Risk Based Monitoring in Clinical Trial: An Application with Neural Networking

Atanu B[1], Savanur S[2] and Suman K[3]*

[1]Centre for Cancer Epidemiology, Tata Memorial Centre, India

[2]Department of Biometrics, Chiltern Clinical Research Ltd, India

[3]Department of Biostatistics, Quintiles IMS,India

Submission: September 18, 2017; **Published:** November 10, 2017

*Corresponding author: Suman Kapoor, Department of Biostatistics, Quintiles IMS, India; Email: kapoor.suman@gmail.com

Abstract

There are limited literature on statistical methods to deal with Risk based monitoring (RBM) in clinical trial. Recently, it is becoming one of the emerging areas where a party has limited resources for completion of new approaches to trial oversight. This paper is dedicated to provide the application of neural network-ing as a choice to maintain the RBM in clinical research. It is expected that pharmaceutical industries will adopt this procedure in near future for continuous assessment of RBM in their routine clinical trial.

Keywords: Neural Networking; Clinical Trials; RBM; Risk Control; Quality Assessment

Abbreviations: RBM: Risk based monitoring; FDA: Food and Drug Administration; ANN: Artificial Neural Networks; SDV: Source Data Verification; CSM: Central Site Monitoring; HNC: Head and Neck Cancer

Introduction

The drug safety and quality are main concerned in any Clinical trial. Both the component helps to gain the faith to the regulatory agency. Recently the drug development cost increasing rapidly due to maintaining the longer follow-up, several routine procedures, work pressure to the trial staff and strict inclusion criteria to the patient recruitment. It is the challenge to improve the trial performance through cost reduction. The "Draft guidance for industry: Oversight of clinical investigations - a risk-based approach to monitoring" was presented by food and drug administration (FDA) in 2011. This comprehensive approach gives the scope to analyze and continuously improve trial data safety and quality with reduction in monitoring cost [1,2]. The challenges lies to implement and establish the risk-based monitoring (RBM) in required way [3].

There are limited literature on statistical methods to deal with Risk based monitoring (RBM) in clinical trial. Recently, it is becoming one of the emerging areas where a party has limited resources for completion of new approaches to trial oversight. The pharmaceutical industry now realized the extensive requirement to monitor an assessment of site risk. Many companies now involved through implementing the monitoring process with central manual with data review process. The objective of data review process is to identify any specific pattern that leads to the strategic concentration to involve the additional re-sources and e ort required at the best place. The

data review process is based on checking the information from hard copy that is really time consuming as well there are higher chance of error.

Neural Networks are a machine learning framework process. It attempts to mimic through the learning pattern of biological neural networking. The process of natural Biological neural networks are interconnected through the neurons with dendrites that receive inputs. It is based on the input procedure. The input provides the output signal through an axon to another neuron. The mimic process happens in data driven method and known as Artificial Neural Networks (ANN). Thereafter we will refer it as neural network [4-9].

The monitoring is an ongoing procedure in any clinical trials and it is the responsibility for investigator. The objective of the monitoring is to control the quality of clinical trial and ensure the safety of the participants. There are several approaches to manage the quality of clinical trial. The traditional approach is based on site monitoring through 100% source data verification (SDV). The SDV assist for safety to the participants and also produce the quality data. However, it is time consuming and costly procedure. The quality control can be managed through the central monitoring through continual process to train the study population and routine site audits [10]. In last two decades the difficulties and cost to conduct clinical trial increased rapidly particularly for the development of new drug. The RBM can be

an alternative choice to facilitate the monitoring procedure through reduced cost. The regulatory authorities like Food and Drug Administration [11,12] has also recommended to conduct the monitoring procedure through RBM paradigm.

