addition to the modifications required for controlling, monitoring and containing outbreaks which are challenges that parallel the human life on the planet.

References

1. Heymann DL, Chen L, Takemi K, Fidler DP, Tappero JW, et al. 2015. Global health security: the wider lessons from the West African Ebola virus disease epidemic. *Lancet* 385(9980): 1884-1901. https://doi.org/10.1016/s0140-6736(15)60858-3
2. Omoleke SA, Mohammed I, Saidu Y. 2016. Ebola viral disease in West Africa: a threat to global health, economy and political stability. *J Public Health Afr* 7(1): 534. https://doi.org/10.4081/jphia.2016.534
3. Outbreak: 10 of the worst pandemics in history.
4. Eissa MEA. 2019. The attribute control charts for outbreak trends of selected states in the USA: a brief report of the insight into the pattern. *International Medicine* 1(1): 11-14.
5. National Outbreak Reporting System (NORS). 2018. NORS Dashboard. Centers of Diseases Control and Prevention.
6. GraphPad-Prism Software Inc. 2013. GraphPad Prism, Version 6.0. GraphPad-Prism Software Inc., San Diego, CA, USA.
7. MINITAB. 2014. MINITAB Release 17: Statistical Software for Windows. MINITAB Inc., PA, USA.
8. XLSTAT 2014.5.
9. Eissa ME. 2018. Adulterated pharmaceutical product detection using statistical process control. *Bangladesh Pharmaceutical Journal* 21(1): 7-15. https://doi.org/10.3329/bpj.v21i1.37900
10. Eissa MEA. 2018. Microbiological quality of purified water assessment using two different trending approaches: a case study. *Sumerianz Journal of Scientific Research* 1(3): 75-79.
11. Eissa ME. 2018. Variable and attribute control charts in trend analysis of active pharmaceutical components: process efficiency monitoring and comparative study. *Experimental Medicine* 1: 31-44. https://doi.org/10.31058/j.em.2018.11003
12. Eissa ME. 2018. Use of control charts as a quality control tool for monitoring of microbiological infection risk during surgery: an expository case. *Clinical Journal of Surgery* 2(1): 1-3.
13. Rothman KJ. 2012. Epidemiology: an Introduction. Oxford University Press, NY, USA.
14. Hand DJ. 2012. Biosurveillance: methods and case studies edited by Taha Kass-Hout, Xiaohui Zhang. *International Statistical Review* 80(1): 193-194. https://doi.org/10.1111/j.1751-5823.2012.00179_17.x
15. Asrul MAA, Naing NN. 2012. Comparison between negative binomial and poisson death rate regression analysis: AIDS mortality co-infection patients. *IOSR Journal of Mathematics (IOSR-JM)* 3(6): 34-38.
16. An Q, Wu J, Fan X, Pan L, Sun W. 2016. Using a negative binomial regression model for early warning at the start of a hand foot mouth disease epidemic in Dalian, Liaoning Province, China. *PloS One* 11(6): e0157815.https://doi.org/10.1371/journal.pone.0157815
17. Mondaini RP. 2010. Biomat 2009: International Symposium on Mathematical and Computational Biology, World Scientific, Brasilia, Brazil.
18. Bauckhage C, Kersting K, Hadiji F. 2015. Maximum entropy models of shortest path and outbreak distributions in networks. arXiv preprint arXiv:1501.04232.
19. Prasad B. 2017. Incorporating the host-pathogen kinetics of time-dose-response into epidemic outbreak models. Drexel University, PA, USA.
20. Woodall WH, Montgomery DC. 1999. Research issues and ideas in statistical process control. *Journal of Quality Technology* 31(4): 376-386. https://doi.org/10.1080/00224065.1999.11979944

21. Ramana EV, Reddy PR. 2012. Integration of control charts and data mining for process control and quality improvement. *International Journal of Advances in Engineering & Technology* 2(1): 640-648.

22. Mohammed MA, Laney D. 2006. Overdispersion in health care performance data: Laney's approach. *Qual Saf Health Care* 15(5): 383-384. https://doi.org/10.1136/qshc.2006.017830

23. Sellers KF. 2012. A generalized statistical control chart for over-or under-dispersed data. *Quality and Reliability Engineering International* 28(1): 59-65. https://doi.org/10.1002/qre.1215

24. Minitab. 2017. Overdispersion and underdispersion.

25. Eissa ME, Seif M, Fares M. 2015. Assessment of purified water quality in pharmaceutical facility using six sigma tools. *International Journal of Pharmaceutical Quality Assurance* 6(2): 54-72.

26. Lotze TH, Shmueli G, Yahav I, Kass-Hout T, Zhang, X. 2010. Simulating and evaluating biosurveillance datasets. *B*iosurveillance: Methods and Case Studies, Chapman and Hall: Boca Raton, FL, USA.

27. Suma AP, Suresh KP, Gajendragad MR, Kavya BA. 2017. Outbreak prediction of anthrax in Karnataka using poisson, negative-binomial and zero truncated models. *International Journal of Scientific Research* 6(3): 32-36.

28. Ferguson NM, Keeling MJ, Edmunds WJ, Gani R, Grenfell BT. et al. 2003. Planning for smallpox outbreaks. *Nature* 425(6959): 681-685. https://doi.org/10.1038/nature02007

29. Walton NA, Poynton MR, Gesteland PH, Maloney C, Staes C, et al. 2010. Predicting the start week of respiratory syncytial virus outbreaks using real time weather variables. *BMC Med Inform Decis Mak* 10: 68. https://doi.org/10.1186/1472-6947-10-68

30. Mikulak RJ, McDermott R, Beauregard M. 2008. The basics of FMEA. Productivity Press, Taylor & Francis Group, NY, USA.