Central statistical monitoring

The basic idea about CSM can be cited with Venet et al. [12,13]. The CSM can be applied in metacentric clinical trial. The procedure is to compare the distribution of variables of all including visit level, site level and patients' level with reference to all units. If any specific pattern of units observed then further monitoring is required to take place. The statistical methodology through application of data exploration by univariate and multivariate statistical modeling can also be performed. There are illustration about application of Mahalanobis distance to detect the outliers in the data sets [13]. The graphical representation also adopted to check the difference in variation or the shape of distribution for CSM [14]. In this study, the statistical methods discussed above can explore whether the specific clinical sites should receive more attention for data review or not.

Challenges for RBM

Based of our knowledge there is not any dedicated software packages avail-able for RBM. However, there is a statistical package available for Central Site Monitoring (CSM) is JMP Clinical. (SAS Institute Inc., Cary, NC, USA) and some codes of SAS can also be performed in R. Recently, there is tendency to conduct CSM and RBM jointly. It is our view that CSM can supervised through the adoption of neural networking as jointly with RBM or independently. The data extract procedure can also be taken like earlier studies [15-17]. It is natural that the clinical trial data measured repeatedly for patients through follow-up measurement. The follow-up measurement data are prone to have measurement errors, missing data and variability over time. These can make negative impact on data quality. The exploratory data analysis during monitoring is required to perform to measure the frequency of measurement errors, missing data and variability over time. Based on presences of these factors the sites can be separated and allocated differently for prediction about risk in near future.

Data methods

The simulation techniques adopted within this study are given below. All simulations were carried through the freely available R statistical software, thus allowing all researchers access to any suitable methods identified.

Generating the datasets

There is no consensus on the most appropriate approach with data driven method in RBM. The simulation study was preferred to assess the performance of the data driven method. The hypothetical datasets were generated to resemble the skewed distributions seen in a motivating on site clinical trial monitor-ing. The clinical trial with Phase II, Blinded Endpoint

(PROBE), Multicenter, Randomized, Prospective pro les are selected to continue the study. Participant were randomized into (1:1) to either receive chemotherapy treatment A or treatment B to treat the head and neck cancer (HNC) patients. The primary objective of the study is to prolonged the duration of median survival. A total of 30 sites data are generated to conduct the trial. The sites names are randomly allotted with 'A-Z' and AA,BB,CC and DD respectively. The subject's participation is considered with 6 months duration. An arbitrarily risk scoring process was generated to address the application of statistical method. The risk score for different sites are considered. Proper validated risk scores can be generated based on study objectives and requirement. In this study the site specific risk
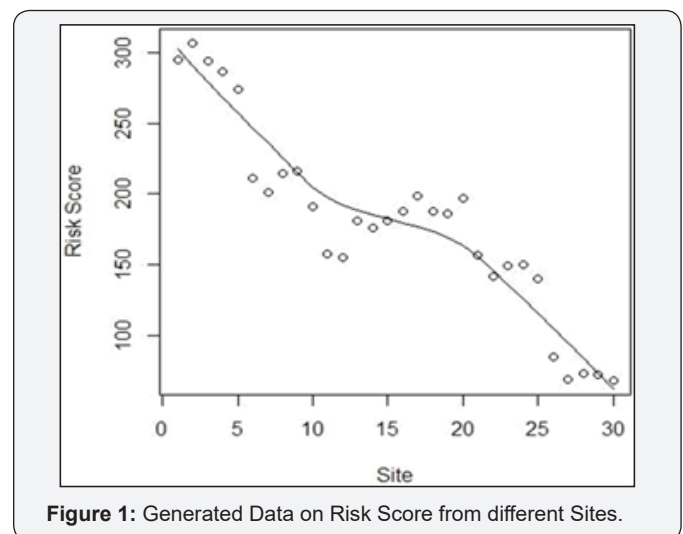

Figure 1: Generated Data on Risk Score from different Sites.

score is generated through the consideration of four parameters like (I) Deviation - Onsite Death during chemotherapy(A). (II) Protocol Deviation (B) (III) Unreported AEs based on 90 day follow-up (C) (IV) Death during chemotherapy (D). Each risk parameter in the Risk scoring process was generated based on the degree of impact it could have on the outcome of study conduct. Details of the risk scoring process are given in Table 1. It is assumed that the Risk Score for 30 site follows the normal distribution with mean value 50. The data generated with range (31,70). The generated data for 30 sites are given in Table 2 and in Figure 1.

Table 1: Risk Score System for continuous assessment.

Item	Parameter	Score
1	$A = \dfrac{\text{Number of Death}}{\text{Number of patient treated in a specic duration}}$	$A \times 4 = A'$
2	$B = \dfrac{\text{Number of Protocol Deviation}}{\text{Total Number of patient Treated}}$	$B \times 2 = B'$
3	$C = \dfrac{\text{Number of Unreported AEs}}{\text{Number of patient treated in a specific duration}}$	$C \times 2 = C'$
4	$D = \dfrac{\text{Number of death}}{\text{Number of patient treated in a specific duration}}$	$D \times 4 = D'$
	Risk Score	$A'+B'+C'+D'$

Table 2: Generated Data on Risk Score from different Sites.

Site	A'	B'	C'	D'	Risk	Risk Score	Prediction
A	66.16	59.02	81.07	1.23	1	207.24	-
B	34.42	48.12	65.65	1.91	0	150.18	-
C	57.32	63.11	80.21	1.4	1	203.63	-
D	42.71	45.75	61.04	1.43	0	153.5	-
E	66.95	18.58	87.7	1.31	1	178.24	-
F	24.9	57.47	0.15	0.01	0	88.53	-
G	48.43	26.81	57.23	1.18	0	139.47	-
H	24.5	32.9	29.71	1.21	1	95.11	-
I	40.65	55.5	47.56	1.17	0	152.71	-
J	25.08	39.78	14.09	0.56	0	88.94	-
K	64.13	25.68	43.51	0.68	0	144.32	-
L	59.44	60.47	92.54	1.56	1	224.45	-
M	61.05	26.36	58.93	0.97	0	159.34	-
N	27.27	61.58	47.6	1.75	0	150.44	-
O	63.06	39.2	18.5	0.29	0	135.77	-
P	50.5	28.22	39.77	0.79	0	134.49	0
Q	43.55	39.57	39.36	0.9	0	139.48	0
R	43.38	60.85	32.78	0.76	0	155	0
S	20.54	61.69	31.57	1.54	0	132.81	0
T	58.89	26.08	49.66	0.84	0	154.62	1
U	23	31.76	11.48	0.5	0	87.24	0
V	32.2	52.42	42.44	1.32	0	149.06	0
W	23.33	48.58	2.23	0.1	0	97.13	0
X	27.85	51.97	49.6	1.78	0	153.42	0
Y	65.3	48.84	54.65	0.84	0	193.8	0
Z	47.45	27.03	53.61	1.13	0	154.1	1
AA'	63.29	36.5	95.95	1.52	1	222.74	1
BB'	45.73	55.84	63.77	1.39	0	193.34	1
CC'	59.42	30.86	20.83	0.35	0	140.1	0
DD'	58.84	54.51	108.71	1.85	1	252.07	1

In this work the perception is used to handle the one or more inputs, bias, activation function and a single output. It receives inputs, multiplies them by some weight, and then passes them into an activation function to produce an output. The linear model is used as activation function in this work. The risk score is modeled through risk indicator. The perceptron is adopted to include different layers, creating a multi-layer perceptron model of a neural network. The data convergence is not possible to conduct in the neural networking. It is required to input the normalized data as an input in neural networking. In this work the normalized data is incorporated to get the prediction of risk function. The data is splitted into equal part with training and test set. Initially, we run our data into training set and thereafter in the test set to test the performance of the test set. Initial 15 sites observations are considered for training set and rest as test set. The neural network function in free available software R is used to perform the neural networking (Figure 2).

The plot(nn) command is used to visualize the neural net data. The weighted vectors between neurons are presented through weighted vectors. The bias included is represented through blue lines. However, it is difficult to interpret the weight. The blue line represents the bias.
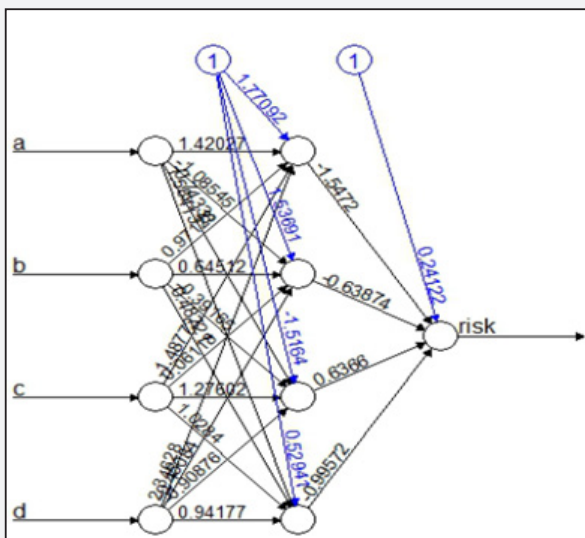
Figure 2 : Neural Networking with 4 hidden nodes.
Error: 0.61489 Steps: 10

Illustration

Risk scoring is the process of analyzing a site about clinical trial performance in order to assess the allocation of more resources of the site. It can take numerous approaches on analyzing the risk of a site. It basically comes down to select the correct independent variables (e.g. Deviation, Protocol Deviation, Unreported AEs and Death during study. A RBM can be represented by machine learning, logistic regression, linear regression or a combination of these. The work is demonstrated through the application of open source soft-ware R. The package neural net is used to perform this analysis. The dataset contains information on different sites who involved in a chemotherapy trial. The variables "a" (Deviation),"b"(Protocol Deviation), "c"(Unreported AEs)," d"(Death during study) and risk score is available. The objective is to devise a model which predicts, based on the input variables "a","b","c" and "d", whether or not a trisk will occur in a specified time point. The dataset will be split up in a subset used for training the neural network and another set used for testing. As the ordering of the dataset is completely random, we do not have to extract random rows and can just take the first x rows.

The neural network is performed with 4 hidden nodes. There are several thumb rules to select the number of hidden. The threshold value of the output is decided with 10%. The output can be visualize in the gure1. The initial 15 sites data (split from the full data set) are adopted as train data and thereafter the test data with ret 15 sites are used to predict their performance. This is an illustration of a total of 30 sites risk scores that are simulated through the parameters A',B',C' and D' are given in Table 2. The simulated data for A',B', C' and D' are generated through random number generation and generated data are given in Figure 3. The Risk score is obtained as =A'+B'+C'+D'. An arbitrary choice of Risk Score more than 200 is considered as Risk of site is present otherwise it is absent. The parameter Risk is valued based on the status of risk score more than 200 as 1 otherwise 0.This is an arbitrary choice to train the neural networking data to provide the prediction from neural networking. The input data for the neural networking is considered through the parameters (I)Site(II) A'(III)B'(IV)C'(V) D' (VI) Risk (VII)Risk Score. The output obtained through application of neural networking is the parameter "Prediction". The prediction parameter for the sites P to DD' is obtained through neural networking and given in Table 2. The prediction value of "0" shows that the sites is having no risk and it is not required further close monitoring. The prediction value of "1" (e.g. site AA') is having risk that it is required for further close monitoring.
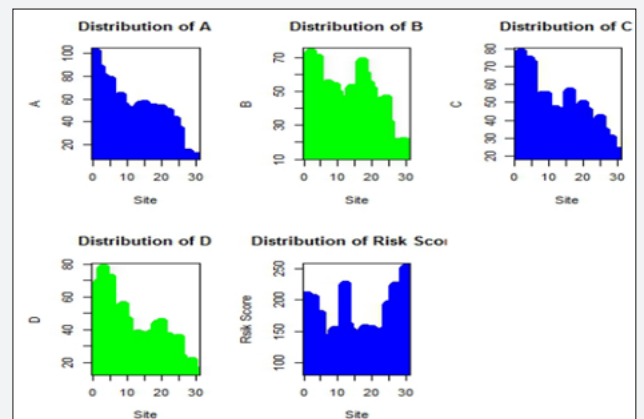

Figure 3 : Distribution of Simulated Data.

Appendix-I

	site	a	b	c	d risk Risk. Score		
15	15	63.06	39.20	18.50	0.29	0	135.77
16	16	50.50	28.22	39.77	0.79	0	134.49
17	17	43.55	39.57	39.36	0.90	0	139.48
18	18	43.38	60.85	32.78	0.76	0	155.00
19	19	20.54	61.69	31.57	1.54	0	132.81
20	20	58.89	26.08	49.66	0.84	0	154.62
> set.seed(1234567890)							
> library("neuralnet")							

		a	b	c	d		
		16 50.50	28.22	39.77	0.79		
		17 43.55	39.57	39.36	0.90		
		18 43.38	60.85	32.78	0.76		
		19 20.54	61.69	31.57	1.54		
		20 58.89	26.08	49.66	0.84		
		21 23.00	31.76	11.48	0.50		

Discussion and Conclusion

In RBM [18-21], the sharing information about risk and risk management between decision makers are required to be established [22,23]. The information about severity, probability, treatment, control, acceptability and detectability are required to be communicate. The communication between decision makers requires to build on risk control to reduce and/or accept the risks. During the protocol buildup, the past therapeutic experience and literature review about the acceptable risk is required to be developed. It can adopt as a threshold to raise a trigger. Different domain can be considered to provide the priority like cost-benefit association and finalize the optimum level of risk to be controlled. The acceptable levels of risk need to be defined to control the risk. The layout about steps to reduce or eliminate the risk needs to be incorporated. The appropriate balance between risk and resources also important. The processes for easing of quality risk when it goes over an acceptable level in known as risk reduction. The intensity and mode of monitoring need to be de ne and proper action can be taken based on relevant protocol consideration and literature review. A decision to accept the risk is known as Risk acceptance. The residual risk acceptance can be a formal decision.

It is difficult to provide accurate prediction for sites those having huge number of missing data in clinical trial. The ranking of this sites on risk based can be conducted through isolation of the presences of missing data in the sites. The handling with missing data to set the sites with risk based is challenging. It is required to the make extension of statistical methodology to rank the sites with having risk in presence of missing data. These following points can be concluded from this analysis. The acceptance of RBM in routine work is not cumbersome. It can be simply adopted through data-review process. The application of neural networking can considerably help in continually assessing risk of sites based on monitor's feedback. The application of data driven method like neural networking can speed up centralized manual review. This illustrated procedure provides path about adoption of data-driven models to assess risk. The centralized manual data review can be complemented by statistical-driven model. It can incline the distinguishing a site with risk pro le. The approaches with RBM can enabled through the use of clinical technology solution to build planning, data visualization and data -driven methods.

Conflict of Interest

The authors declare that they have no conflict of interest.

References

1. Food and Drug Administration and others (2013). Guidance for industry: oversight of clinical investigations-a risk-based approach to monitoring. Silver Spring, MD: FDA.

2. Franco P, Hronec M, Karacsony A (2013) Risk-based monitoring: Reduce clinical trial costs while protecting safety and quality.

3. Nayak VS, Khan MS, Bharat KS (2016) Quality management in risk based monitoring. IJCT 3(2): 59-63.

4. Sholom MW, Ioannis K (1989) An empirical comparison of pattern recognition, neural nets, and machine learning classification methods. Readings in machine learning. Citeseerx 781-787.

5. Jen LY, Fine TL (1998) Neural-network design for small training sets of high dimension. IEEE Trans. Neural Netw 9(2): 266-280.

6. Schmidhuber J (2015) Deep Learning in Neural Networks: An Overview. Neural Netw 61: 85-117.

7. Yanbo Huang (2009) Advances in Artificial Neural Networks-Methodological Development and Application. Algorithms 2(3): 973-1007.

8. Kozma R, Bressler S, Perlovsky L, Venayagamoorthy GK (2009) Advances in neural networks research: An introduction. Neural Netw 22(5-6): 489-490.

9. Bing C, DM Titterington (1994) Neural Networks: A review from a Statistical Perspective. Statist Sci 9(1): 2-30.

10. Journot V, Pignon JP, Gaultier C, Daurat V, Bouxin-Métro A, et al. (2011) Validation of a risk-assessment scale and a risk-adapted monitoring plan for academic clinical research studies-The Pre-Optimon study. Contemp Clin Trials 32(1): 16-24.

11. Morrison BW, Cochran CJ, White JG, Harley J, Kleppinger CF, et al. (2011) Monitoring the quality of conduct of clinical trials: a survey of current practices Clin Trials 8(3): 342-349.

12. Venet D, Doffagne E, Burzykowski T, Beckers F, Tellier Y, et al. (2012) A statistical approach to central monitoring of data quality in clinical trials Clin Trials 9(6): 705-713.

13. Buyse M, George SL, Evans S, Geller NL, Ranstam J, et al. (1999) The role of biostatistics in the prevention, detection and treatment of fraud in clinical trials Stat Med 18(24): 3435-3451.

14. Kirkwood AA, Cox T, Hackshaw A (2013) Application of methods for central statistical monitoring in clinical trials. Clin Trials 10(5): 783-806.

15. Sheetz N, Wilson B, Benedict J, Huffman E, Lawton A, et al. (2014) Evaluating source data verification as a quality control measure in clinical trials. Therapeutic Innovation Regulatory Science 48(6): 671-680.

16. Evans S (2001) Statistical aspects of the detection of fraud. Fraud and misconduct in medical research. (3rd edn) London: BMJ Publishing Group, pp. 186-204.

17. Buyse M (2014) Centralized statistical monitoring as a way to improve the quality of clinical data. Applied Clinical Trials.

18. Prajna PKumar, Jeroze D, Manoj P (2016) Risk Based Monitoring (RBM): A global study focusing on perception and merits among clinical investigational sites. Contemporary Clinical Trials Communications 4: 155-160.

19. Guideline for Good Clinical Practice E6 (R2). ICH Integrated Addendum to ICH E6 (R1).

20. Marc B (2013) Reflection Paper on Risk-based Quality Management in Clinical Trials. Applied Clinical Trials.

21. Hurley C, Shiely F, Power J, Clarke M, Eustace JA, et al. (2016) Risk based monitoring (RBM) tools for clinical trials: A systematic review. Contemp Clin Trials 51: 15-27.

22. Alemayehu D, Alvir J, Levenstein M, Nickerson D (2013) A data-driven approach to quality risk management. Perspect Clin Res 4(4): 221-226.

23. Davis JR, Nolan VP, Woodcock J, Estabrook RW (1999) Assuring Data Quality and Validity in Clinical Trials for Regulatory Decision Making; Workshop Report. Washington D.C: National Academy Press, USA.

Your next submission with Juniper Publishers
will reach you the below assets

- Quality Editorial service
- Swift Peer Review
- Reprints availability
- E-prints Service
- Manuscript Podcast for convenient understanding
- Global attainment for your research
- Manuscript accessibility in different formats
  (Pdf, E-pub, Full Text, Audio)
- Unceasing customer service

Track the below URL for one-step submission
https://juniperpublishers.com/online-submission.php